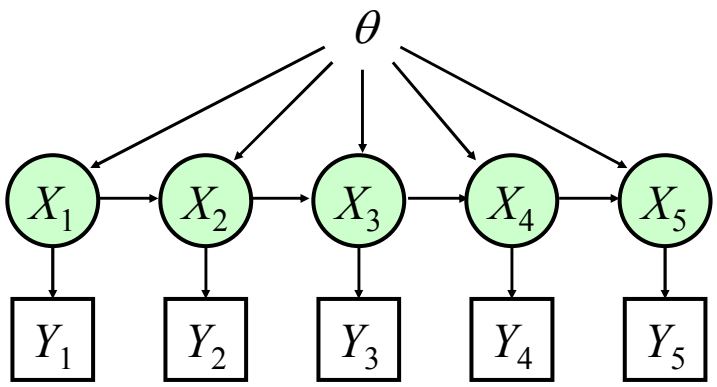
第1回 Latent Dynamics Workshop

# 階層ベイズモデリングによる時系列からの再構成

京都大学　大学院情報学研究科
石井　信

## 状態空間モデル（第一のlatent dynamics)



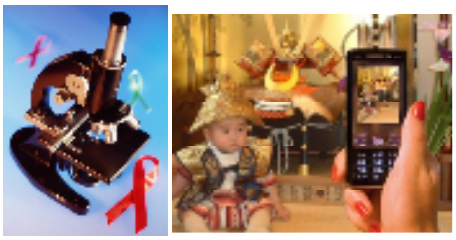
# 遮蔽物を除去しながらの
# ベイズ超解像法

Kanemura, A., Maeda, S., Fukuda, W., Ishii, S.
*Journal of Systems Science and Complexity*, **23**(1), 116-136, 2010.

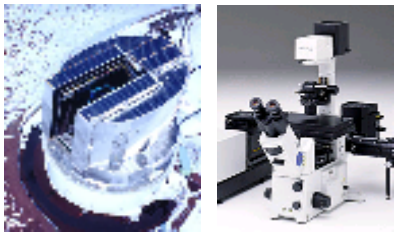
兼村厚範: ATR脳情報解析研究所

## 超解像

- 超解像:撮像機器に内在する物理的限界を超えた高解像度の画像を生成する
  - 高解像度
    1) ピクセル数増大, 2) ボケ除去, 3) ノイズ低減



安いハードウエア
超解像で画質向上

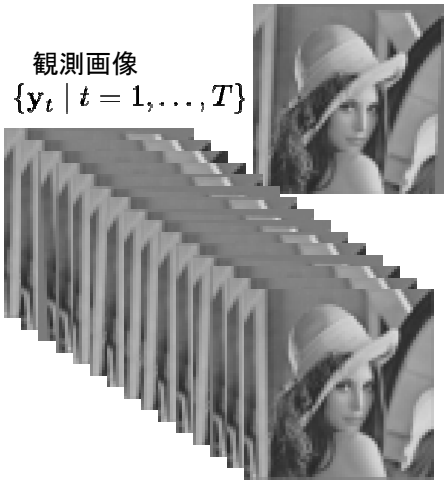


高性能ハードウエア
超解像で物理的限界を突破

# マルチフレーム超解像法

- 複数の互いに位置ズレの異なる観測画像を用いて高精細な原画像を推定 (Tsai & Huang, 1984)

観測画像
$\{\mathbf{y}_t \mid t = 1, \ldots, T\}$



1枚を拡大した画像 $\hat{\mathbf{x}}$

16枚を合成した超解像画像 $\hat{\mathbf{x}}$

# 超解像問題のベイズ推定

- データ $D = \{\mathbf{y}_t \mid t = 1, \ldots, T\}$ が与えられたもとでの原画像 **x** の事後確率 $p(\mathbf{x} \mid D)$
  - Bayesの定理

事前確率　　尤度

$$p(\mathbf{x} \mid D) = \frac{p(\mathbf{x})p(D \mid \mathbf{x})}{p(D)} = \frac{p(\mathbf{x})\prod_t p(\mathbf{y}_t \mid \mathbf{x})}{p(D)}$$

  - 高解像度画像の推定量

$$\hat{\mathbf{x}} = \mathrm{E}[p(\mathbf{x} \mid D)]$$

# 事前確率と尤度

$$p(\mathbf{x}\,|\,D) = \frac{p(\mathbf{x})p(D\,|\,\mathbf{x})}{p(D)} = \frac{\overset{\text{事前確率}}{p(\mathbf{x})}\prod_t \overset{\text{尤度}}{p(\mathbf{y}_t\,|\,\mathbf{x})}}{p(D)}$$

$p(\mathbf{x})$ と $p(\mathbf{y}|\mathbf{x})$ の設計が重要

– 事前確率：正則化

例 $p(\mathbf{x}) = \mathrm{Gauss}(\mathbf{x}\,|\,\mathbf{0}, A^{-1})$ （平滑化制約；Tikhonov正則化）

– 尤度: 観測過程

例 $p(\mathbf{y}_t\,|\,\mathbf{x}) = \mathrm{Gauss}(\mathbf{y}_t\,|\,W_t\mathbf{x}, \beta^{-1}I)$ （線形変換＋ガウスノイズ）

$$\Rightarrow \hat{\mathbf{x}} = \left((\rho/\beta)A + \sum_t W_t^{\mathrm{T}} W_t\right)^{-1} \sum_t W_t^{\mathrm{T}} \mathbf{y}_t$$

（正則化付逆問題の典型的な解; Hardie et al. (1997), Elad (1997)）

# 階層化：隠れ変数の導入

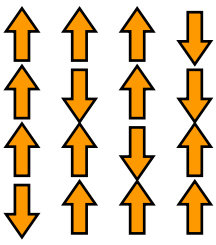
- 隠れ変数を導入してモデルを階層化する
  - 事前分布に導入

$$p(\mathbf{x}) = \sum_{\boldsymbol{\eta}} p(\boldsymbol{\eta})p(\mathbf{x}\,|\,\boldsymbol{\eta})$$

    - 隠れ変数 $\boldsymbol{\eta}$ は画像中のエッジ
  - 尤度に導入

$$p(\mathbf{y}\,|\,\mathbf{x}) = \sum_{\mathbf{z}} p(\mathbf{z})p(\mathbf{y}\,|\,\mathbf{x},\mathbf{z})$$

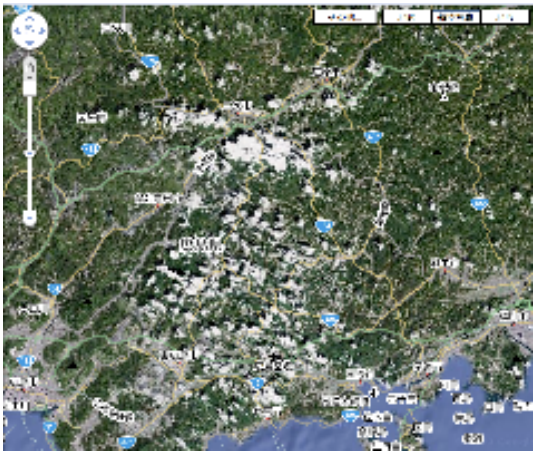
    - 隠れ変数 $\mathbf{z}$ は遮蔽パタン



- 事後分布計算が困難
  - 試験分布を用いて変分法的に学習（平均場近似）

# 遮蔽

- 観測はしばしば遮蔽される
  - 衛星画像中の雲
  - 観光地を横切る人
  - CCDが
    部分的にノイジー



(Google Maps)

# 遮蔽

- 観測はしばしば遮蔽される
  - 衛星画像中の雲
  - 観光地を横切る人
  - CCDが
    部分的にノイジー

# 非階層尤度

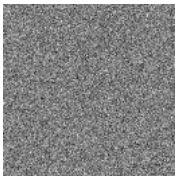
$$p(\mathbf{y}_t \mid \mathbf{x}) = \mathrm{Gauss}(\mathbf{y}_t \mid W_t\mathbf{x}, \beta^{-1}I)$$

512x512画像 $\mathbf{x}$

通常の尤度モデルは
一様な精度（逆分散）$\beta$ を仮定

128x128画像 $W_t\mathbf{x}$

＋



ノイズ $\boldsymbol{\varepsilon}_t$

＝

観測 $\mathbf{y}_t$

# 階層尤度

- 尤度を階層化し隠れ変数を導入 $p(\mathbf{y}_t \mid \mathbf{x}, \mathbf{z}_t) = \mathrm{Gauss}(\mathbf{y}_t \mid W_t\mathbf{x}, B(\mathbf{z}_t))$

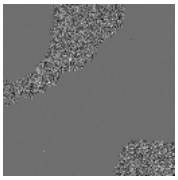
512x512画像 $\mathbf{x}$

128x128 ノイズパタン $\mathbf{z}_t$

黒: 低精度 10dB
白: 高精度 40dB

128x128 画像 $W_t\mathbf{x}$

＋



ノイズ $\boldsymbol{\varepsilon}_t$

＝

128x128x$T$ 観測画像 $\{\mathbf{y}_t\}$

# ノイズパタン

- ノイズパタンをスピンでモデリング
  - イジングスピン

    $\mathbf{z} = \{z_i \in \{+1, -1\} \mid i = 1, \ldots, P_L\}$

128x128 ノイズパタン $\mathbf{z}$

  - 事前分布は
    - エネルギー関数で記述

例:

$J_i = 0.01, J_{ij} = 1, J = 0$

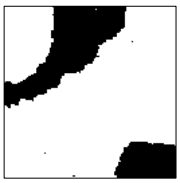
$$p(\mathbf{z}_t \mid \mathbf{z}_{t-1}, \varphi_{t-1}) = \frac{1}{Z} \exp\{-E(\mathbf{z}_t, \mathbf{z}_{t-1}, \varphi_{t-1})\}$$

$$E(\mathbf{z}_t, \mathbf{z}_{t-1}, \varphi_{t-1}) = -\sum_i J_i z_{ti} - \sum_{i \sim j} J_{ij} z_{ti} z_{tj} - J \mathbf{z}_t (G(\varphi_{t-1}) \mathbf{z}_{t-1})$$

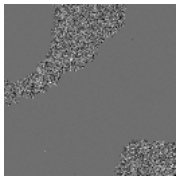
パターン内相関　　パターン間相関

    - $G$ はパターンの状態遷移を表す行列
    - 白領域の方が大きい　⇒　$J_i > 0$
    - パターン内に相関がある ⇒　$J_{ij} > 0$

# 観測ノイズ



128x128 ノイズパタン $\mathbf{z}_t$



ノイズ $\boldsymbol{\varepsilon}_t$

- ピクセル $i$ での観測ノイズ

$$\beta_{ti}(z_{ti}) = \begin{cases} \beta_{\mathrm{H}} & z_{ti} = +1 \\ \beta_{\mathrm{L}} & z_{ti} = -1 \end{cases} \qquad \beta_{\mathrm{H}} > \beta_{\mathrm{L}}$$

  - $z$ に応じて可変
- 観測モデル

$$\mathbf{y}_t = W_t \mathbf{x} + \varepsilon_t, \quad \boldsymbol{\varepsilon}_t = \mathrm{Gauss}(\mathbf{0}, B_t)$$

$$B_t(\mathbf{z}_t) = \mathrm{diag}(\beta_1(z_{t1}), \ldots, \beta_{P_L}(z_{tP_L}))$$

  - ピクセルごとにノイズ量が異なる

# 変分学習

- $p(\mathbf{x},\{\mathbf{z}_t\}\,|\,\{\mathbf{y}_t\})$ を $q(\mathbf{x},\mathbf{z}) = q(\mathbf{x})\prod_{t,i} q(z_{ti})$ で近似
- 学習則：

$$q^*(\mathbf{x}) = \mathrm{Gauss}(\mathbf{x}\,|\,\boldsymbol{\mu},\Sigma) \qquad \Sigma = \left(\rho A + \sum_t W_t^{\mathrm{T}}\langle B_t\rangle W_t\right)^{-1}$$

$$\boldsymbol{\mu} = \Sigma\sum_t W_t^{\mathrm{T}}\langle B_t\rangle \mathbf{y}_t$$

$$\langle B_t\rangle = \mathrm{diag}(\langle\beta_{t1}\rangle,\ldots,\langle\beta_{tP_{\mathrm{L}}}\rangle)$$

$$\langle\beta_{ti}\rangle = \beta_{\mathrm{H}}q(z_{ti}=+1) + \beta_{\mathrm{L}}q(z_{ti}=-1)$$

– 精度の期待値（事後確率による $\beta_{\mathrm{H}}$ と $\beta_{\mathrm{L}}$ の内分）により，効果的に正則化の強さを調節

# 変分学習

- 隠れ変数 **z** の最適試験分布

$$q^*(z_{ti}) = \mathrm{Ber}(z_{ti}\,|\,v_{ti}), \quad v_{ti} = \mathrm{sig}(2\lambda_{ti})$$

– シグモイド関数によるソフトな推定

– パラメタ $\lambda_{ti}$ は

$$\lambda_{ti} = J_i + \sum_{j\in N(i)} J_{ij}\langle z_{tj}\rangle$$ パターン内相関

パターン間相関

$$+ J[G(\varphi_{t-1})(2\mathbf{v}_{t-1}-1) + G(\varphi_t)^{\mathrm{T}}(2\mathbf{v}_{t+1}-1)]_i$$

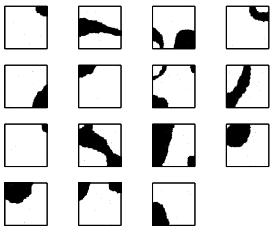
$$+\frac{1}{4}\log\frac{\beta_{\mathrm{H}}}{\beta_{\mathrm{L}}} - \frac{1}{4}(\beta_{\mathrm{H}}-\beta_{\mathrm{L}})\left(([\mathbf{y}_t - W_t\boldsymbol{\mu}]_i)^2 + \mathbf{w}_{ti}^{\mathrm{T}}\Sigma\mathbf{w}_{ti}\right)$$

ピクセルごとのノイズ分散

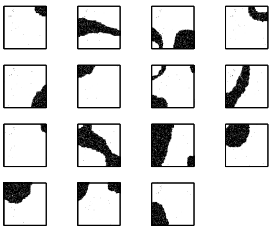
# 実験 1

- 128x128観測画像が15枚
- 4倍に拡大した512x512画像を推定
- 位置パラメタは既知
- ノイズパタンが観測画像ごとに異なる

真のノイズパタン **z**



推定された $q(\mathbf{z})$



# 実験 1: 結果

True
PSNR Inf

Obs. Mean
PSNR 21.57

Ours
PSNR 32.45

Uniform Low Prec.
PSNR 27.80

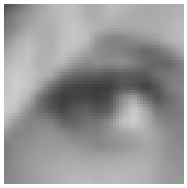
Uniform Opt. Prec.
PSNR 29.51

Uniform High Prec.
PSNR 20.41

# 実験 1: 結果(右目拡大)

True
PSNR Inf

Obs. Mean
PSNR 20.85

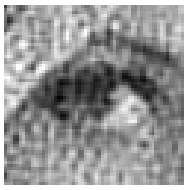


Ours
PSNR 31.38

Uniform Low Prec.
PSNR 25.57

Uniform Opt. Prec.
PSNR 28.06

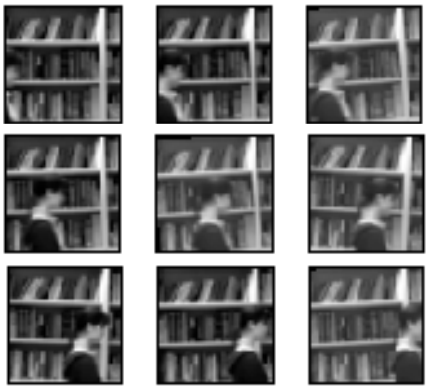
Uniform High Prec.
PSNR 21.26



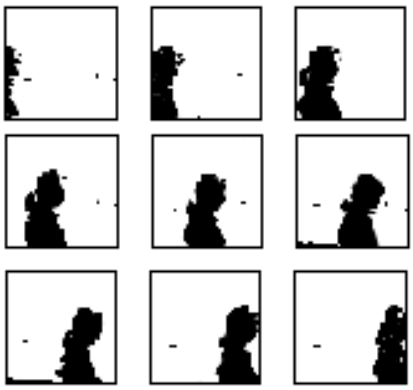
画素ごとに
正則化の強さが
変わることで
観測情報を
効率的に利用

# 実験 3

- 実画像2: x4超解像
  - どのフレームにも遮蔽物が存在



観測画像列



推定されたパタン $\hat{\mathbf{z}}$

# 実験 3: 結果

観測画像（1/9枚目）

観測の平均

観測メディアン

遮蔽を仮定しない推定

階層ベイズ推定

# このパートのまとめ

- 階層ベイズモデリングの技法により、構造尤度を導入したモデル化を行った
- 困難な積分を変分学習により近似的に実施
- 遮蔽がある状況においても、遮蔽物を除去しながらの超解像が可能
- 遮蔽物のダイナミクスを考えることでさらに精度を上げることができる

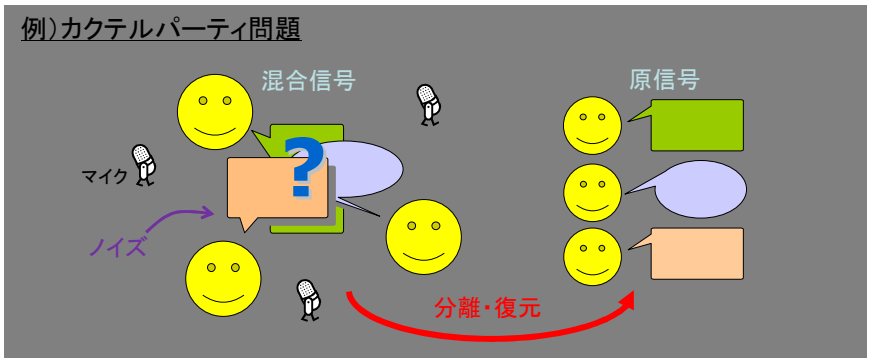
# 明滅する信号源からの
# ブラインド信号分離

Hirayama, J., Maeda, S., Ishii, S. *IEEE Transactions on Neural Networks*, **18**(5), 1326-1342, 2007.

平山淳一郎: Helsinki University

## ブラインド信号分離 (BSS)

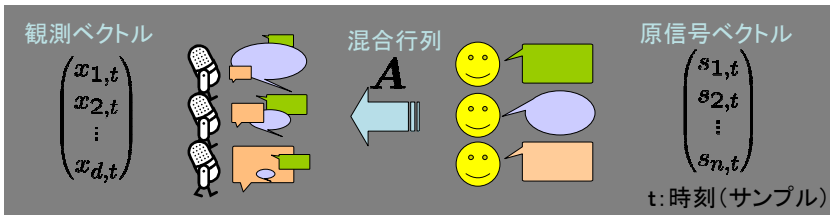
- 混合信号(＋ノイズ)から原信号を分離・復元したい
  - ただし，混合過程，原信号ともに未知の状況を想定




原信号間の独立性を手がかりにした解法: 独立成分分析（ICA）

# パラメトリックICA（ノイズあり）

- 混合過程：正規線形モデル




$$\boldsymbol{x}_t = \boldsymbol{A}\boldsymbol{s}_t + \boldsymbol{\epsilon}_t$$ 正規ノイズ

- 原信号の生成モデル：成分ごとに独立な非正規分布

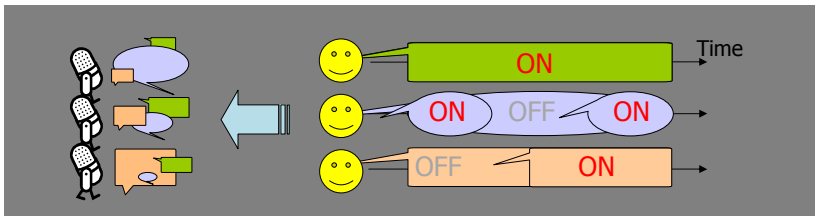
$$p(\boldsymbol{s}_t) = \prod_{i=1}^{n} p_s(s_{i,t})$$

- ラプラス分布
- t分布
- 混合正規分布，etc.

観測 $\mathbf{x}_1$, $\mathbf{x}_2$, …に基づく未知変量 $\mathbf{A}$, $\mathbf{s}_1$, $\mathbf{s}_2$, …の推定問題に帰着

# 信号源のONとOFF

- 通常，原信号の数は「既知」かつ「固定」
- しかし，非定常環境ではしばしば時間的に「ON・OFF」が切り替わる




- 一般性のある状況だが，陽に考慮した手法はない
- 出現信号の切り替えを取り入れることで，
  - 高ノイズ下での復元性能の向上
  - 出現パターンの抽出による知識発見

が見込まれる

# 提案手法：Switching ICA

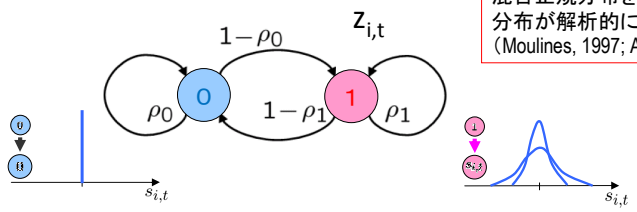
- 原信号のモデルにON/OFFの切り替え構造を導入
  - $z_{i,t} \in \{0, 1\}$ ： ON/OFFを表す「隠れ変数」
  - $z_{i,t}$ は2状態のマルコフ連鎖に従う

$$p(s_{i,t} \mid z_{i,t}) = (1 - z_{i,t})\,\delta(s_{i,t}) + z_{i,t}\,p_s(s_{i,t})$$

ディラックのデルタ関数
（OFF信号のモデル）

ON信号のモデル

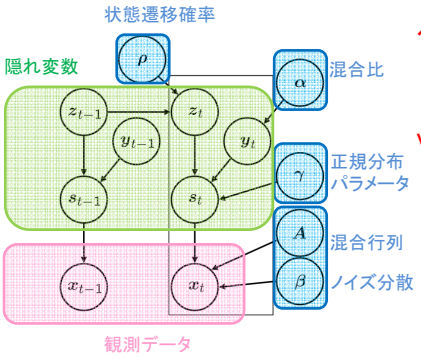
混合正規分布を仮定：事後分布が解析的に計算可能
（Moulines, 1997; Attias, 1999）




「デルタ分布」と「混合正規分布」の隠れマルコフモデル(HMM)

# 学習アルゴリズム

- 変分ベイズ（VB）法（Attias, 1999）に基づき学習則を導出
  - 平均場近似に基づく決定論的な反復推定アルゴリズム




$y_t$: 混合正規分布（ON信号）における混合要素インデックス（隠れ変数）

ベイズ事後分布（計算困難）

P(隠れ変数, パラメータ | 観測データ)

≈ KL[ Q || P ] 最小化

VB近似事後分布

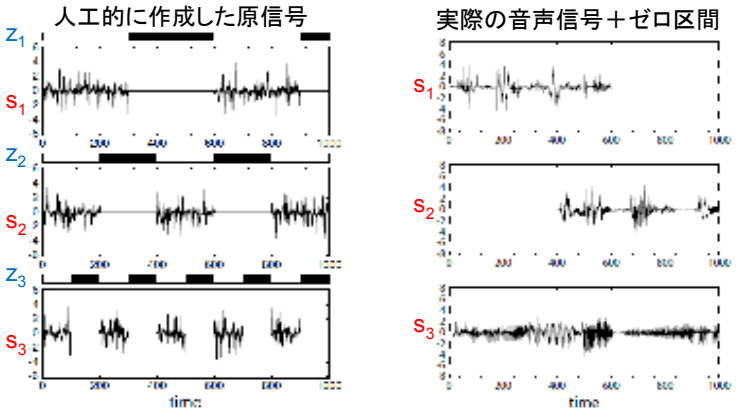
Q(隠れ変数 | 観測データ)

×Q（パラメータ | 観測データ）

条件付独立性より，各 ごとに分解される

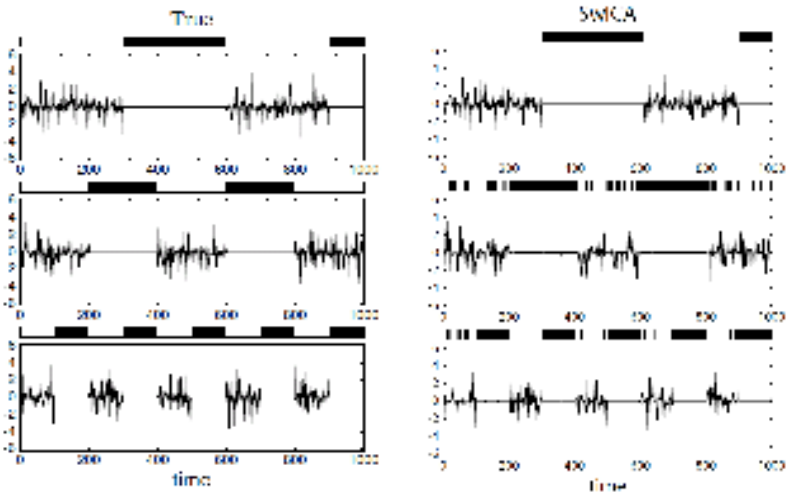
各 Q( ), Q( ) ごとに交互にKL最小化
（解析解が得られる）することで解が得られる

# 計算機実験




- $x_t = As_t + \epsilon_t$ により混合（混合数3）
- 混合行列Aは試行ごとにランダムに設定

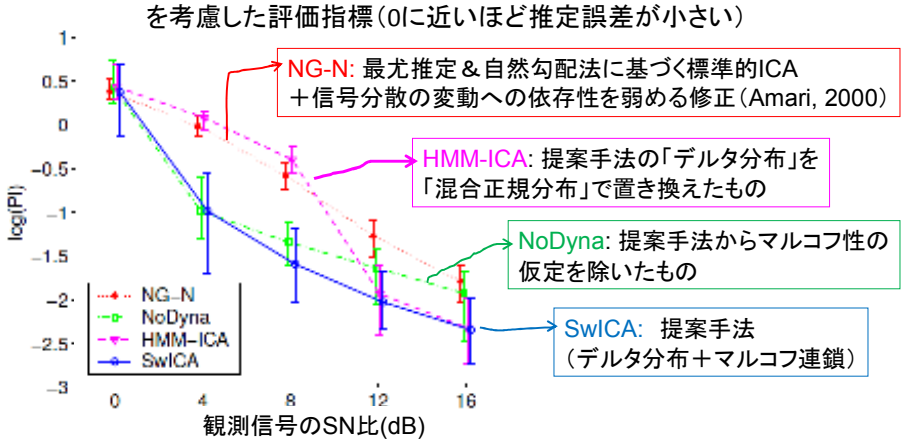
# 原信号の復元例（人工信号）




原信号を良く再現；ON/OFFパターンを比較的良く再現

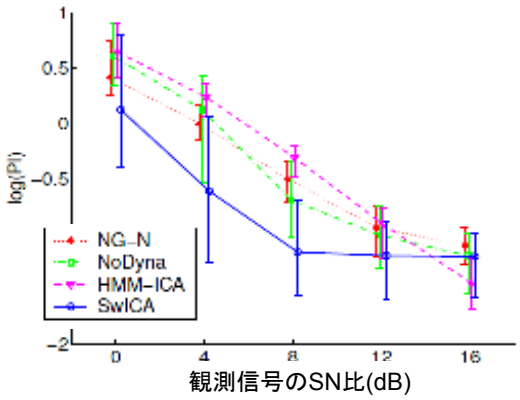
# 混合行列の推定結果（人工信号）

- Amari's Performance Index (Amari,1996)による評価
  - 混合行列の推定における列の入れ替えと符号に関する不定性を考慮した評価指標（0に近いほど推定誤差が小さい）

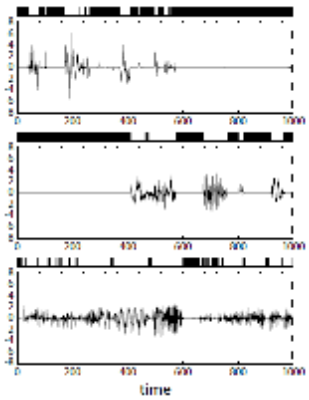



「デルタ分布＋マルコフ連鎖」によるON/OFFのモデル化が有効

# 混合行列の推定結果（音声信号）




推定結果の例




既存手法と比較して，高ノイズ下での推定精度を改善可能

## このパートのまとめ

- ブラインド信号分離（BSS）と独立成分分析（ICA）
  - BSS:  混合過程が未知，原信号が未知，両方を復元
  - ICA:  信号間の独立性を手がかりにした統計的推定手法
- 非定常環境で原信号の数や種類が切り替わる問題
- パラメトリックICAに原信号のON・OFFを導入
  - デルタ分布(OFF)と混合正規分布(ON)によるHMM原信号
- セミマルコフモデルへの拡張
  - 「デルタ分布＋セミマルコフ」は信号復元＆ON／OFFパターン推定の双方に有効

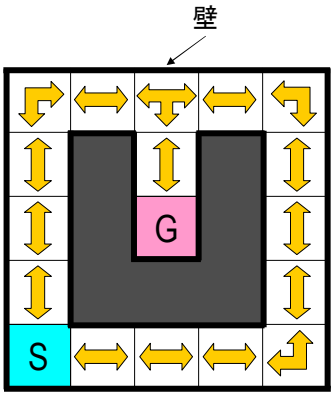
## ベイズフィルタによる<br>脳における信念形成機構の解明

Yoshida, W., Ishii, S. *Neuron*, **50**(5), 781-789, 2006.

吉田和子：University College London

# 部分観測状況での意思決定

- 部分観測迷路
  - 状態：迷路の絶対座標
    - 観測できない
  - 観測：四方の壁の有無
    - 同じ観測でも異なる状態が存在
  - 行動：前後左右への進行

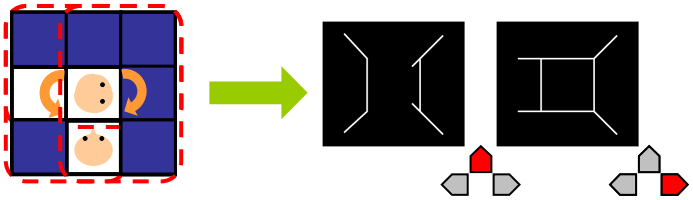



同じ行動でも最適行動が違うので
性能が劣化してしまう

# 部分観測迷路課題

- グリッド迷路内を3次元の観測に基づき探索する

2次元迷路上の位置　　3次元のwire-frame scene



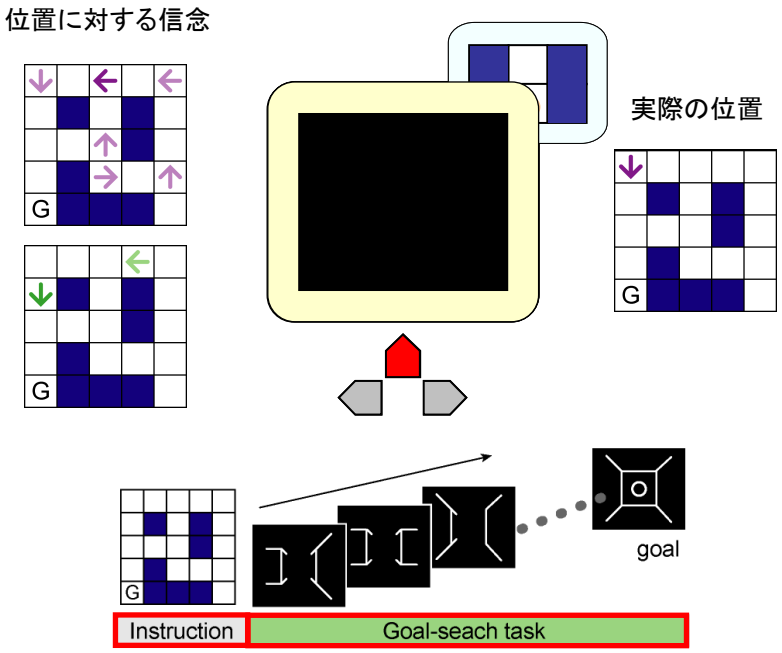
  - 観測できるのは、現在位置周囲の6グリッドのみ
  - 各試行（2秒）で前進・左右回転の3行動から1つを選択
- 迷路の構造（地図）、2次元⇔3次元間の変換（➡）は実験前日の訓練により学習済み

# Goal-search (GS) 課題

- 未知のスタートから既知のゴールへの探索課題
  - 教示では迷路上のゴール位置(‘G’)が提示される
  - 迷路上の現在位置は隠れ状態であり観測不可能

  - ゴールに到達すると課題終了

効率的なゴール到達には、記憶した地図と観測履歴に基づき、隠れ状態を推定することが重要



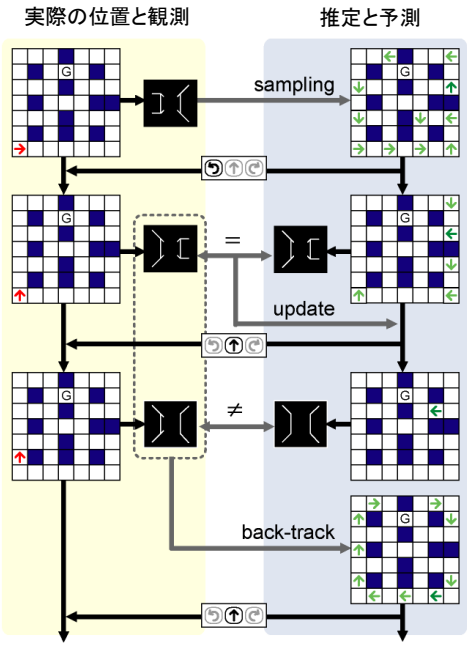
# 機能的核磁気共鳴画像

- 血中の脱酸化ヘモグロビン量（BOLD効果）を調べる
- 間接的に神経細胞集団の活動量を調べる
- 被験者13名
  （男性11名、女性2名）

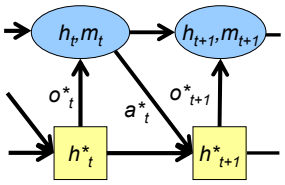
島津マルコーニ社製
1.5 tesla fMRI装置

# 隠れ状態推定過程

(1)候補のサンプリング
　観測に合致する位置の候補をサンプリングし、推定位置を決める
(2)推定位置の遷移
　行動を選択し、次時刻での位置と観測を予測する
(3)観測の照合
　実際と予測した観測を照合する
　同じ場合：*Updating (UD)*
　　予測位置を次時刻の推定位置とする
　異なる場合：*Backtracking (BT)*
　　後ろ向き推論により、候補位置を再構成
(4)t=t+1とし、(2)から繰り返す



# 確率モデル化と推定




- 被験者のモデル
  - 隠れ状態を逐次ベイズ推定
  - 行動は準最適(ε 貪欲)
- 被験者の内的な推定状態($h$)とオペラント($m$)は実験者にとって観測できない
  - 隠れマルコフモデルとして定式化
  - 隠れ変数は、被験者行動($a^*$)に基づきベイズフィルタで推定
  - パラメータは最尤推定
- 隠れ状態の推定負荷を2種の統計量で相関解析
  - 信念(隠れ状態の事後分布)エントロピー
  - Backtracking確率

# HMMに基づくベイズ推定

- 仮定したHMMに対するベイズ推定
  - 被験者の認知状態を行動(観測)データから推定する
    $h_{1:T}$ と $m_{1:T}$ $\quad$ $o^*_{1:T}$ と $a^*_{1:T-1}$
  - 逐次ベイズ推定

$$P(h_{1:t+1}, m_{1:t+1} | a^*_{1:t}) = \frac{\boxed{P(h_{t+1}, m_{t+1}, a^*_t | h_{1:t}, m_{1:t}, a^*_{1:t-1})} P(h_{1:t}, m_{1:t} | a^*_{1:t-1})}{P(a^*_t | a^*_{1:t-1})}$$

    - 観測シーン $o^*_{1:T}$ は環境のダイナミクスを構成する
  - 尤度

$$P(h_{t+1}, m_{t+1}, a^*_t | h_{1:t}, m_{1:t}, a^*_{1:t-1}) = P(a^*_t | h_t) P(m_{t+1} | h_t, m_{1:t}, a^*_{1:t}) P(h_{t+1} | m_{1:t+1}, a^*_{1:t})$$

- 尤度の計算（再掲）

$$P(a_t^*|h_t)\,P(m_{t+1}|h_t, m_{1:t}, a_{1:t}^*)\,P(h_{t+1}|m_{1:t+1}, a_{1:t}^*)$$

  - 行動選択
    - 最適行動選択確率　$\alpha(<1)$
  - 次のオペラントの決定
    - $\hat{o}_{t+1} = o_{t+1}^*$ であれば、*updating* を選択（ $m_{t+1} = 0$ ）
    - 観測が一致しなければ、*back-tracking* を選択（ $m_{t+1} = 1$ ）
  - 次の推定位置の評価
    - Updating

$$P(h_{t+1}|m_{t+1} = 0, m_{1:t}, a_{1:t}^*)$$
$$\propto \Sigma_{h_t} P(h_{t+1}|h_t, a_t^*) P(m_{t+1} = 0|h_t, a_t^*) P(a_t^*|h_t) P(h_t|m_{1:t}, a_{1:t-1}^*)$$

    - Back-tracking

$$P(h_{t+1}|m_{t+1} = 1, m_{1:t}, a_{1:t}^*) = \begin{cases} 1/N_{t+1} & \text{if} \quad h_{t+1} \in H_{t+1} \\ 0 & \text{otherwise,} \end{cases}$$

# モデルから推定される認知負荷

- 隠れ状態推定に関わる主要な情報処理
  - Hidden current position (HCP) entropy
    - 推定位置の操作・保持に関わる計算負荷
    - $h_t$ の確率分布についての情報量

$$\mathcal{J}_t = -\Sigma_{m_{1:t}} P(m_{1:t}|a_{1:T-1}^*) \Sigma_{h_t} P(h_t|m_{1:t}, a_{1:t-1}^*) \log P(h_t|m_{1:t}, a_{1:t-1}^*)$$

  - Back-tracking (BT) probability
    - 位置の再推定時にかかる計算負荷

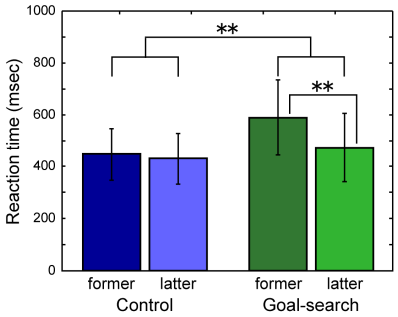
$$\mathcal{M}_t = P(m_t = 0|a_{1:T-1}^*)$$

被験者の認知状態に対する依存性を除去するため、
これらの指標は隠れ状態に対して周辺化して求める

平滑化分布

# 行動データ解析 (1/4)

- GS,VMの各課題を前後半に分けてRTを比較

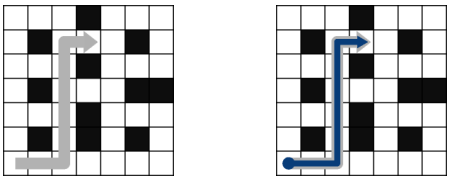



*t* 検定
- GS>VM: $p<0.0001$
- 前半>後半:
  GS: $p<0.001$
  VM: n.s.

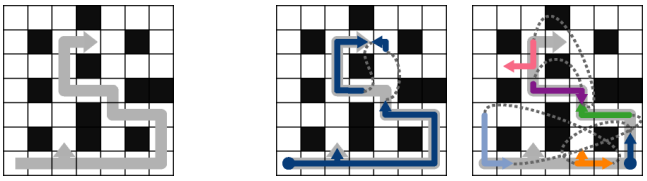
– RTは隠れ状態推定に関わる認知負荷を反映
– ランダム行動により偶然ゴールしているのではない

# 行動データ解析 (2/4)

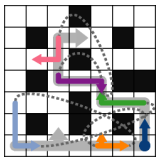
- 被験者が実際にとった経路と再構成経路
  – 成績の良い被験者



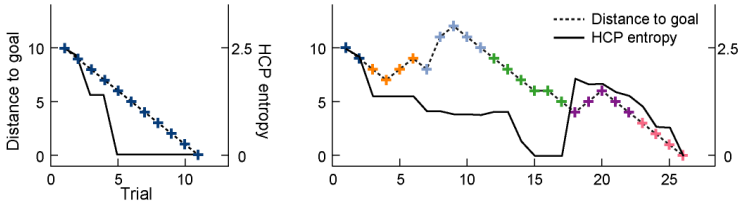
  – 成績の悪い被験者



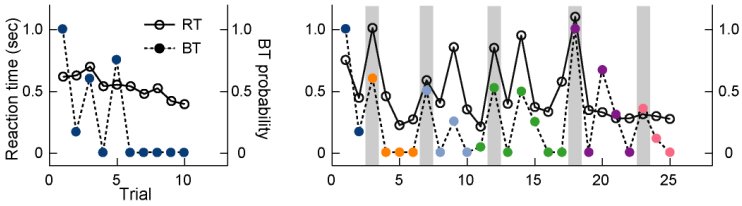
# 行動データ解析 (3/4)



- ゴールまでの最短ステップ数とHCP entropy



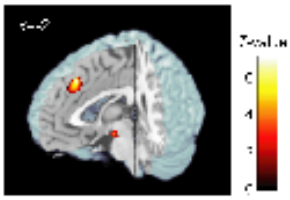

- 反応時間(RT)とBT probability




# 行動データ解析 (4/4)

- モデルによって被験者行動がよく説明できる
  - 被験者行動の再現率: 88.7±2.9%
    - 時刻*t*までの観測・行動履歴から、時刻*t*+1での行動を予測
  - RT成分との一致
    - 高周波成分とBT probability
      - BTとRTピークの一致 (59.1%, 前後試行含86.1%)
    - 低周波成分とHCP entropy
      - 10名の被験者で、RTとentropyが有意に相関($p$<0.05)
      - ゴールまでの距離との相関より有意に高い
  - パラメータ推定結果(最尤推定)
    - 最適行動選択確率: $\alpha$=0.8 (0.81±0.08)
    - BT再推定時の試行数: $n$=1 (1.3±0.6)

# 核磁気共鳴図データ解析



- Backtracking確率と相関
  - 内側前頭前野 (medial PFC)
  - 隠れ状態のサンプリングに関わることを示唆

- 信念エントロピーと相関
  - 前部前頭前野 (anterior PFC: APF)
  - 隠れ状態の保持と更新に関わることを示唆

# このパートのまとめ

- 不確実な環境における脳の意思決定では、不確実さの解消のための推論が必要
- 推論や意思決定のような脳の高次情報処理のモデル化には確率モデルが有効
- 隠れマルコフモデルは人間の行動を良く再現した
- 再現性の高いモデルに基づき、人間の推論に関わる脳の情報処理基盤を論じることができる
  - ヒトの推論はベイズ推定を使っている …かも知れない

E-mail: ishii@i.kyoto-u.ac.jp
URL: http://ishiilab.jp/