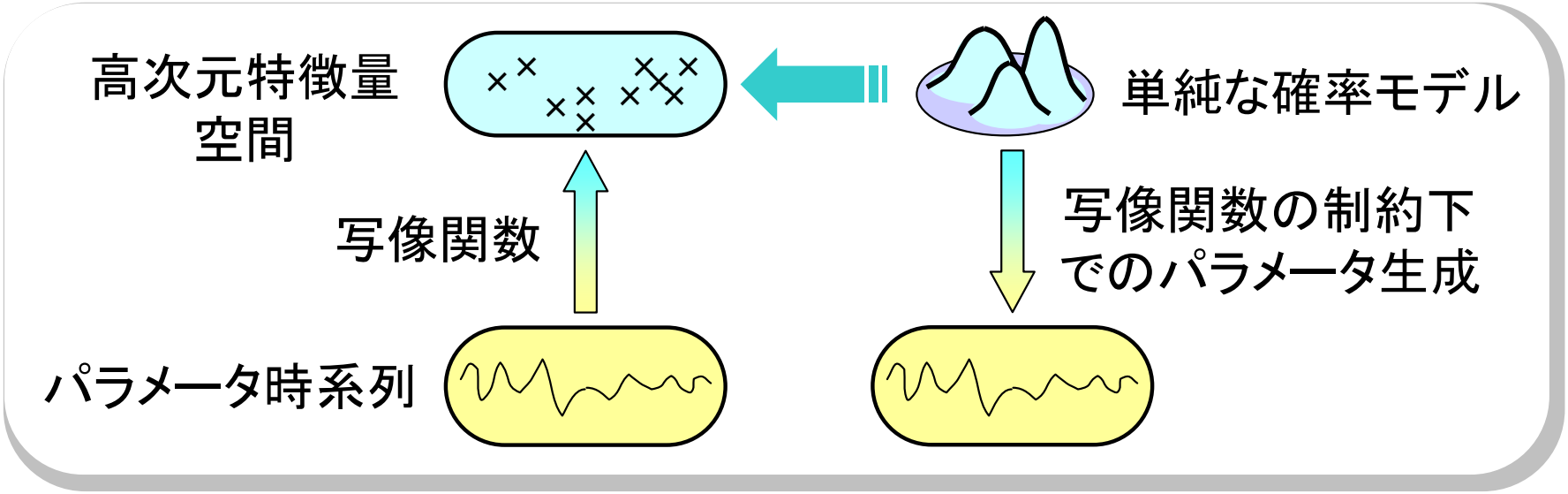
音声音響信号処理
～統計的手法による音声変換～



戸田 智基

奈良先端科学技術大学院大学

情報科学研究科

2012年1月23日(月)

内容

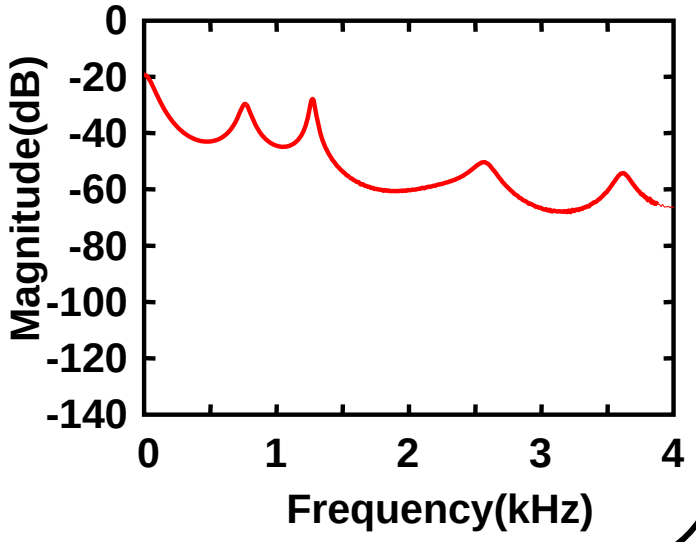
1. 音声変換のしくみ
2. 統計的手法による声質変換
   2.1. 基本的な枠組み
   2.2. フレームベース変換法
   2.3. 系列ベース変換法
3. 応用例

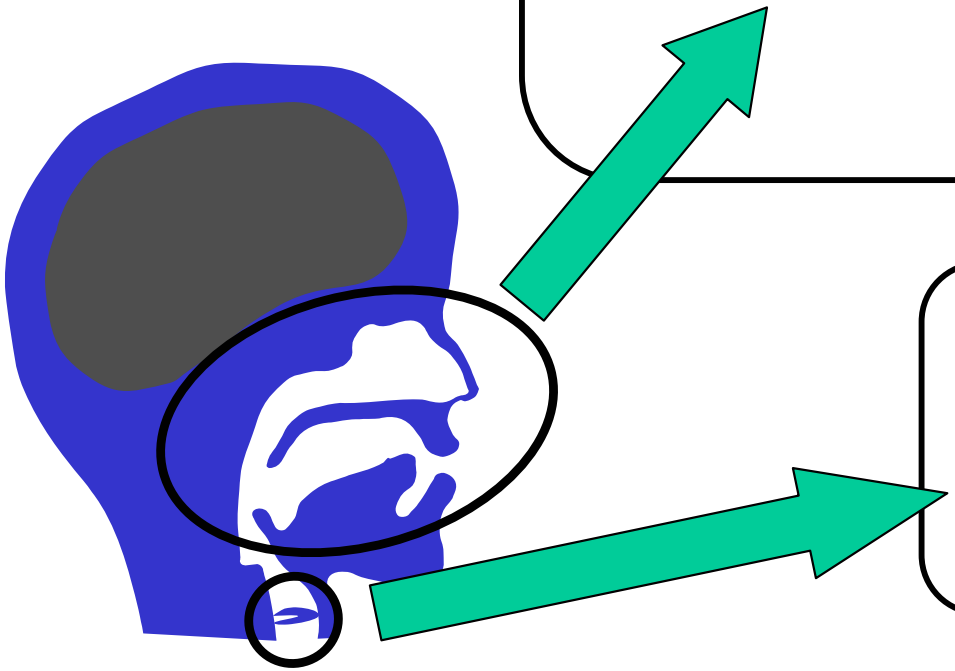
音声の特徴量

声道特徴量

音韻性や声質を
調節する.

- スペクトル包絡





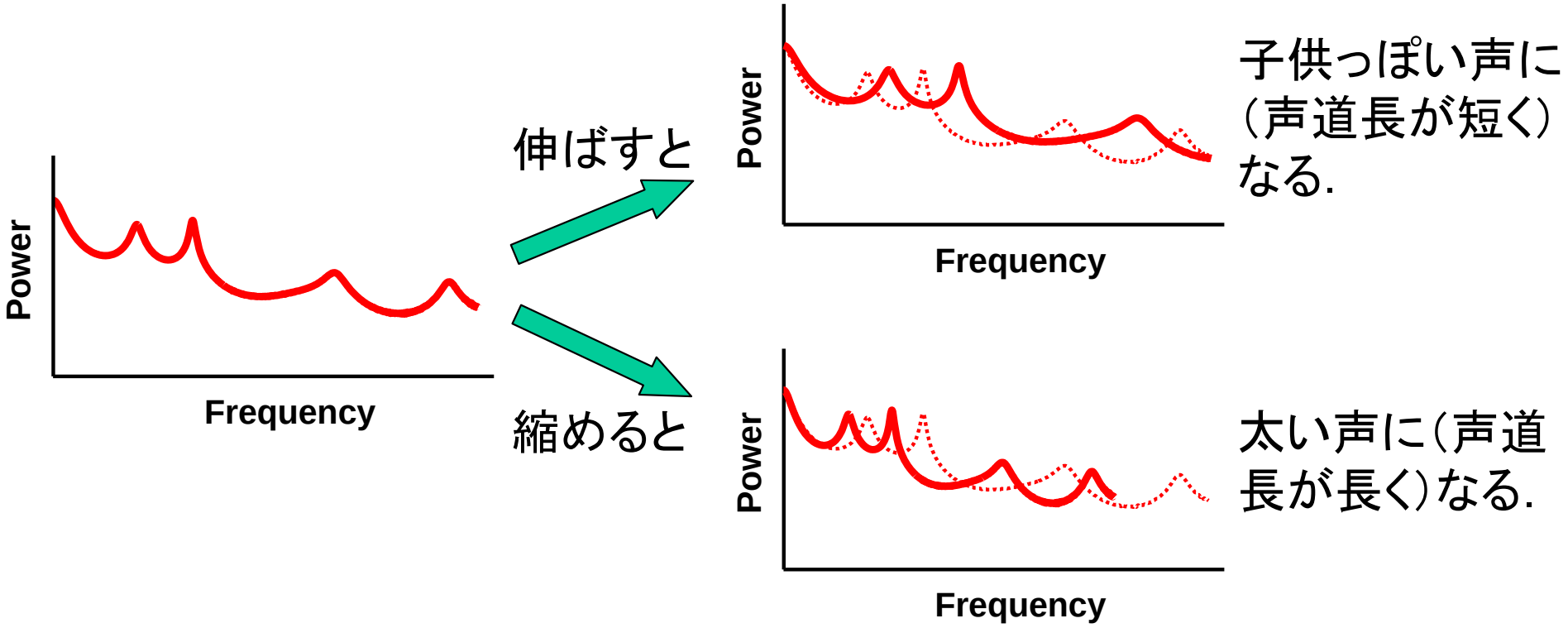
音源特徴量

韻律(イントネーション)を
調節する.

- 基本周波数(F_0)

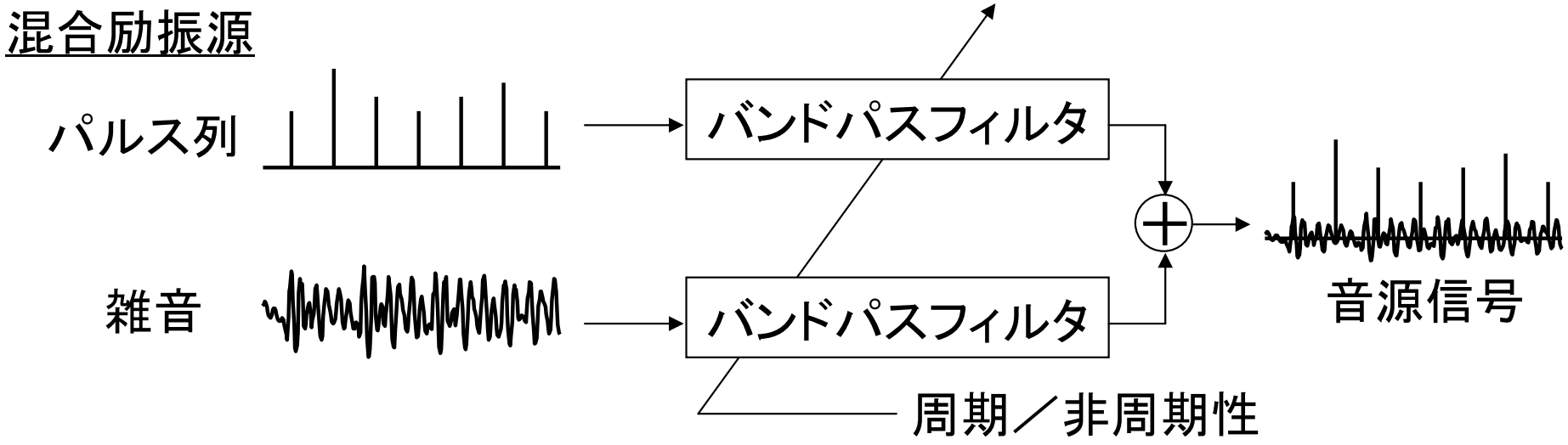
声道特徴量の変換

- 特徴量パラメータとしてスペクトル包絡が用いられる.
- 音韻性や声質などを表す.
- 周波数軸方向に一律に伸縮させることで，音韻性を保ったまま声質を変換することができる.



音源特徴量の変換

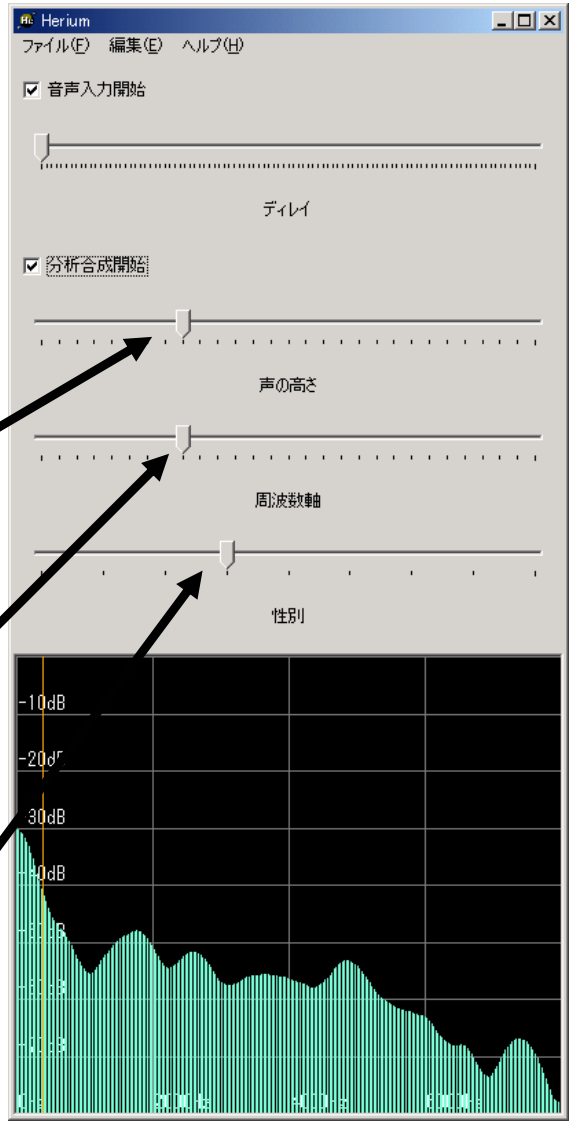
- 特徴量パラメータとして基本周波数(F_0)や周期／非周期性(混合励振源使用時)が用いられる.
- 声の高さや声のかすれなどを表す.
- 基本周波数を高くすれば高い声に，低くすれば低い声に変換することができる.
- 混合励振源使用時には，非周期性を大きくすればかすれた声に変換できる.



リアルタイム音声変換デモ

リアルタイム声質変換ソフト
Herium: High Entertaining Real-time Input Utterance Modifier

作成者： 名城大学　坂野秀樹先生



- 声の高さの調節
  - F_0の変換
- 声質の調節
  - スペクトルの変換（周波数軸の伸縮）
- 性別の調節
  - F_0とスペクトルの変換

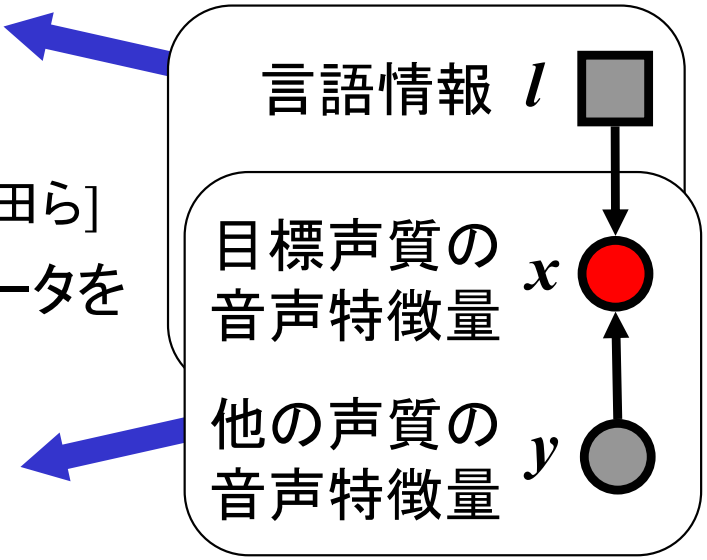
http://www.sp.m.is.nagoya-u.ac.jp/people/banno/spLibs/herium/index-j.html

内容

1. 音声変換のしくみ
2. 統計的手法による声質変換
   2.1. 基本的な枠組み
   2.2. フレームベース変換法
   2.3. 系列ベース変換法
3. 応用例

統計的手法に基づく声質モデリング

- 確率モデルによる音声特徴量のモデル化
  - テキスト音声合成（TTS）
    - 確率密度関数（$p.d.f.$）$P(\boldsymbol{x}\,|\,\boldsymbol{l})$ のモデル化
    - 隠れマルコフモデル（HMM）による手法 [徳田ら]
    - 言語情報が付与された目標声質の音声データを用いて学習
  - 声質変換
    - $P(\boldsymbol{x}\,|\,\boldsymbol{y})$ のモデル化
    - 混合正規分布モデル（GMM）による手法 [Stylianou *et al.*]
    - 言語情報は同一で，所望の声質成分のみが異なる音声データ（パラレルデータ）を用いて学習

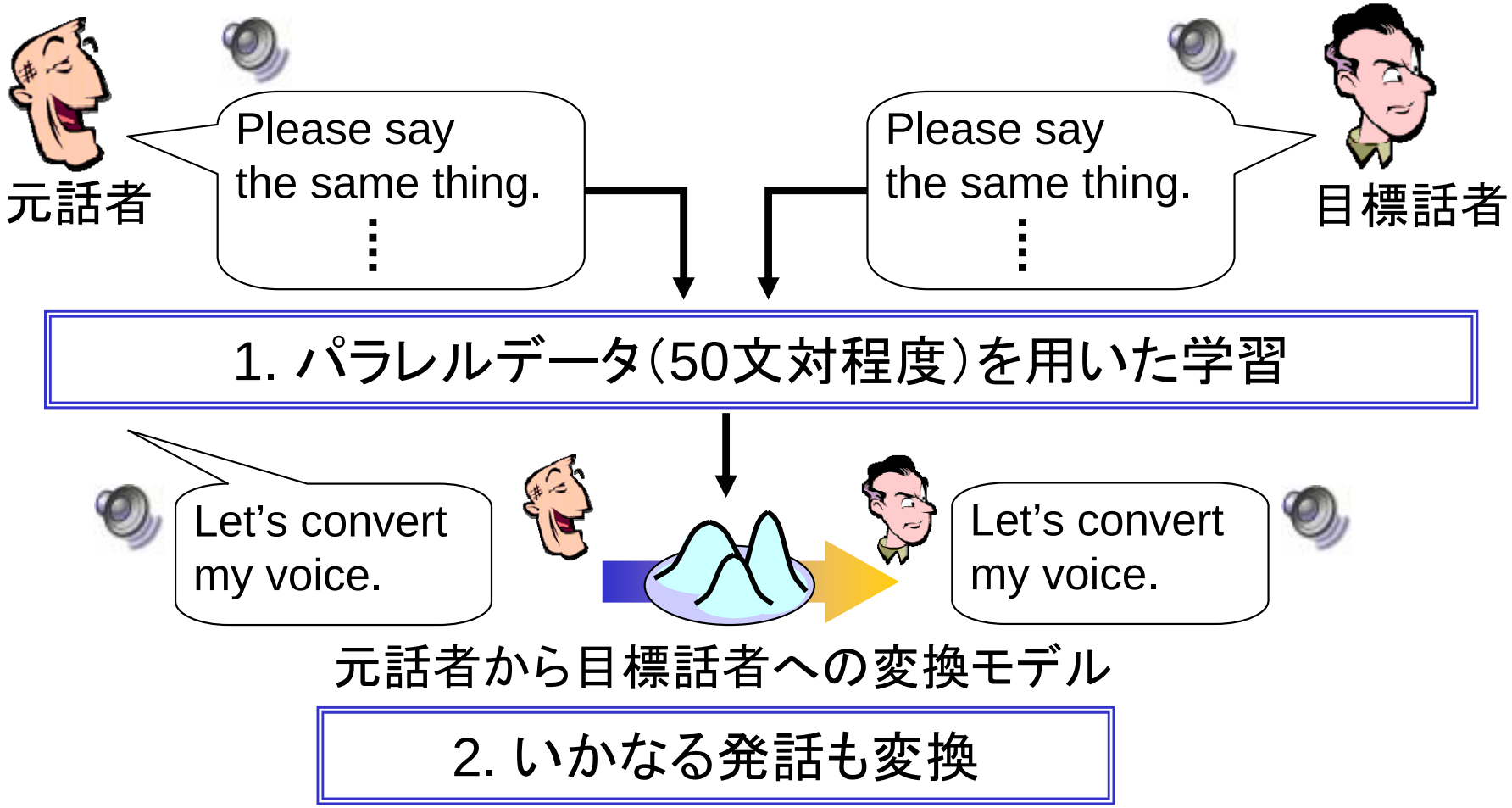


- モデル化される声質成分は**学習データ**により決定

統計的手法に基づく声質変換

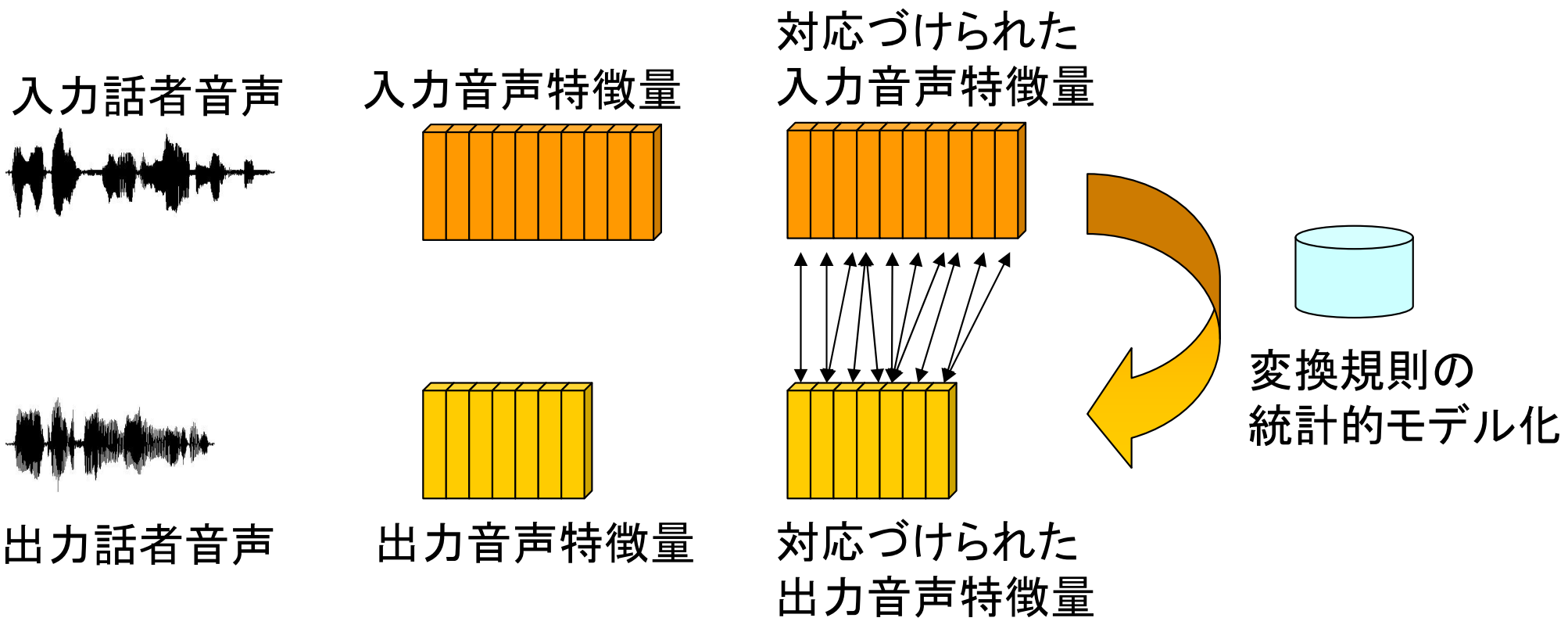
[Abe *et al.*]

- 異なる話者や発話様式間における変換を実現できる.
- 入・出力話者による同一内容発声の音声データ(パラレルデータ)を用いて入・出力音声特徴量間の対応関係を統計的にモデル化する.



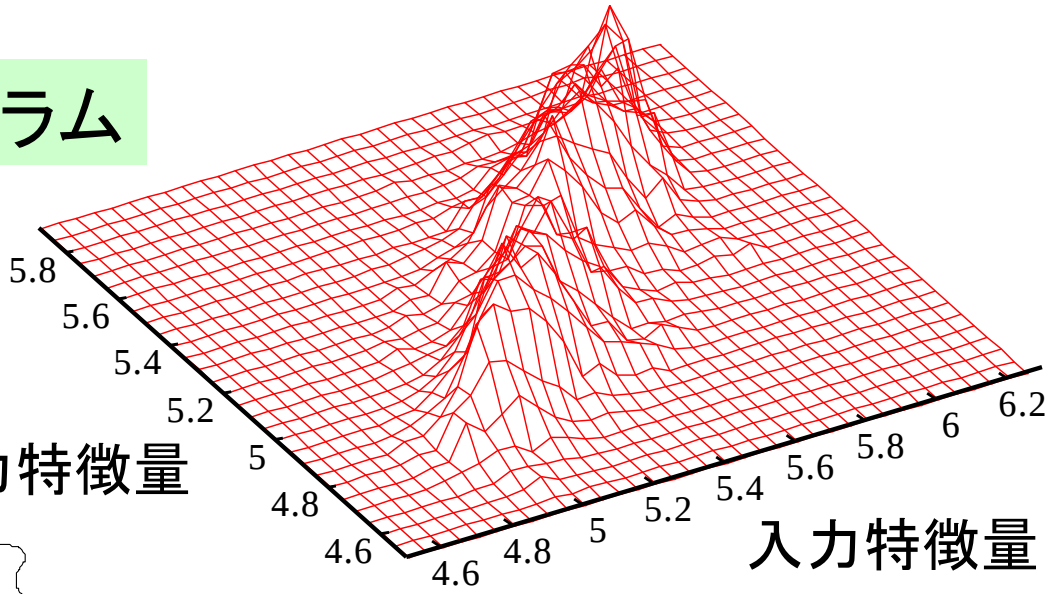
学習手順

音声波形 → 音声特徴量 → 対応付けられた音声特徴量 → 変換規則

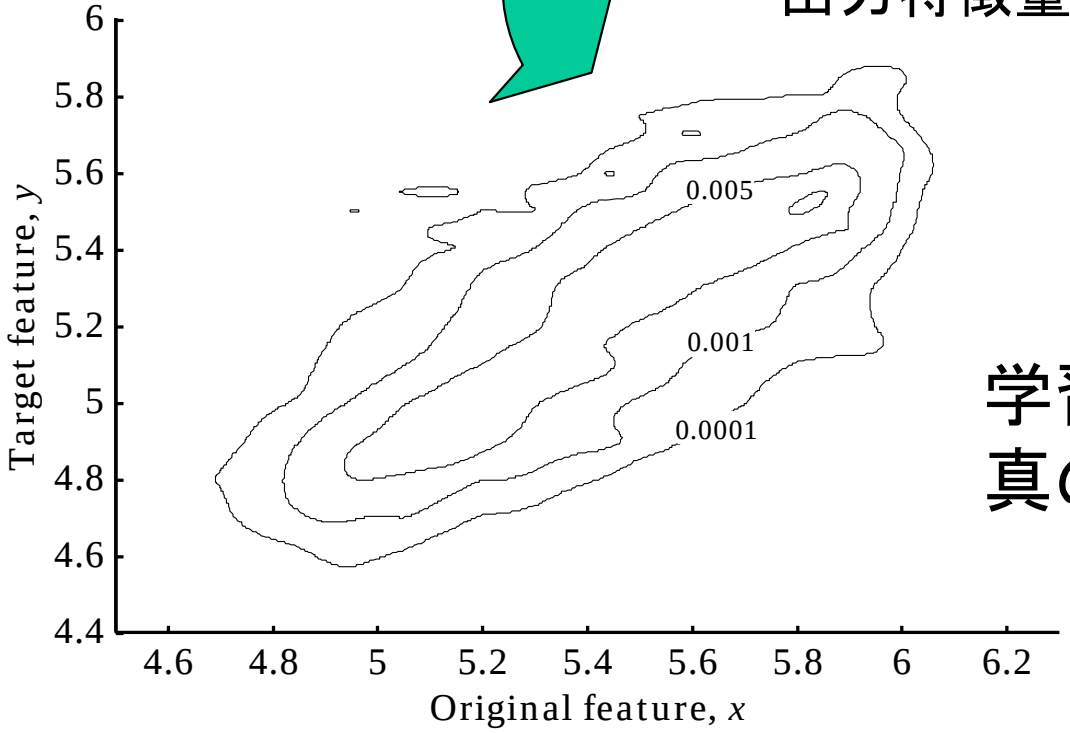


学習データの一例

学習データのヒストグラム



上から見ると



学習データ量が限られているため
真の分布は分からない.

内容

1. 音声変換のしくみ
2. 統計的手法による声質変換
   2.1. 基本的な枠組み
   2.2. フレームベース変換法
   2.3. 系列ベース変換法
3. 応用例

フレームベースの変換関数

- 各フレーム（各時間）において独立に変換処理を行う.
  - 入力特徴量ベクトル: $\boldsymbol{x}_t$
  - 出力特徴量ベクトル: $\boldsymbol{y}_t$
  - 統計モデル: $\boldsymbol{\lambda}$
  - 変換特徴量ベクトル: $\hat{\boldsymbol{y}}_t$

$$\hat{\boldsymbol{y}}_t = F_{\lambda}(\boldsymbol{x}_t)$$

例えば

$$= E_{\lambda}[\boldsymbol{y}_t \mid \boldsymbol{x}_t] = \int \boldsymbol{y}_t P(\boldsymbol{y}_t \mid \boldsymbol{x}_t, \boldsymbol{\lambda}) d\boldsymbol{y}_t$$

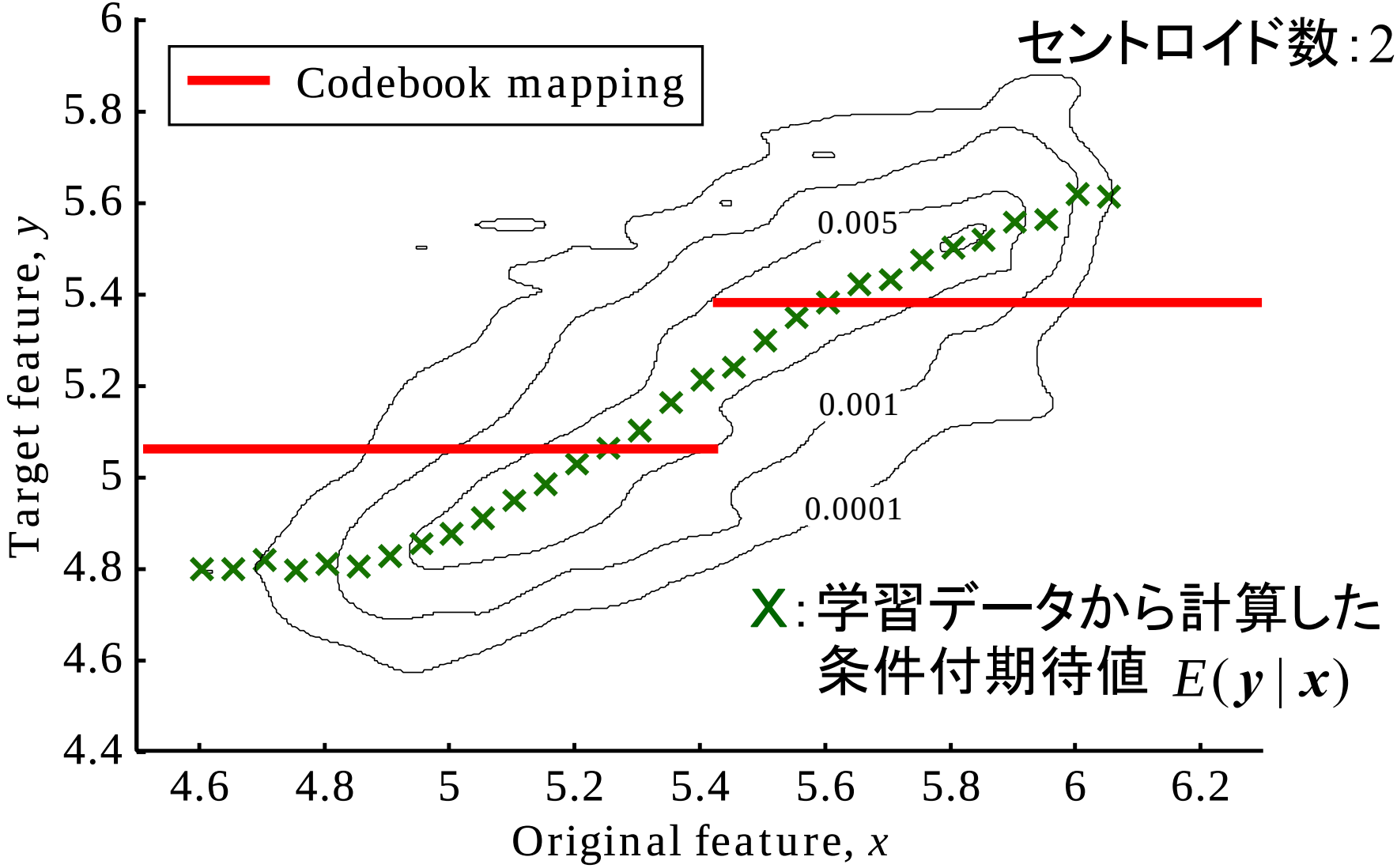
とか

$$= \arg\max_{y_t} P(\boldsymbol{y}_t \mid \boldsymbol{x}_t, \boldsymbol{\lambda})$$

代表的な変換法

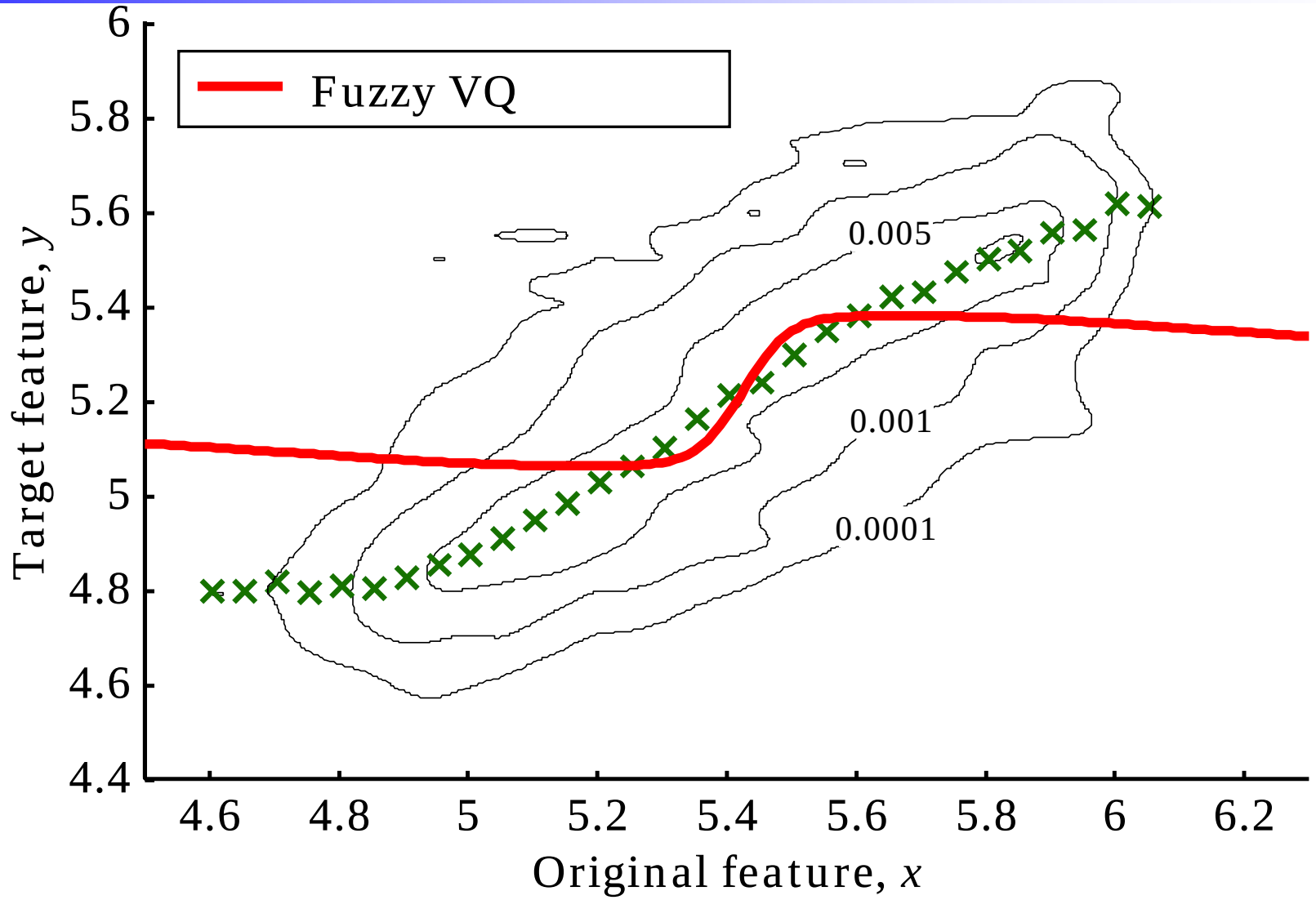
1. コードブックマッピング法 [Abe *et al.*]	$\hat{\boldsymbol{y}}_t = \boldsymbol{\mu}_m^{(y)}$ ハードクラスタリングと離散的なマッピング
2. ファジーベクトル量子化に基づくコードブックマッピング法 [中村 他]	$\hat{\boldsymbol{y}}_t = \sum_{m=1}^{M} \gamma_{m,t}^{(x)} \boldsymbol{\mu}_m^{(y)}$ ソフトクラスタリングと離散的なマッピング
3. ファジーベクトル量子化と差分ベクトルに基づくコードブックマッピング法 [Matsumoto *et al.*]	$\hat{\boldsymbol{y}}_t = \boldsymbol{x}_t + \sum_{m=1}^{M} \gamma_{m,t}^{(x)} \left(\boldsymbol{\mu}_m^{(y)} - \boldsymbol{\mu}_m^{(x)} \right)$ ソフトクラスタリングと連続的なマッピング
4. 線形回帰法 [Valbret *et al.*]	$\hat{\boldsymbol{y}}_t = \boldsymbol{A}_m \boldsymbol{x}_t + \boldsymbol{b}_m$ ハードクラスタリングと連続的かつ高精度な変換
5. 混合正規分布モデルに基づく変換法 [Stylianou *et al.*]	$\hat{\boldsymbol{y}}_t = \sum_{m=1}^{M} \gamma_{m,t}^{(x)} \left(\boldsymbol{A}_m \boldsymbol{x}_t + \boldsymbol{b}_m \right)$ ソフトクラスタリングと連続的かつ高精度な変換

コードブックマッピング法の変換関数



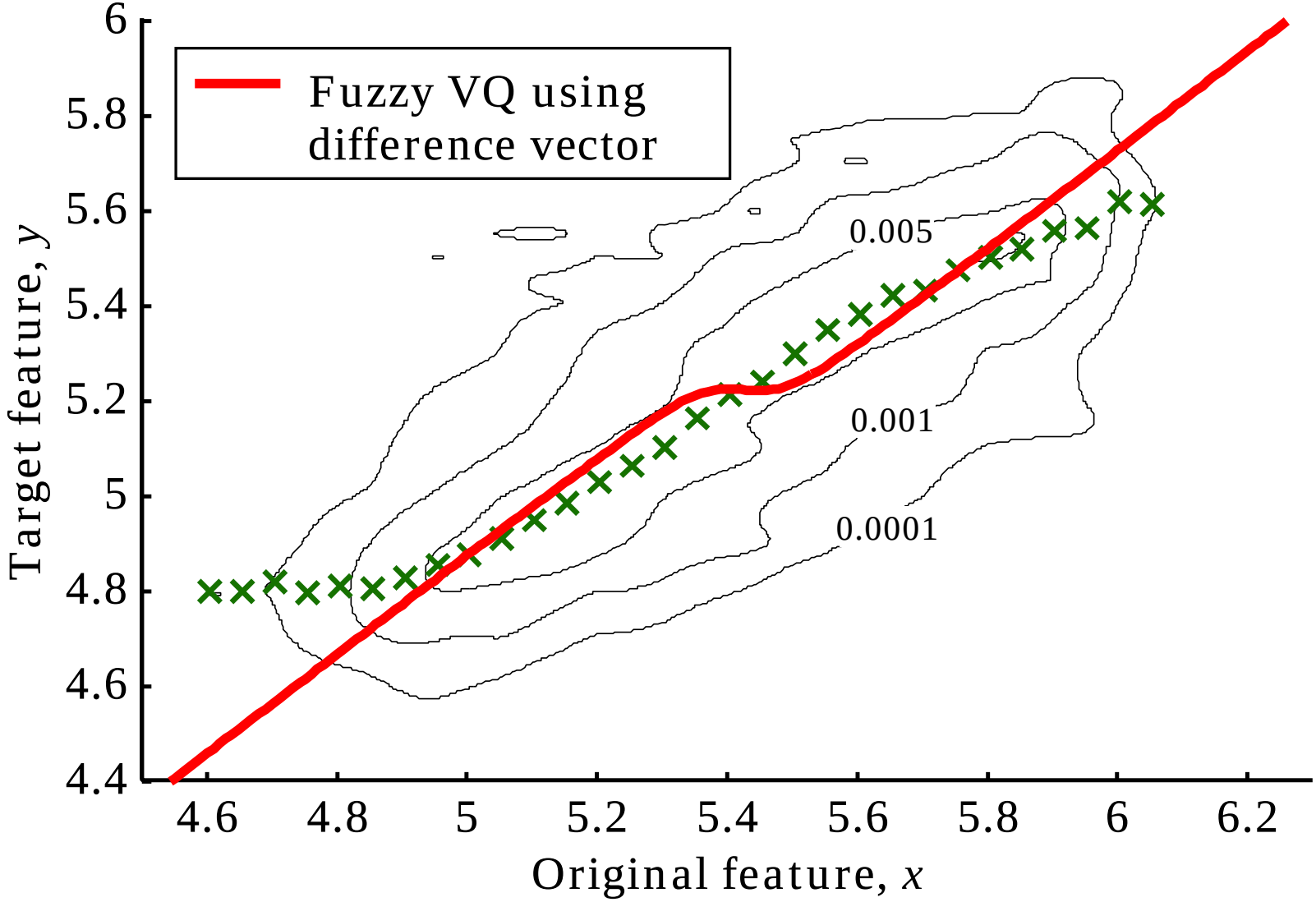
$$\hat{y}_t = \mu_m^{(y)}$$ ハードクラスタリングと離散的なマッピング

ファジーVQ使用時の変換関数



$$\hat{\boldsymbol{y}}_t = \sum_{m=1}^{M} \gamma_{m,t}^{(x)} \boldsymbol{\mu}_m^{(y)}$$ ソフトクラスタリングと離散的なマッピング

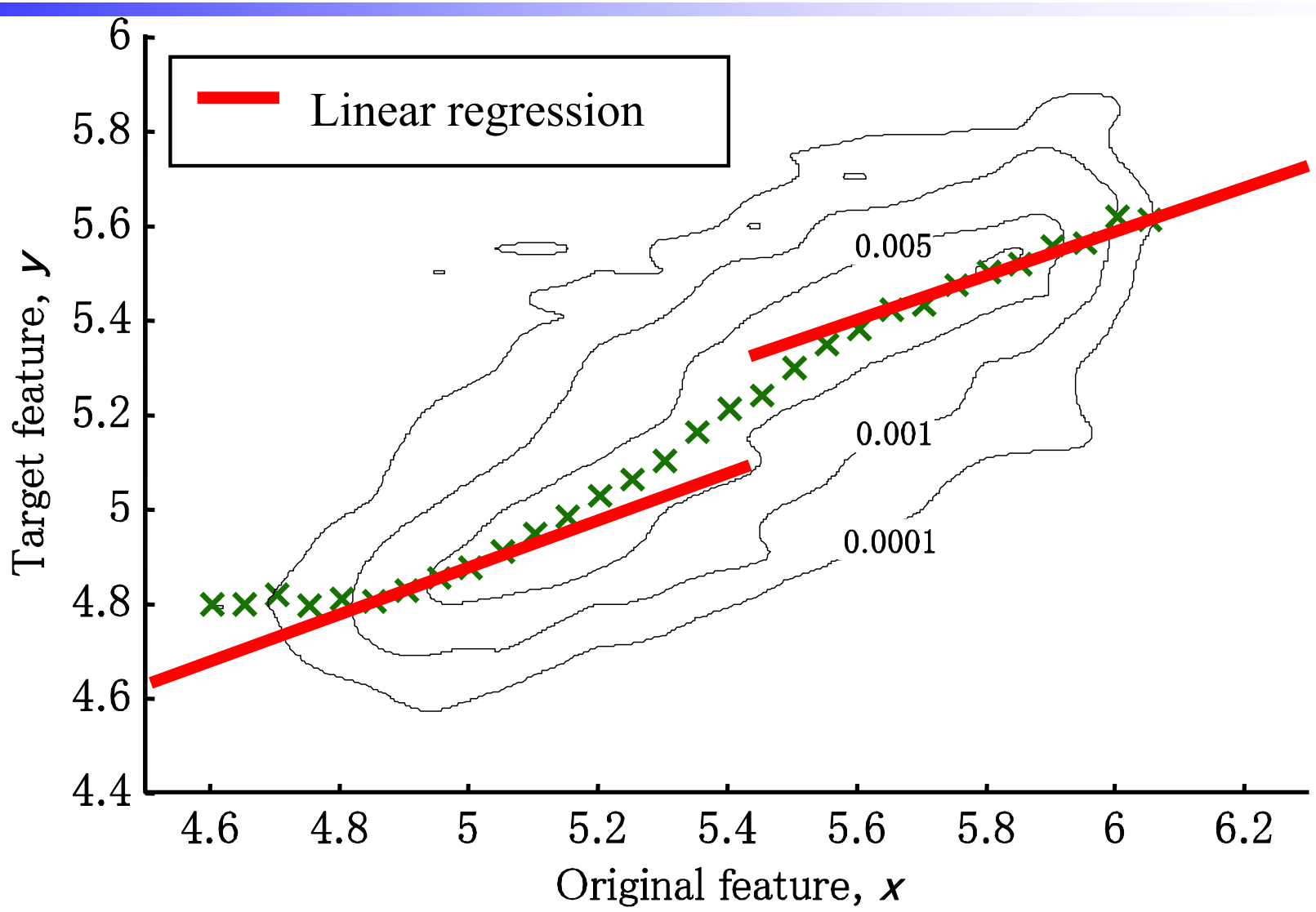
差分ベクトル使用時の変換関数



$$\hat{\boldsymbol{y}}_t = \boldsymbol{x}_t + \sum_{m=1}^{M} \gamma_{m,t}^{(x)} \left(\boldsymbol{\mu}_m^{(y)} - \boldsymbol{\mu}_m^{(x)} \right)$$

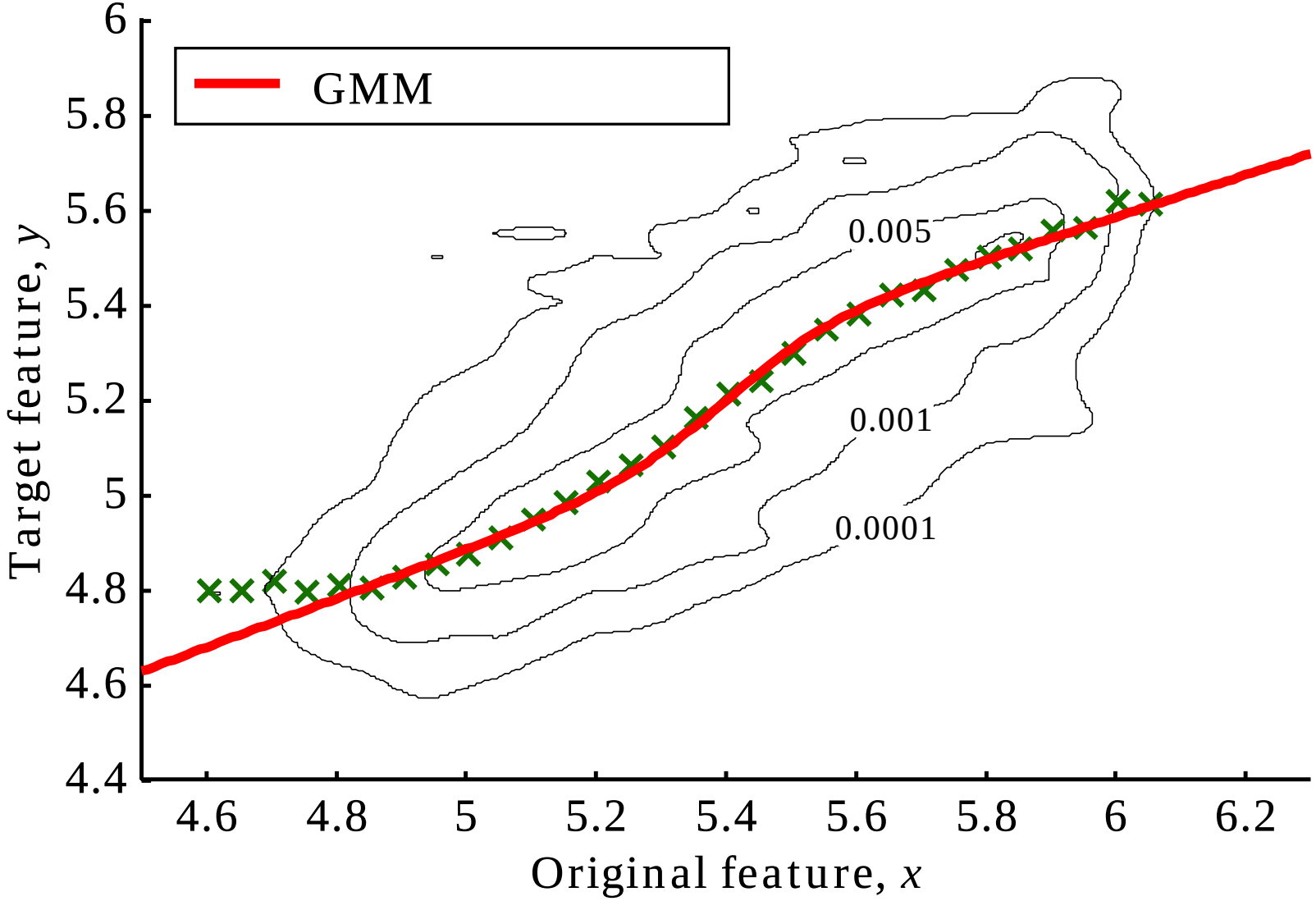
ソフトクラスタリングと連続的なマッピング

線形回帰に基づく変換関数



$$\hat{\boldsymbol{y}}_t = \boldsymbol{A}_m \boldsymbol{x}_t + \boldsymbol{b}_m$$ ハードクラスタリングと連続的かつ高精度な変換

混合正規分布モデルに基づく変換関数



$$\hat{\boldsymbol{y}}_t = \sum_{m=1}^{M} \gamma_{m,t}^{(x)} \left(\boldsymbol{A}_m \boldsymbol{x}_t + \boldsymbol{b}_m \right)$$ ソフトクラスタリングと連続的かつ高精度な変換

混合正規分布モデル（GMM）

- 正規分布の重み付け和により多様な形状の$p.d.f.$をモデル化する.
- モデルパラメータは各分布における混合重み α_m，平均ベクトル $\boldsymbol{\mu}_m$，並びに共分散行列 $\boldsymbol{\Sigma}_m$ である.

制約条件：$\sum_{m=1}^{M} \alpha_m = 1, \quad \alpha_m \geq 0$

1次元の場合

$$p.d.f.: \quad P(x_t \mid \boldsymbol{\lambda}) = \sum_{m=1}^{M} \alpha_m N(x_t; \mu_m, \Sigma_m)$$

m番目の正規分布：$N(x_t; \mu_m, \sigma_m^2) = \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left(-\frac{(x_t - \mu_m)^2}{2\sigma_m^2} \right)$

多次元の場合

$$p.d.f.: \quad P(\boldsymbol{x}_t \mid \boldsymbol{\lambda}) = \sum_{m=1}^{M} \alpha_m N(\boldsymbol{x}_t; \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$$

m番目の正規分布：

$$N(\boldsymbol{x}_t; \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) = \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}_m|}} \exp\left(-\frac{1}{2} (\boldsymbol{x}_t - \boldsymbol{\mu}_m)^{\mathrm{T}} \boldsymbol{\Sigma}_m^{-1} (\boldsymbol{x}_t - \boldsymbol{\mu}_m) \right)$$

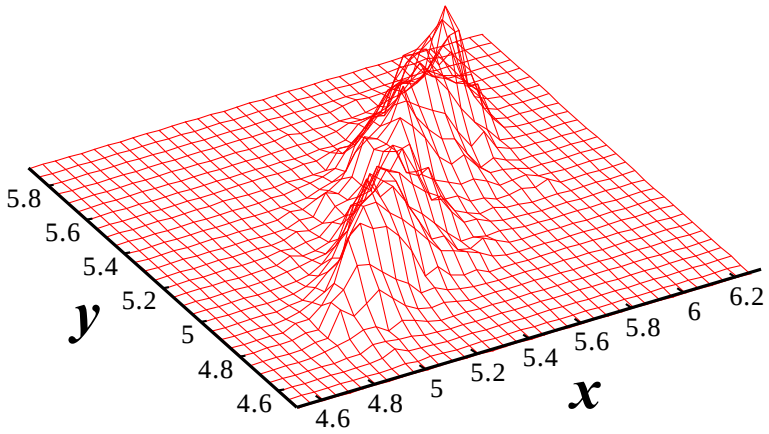
結合確率密度のモデル化

[Kain and Macon]

- パラレルデータから結合ベクトルを作成し，結合$p.d.f.$をGMMでモデル化する．

学習データのヒストグラム

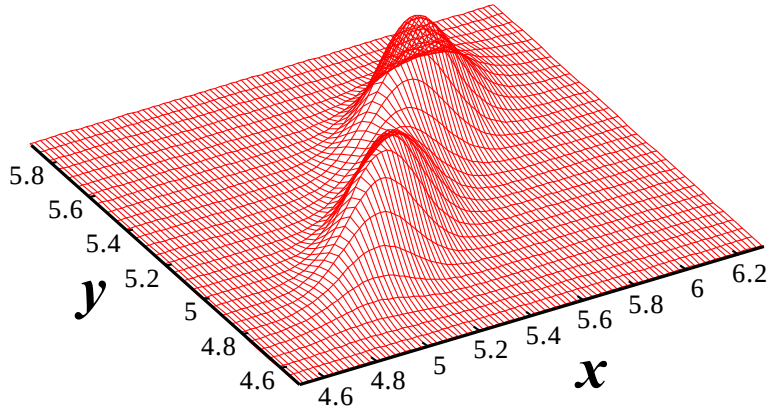
結合ベクトル：$\begin{bmatrix} \boldsymbol{x}_t \\ \boldsymbol{y}_t \end{bmatrix}$



GMM

$$p.d.f.: P(\boldsymbol{x}_t, \boldsymbol{y}_t \mid \boldsymbol{\lambda}^{(x,y)}) = \sum_{m=1}^{M} P\left(m \mid \boldsymbol{\lambda}^{(x,y)}\right) P\left(\boldsymbol{x}_t, \boldsymbol{y}_t \mid m, \boldsymbol{\lambda}^{(x,y)}\right)$$

$$= \sum_{m=1}^{M} \alpha_m N\left(\begin{bmatrix} \boldsymbol{x}_t \\ \boldsymbol{y}_t \end{bmatrix} \mid \boldsymbol{\mu}_m^{(x,y)}, \boldsymbol{\Sigma}_m^{(x,y)}\right)$$



GMMの構造

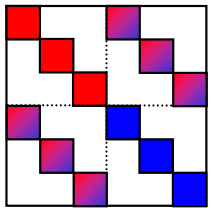
- m番目の混合分布パラメータ

次元間の相関をモデル化しない場合

同一の特徴量パラメータ間で変換を行う際に限って用いることができる.

重み　平均　分散

$$\alpha_m \quad \boldsymbol{\mu}_m^{(x,y)} = \begin{bmatrix} \boldsymbol{\mu}_m^{(x)} \\ \boldsymbol{\mu}_m^{(y)} \end{bmatrix} \quad \boldsymbol{\Sigma}_m^{(x,y)} = \begin{bmatrix} \boldsymbol{\Sigma}_m^{(xx)} & \boldsymbol{\Sigma}_m^{(xy)} \\ \boldsymbol{\Sigma}_m^{(yx)} & \boldsymbol{\Sigma}_m^{(yy)} \end{bmatrix}$$

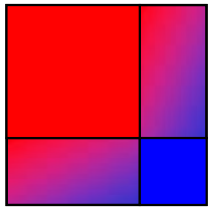


次元間の相関をモデル化する場合

異なる特徴量パラメータ間で変換を行う際には必要不可欠である.

重み　平均　分散

$$\alpha_m \quad \boldsymbol{\mu}_m^{(x,y)} = \begin{bmatrix} \boldsymbol{\mu}_m^{(x)} \\ \boldsymbol{\mu}_m^{(y)} \end{bmatrix} \quad \boldsymbol{\Sigma}_m^{(x,y)} = \begin{bmatrix} \boldsymbol{\Sigma}_m^{(xx)} & \boldsymbol{\Sigma}_m^{(xy)} \\ \boldsymbol{\Sigma}_m^{(yx)} & \boldsymbol{\Sigma}_m^{(yy)} \end{bmatrix}$$



条件付$p.d.f.$

- 結合$p.d.f.$: $P\left(\boldsymbol{x}_t, \boldsymbol{y}_t \mid \boldsymbol{\lambda}^{(x,y)}\right) = \sum_{m=1}^{M} P\left(m \mid \boldsymbol{\lambda}^{(x,y)}\right) P\left(\boldsymbol{x}_t \mid m, \boldsymbol{\lambda}^{(x,y)}\right) P\left(\boldsymbol{y}_t \mid \boldsymbol{x}_t, m, \boldsymbol{\lambda}^{(x,y)}\right)$

  周辺化すると $P\left(\boldsymbol{x}_t \mid \boldsymbol{\lambda}^{(x,y)}\right) = \sum_{m=1}^{M} P\left(m \mid \boldsymbol{\lambda}^{(x,y)}\right) P\left(\boldsymbol{x}_t \mid m, \boldsymbol{\lambda}^{(x,y)}\right)$

- 入力 $\boldsymbol{x}_t$ が与えられた際の条件付$p.d.f.$:

$$P\left(\boldsymbol{y}_t \mid \boldsymbol{x}_t, \boldsymbol{\lambda}^{(x,y)}\right) = \frac{P\left(\boldsymbol{x}_t, \boldsymbol{y}_t \mid \boldsymbol{\lambda}^{(x,y)}\right)}{P\left(\boldsymbol{x}_t \mid \boldsymbol{\lambda}^{(x,y)}\right)}$$

$$= \sum_{m=1}^{M} P\left(m \mid \boldsymbol{x}_t, \boldsymbol{\lambda}^{(x,y)}\right) P\left(\boldsymbol{y}_t \mid \boldsymbol{x}_t, m, \boldsymbol{\lambda}^{(x,y)}\right)$$

GMMでモデル化される.

事後確率

$$\frac{P\left(m \mid \boldsymbol{\lambda}^{(x,y)}\right) P\left(\boldsymbol{x}_t \mid m, \boldsymbol{\lambda}^{(x,y)}\right)}{\sum_{n=1}^{M} P\left(n \mid \boldsymbol{\lambda}^{(x,y)}\right) P\left(\boldsymbol{x}_t \mid n, \boldsymbol{\lambda}^{(x,y)}\right)}$$

正規分布

$$\text{平均} : \boldsymbol{\mu}_{m,t}^{(y|x)} = \boldsymbol{\mu}_m^{(y)} + \boldsymbol{\Sigma}_m^{(yx)} \boldsymbol{\Sigma}_m^{(xx)-1} \left(\boldsymbol{x}_t - \boldsymbol{\mu}_m^{(x)}\right)$$

$$= \boldsymbol{A}_m \boldsymbol{x}_t + \boldsymbol{b}_m$$

$$\text{共分散} : \boldsymbol{\Sigma}_m^{(y|x)} = \boldsymbol{\Sigma}_m^{(yy)} - \boldsymbol{\Sigma}_m^{(yx)} \boldsymbol{\Sigma}_m^{(xx)-1} \boldsymbol{\Sigma}_m^{(xy)}$$

最小平均自乗誤差推定

[Stylianou *et al.*]

- 入力が与えられた際の出力の条件付期待値へと変換する.

$$\hat{\boldsymbol{y}}_t = \int P\left(\boldsymbol{y}_t \mid \boldsymbol{x}_t, \boldsymbol{\lambda}^{(x,y)}\right)\boldsymbol{y}_t d\boldsymbol{y}_t$$

$$= \int \sum_{m=1}^{M} P\left(m \mid \boldsymbol{x}_t, \boldsymbol{\lambda}^{(x,y)}\right) P\left(\boldsymbol{y}_t \mid \boldsymbol{x}_t, m, \boldsymbol{\lambda}^{(x,y)}\right)\boldsymbol{y}_t d\boldsymbol{y}_t$$

$$= \sum_{m=1}^{M} P\left(m \mid \boldsymbol{x}_t, \boldsymbol{\lambda}^{(x,y)}\right) \int P\left(\boldsymbol{y}_t \mid \boldsymbol{x}_t, m, \boldsymbol{\lambda}^{(x,y)}\right)\boldsymbol{y}_t d\boldsymbol{y}_t$$

$$= \sum_{m=1}^{M} \gamma_{m,t}^{(x)}\left(\boldsymbol{A}_m \boldsymbol{x}_t + \boldsymbol{b}_m\right)$$

ソフトクラスタリング

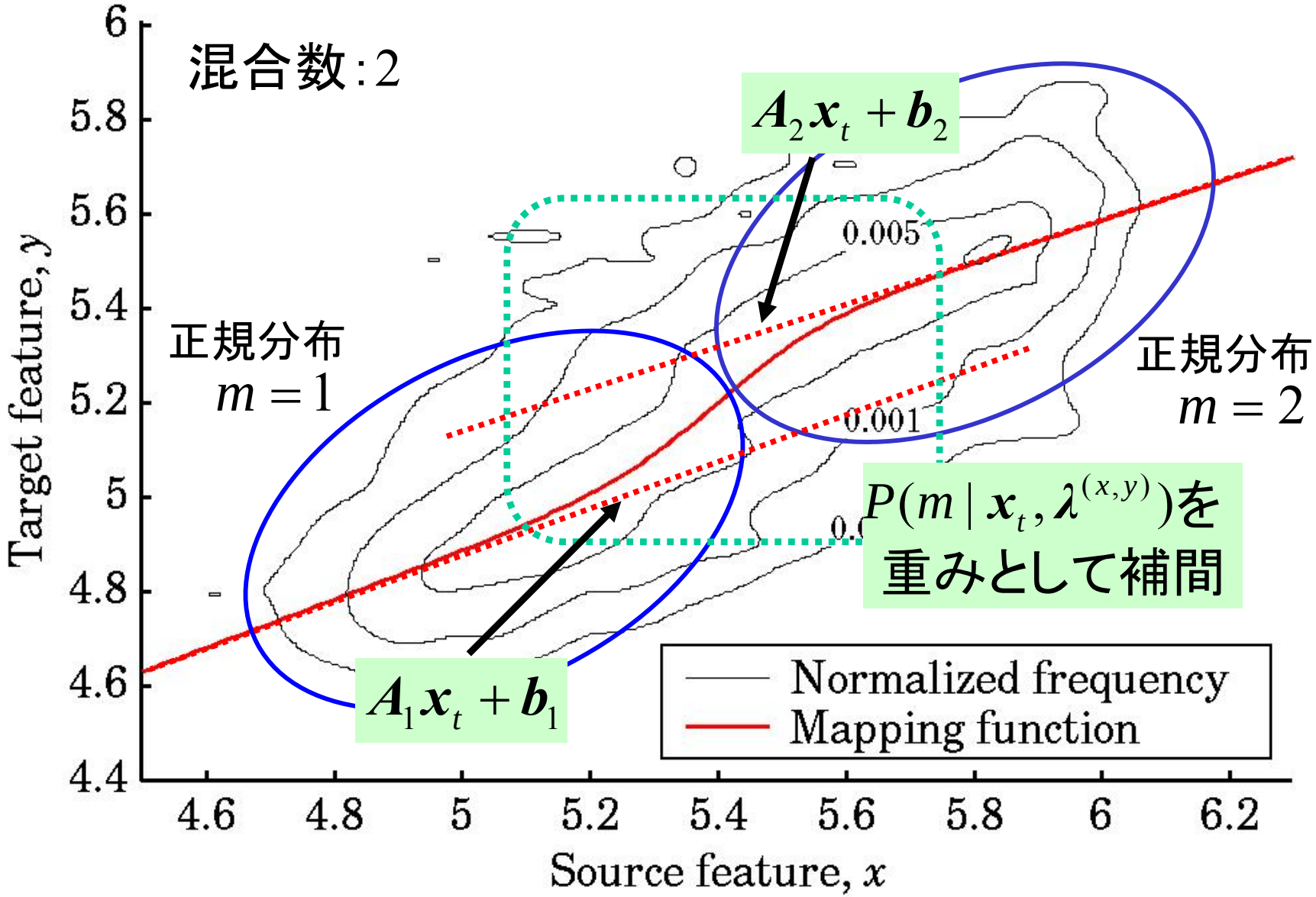
$$\gamma_{m,t}^{(x)} = P\left(m \mid \boldsymbol{x}_t, \boldsymbol{\lambda}^{(x,y)}\right) = \frac{\alpha_m N\left(\boldsymbol{x}_t; \boldsymbol{\mu}_m^{(x)}, \boldsymbol{\Sigma}_m^{(xx)}\right)}{\sum_{n=1}^{M} \alpha_n N\left(\boldsymbol{x}_t; \boldsymbol{\mu}_n^{(x)}, \boldsymbol{\Sigma}_n^{(xx)}\right)}$$

入・出力の相関を考慮した連続的な変換

$$\boldsymbol{A}_m = \boldsymbol{\Sigma}_m^{(yx)}\left(\boldsymbol{\Sigma}_m^{(xx)}\right)^{-1}$$

$$\boldsymbol{b}_m = \boldsymbol{\mu}_m^{(y)} - \boldsymbol{A}_m \boldsymbol{\mu}_m^{(x)}$$

GMMに基づく変換関数の解釈

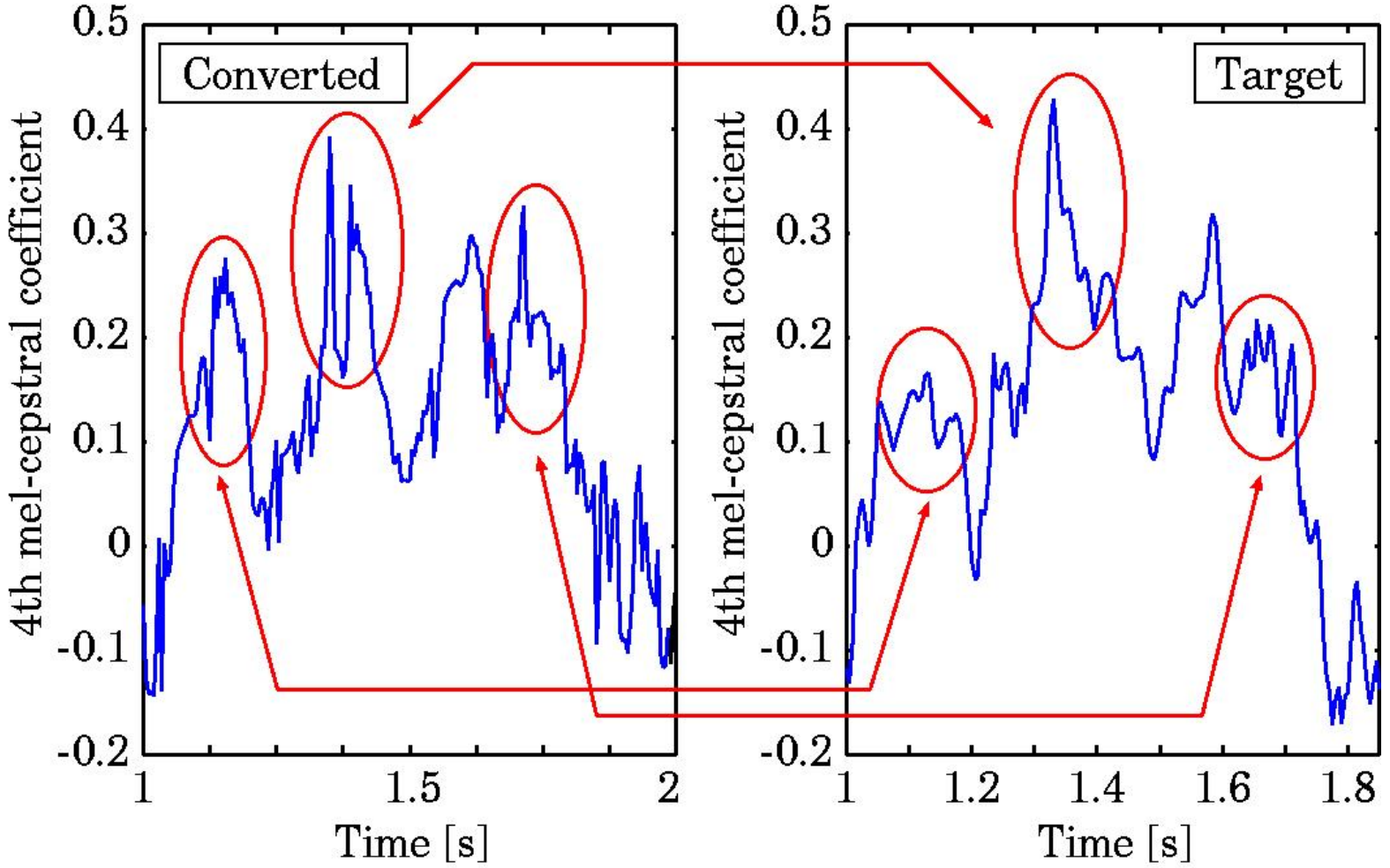


話者変換音声サンプル

- 話者：4名（男：bdl, 男：rms, 女：clb, 女：slt）
- 学習：50文対（完全自動学習）
- 変換法: 最小平均自乗誤差推定法（フレームベース変換法）

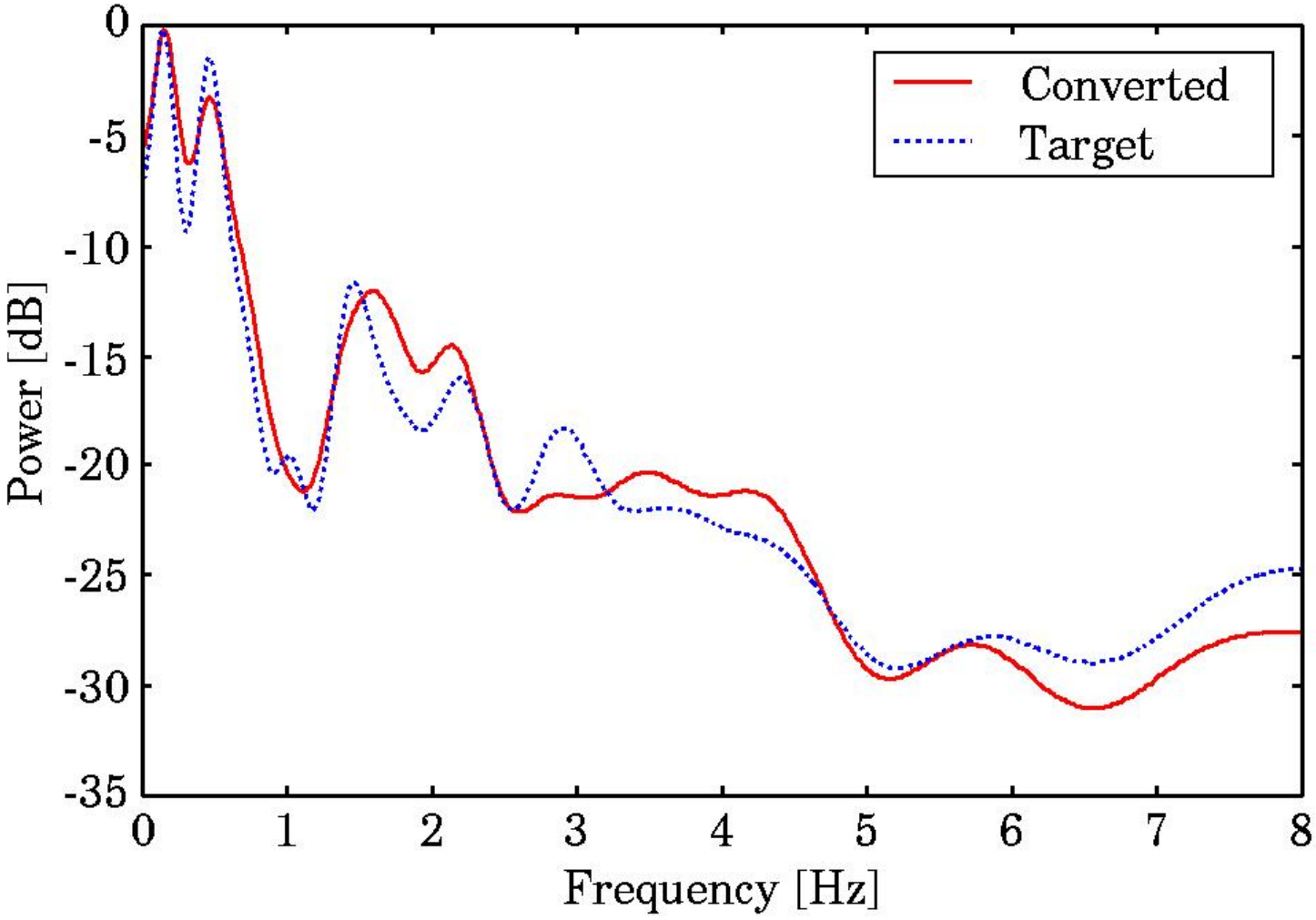
問題点1: 時間的依存関係の無視

- フレーム毎に独立して変換処理が行われるため，不適切なパラメータ遷移が発生する場合がある.



問題点2：過剰な平滑化

- 汎化処理により詳細なスペクトル構造が消失し，変換音声の肉声感が大幅に損なわれる．



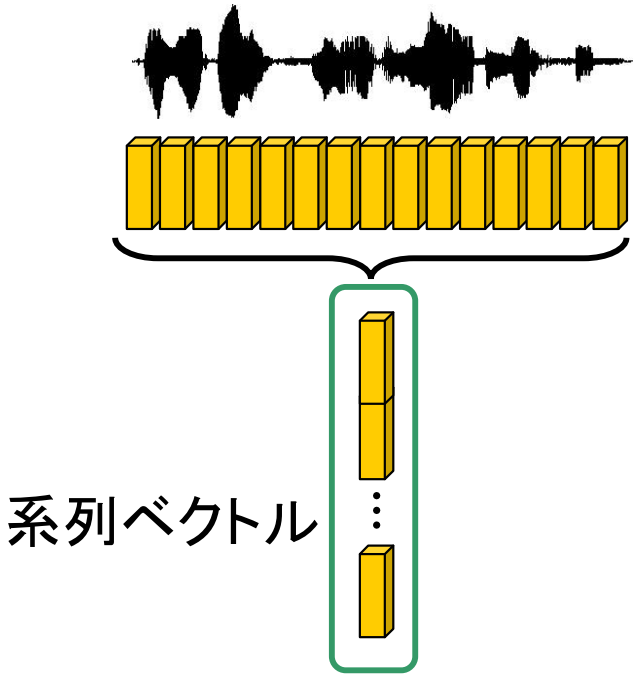
内容

1. 音声変換のしくみ
2. 統計的手法による声質変換
   2.1. 基本的な枠組み
   2.2. フレームベース変換法
   2.3. 系列ベース変換法
3. 応用例

系列ベースの変換関数

- 時系列単位で変換処理を行う.
  - 入力特徴量系列ベクトル: $\boldsymbol{x} = [\boldsymbol{x}_1^{\mathrm{T}} \mid \boldsymbol{x}_2^{\mathrm{T}} \mid \cdots \mid \boldsymbol{x}_T^{\mathrm{T}}]^{\mathrm{T}}$
  - 出力特徴量系列ベクトル: $\boldsymbol{y} = [\boldsymbol{y}_1^{\mathrm{T}} \mid \boldsymbol{y}_2^{\mathrm{T}} \mid \cdots \mid \boldsymbol{y}_T^{\mathrm{T}}]^{\mathrm{T}}$
  - 統計モデル: $\boldsymbol{\lambda}$
  - 変換特徴量系列ベクトル: $\hat{\boldsymbol{y}} = [\hat{\boldsymbol{y}}_1^{\mathrm{T}} \mid \hat{\boldsymbol{y}}_2^{\mathrm{T}} \mid \cdots \mid \hat{\boldsymbol{y}}_T^{\mathrm{T}}]^{\mathrm{T}}$

$$\hat{\boldsymbol{y}} = F_{\lambda}(\boldsymbol{x})$$
$$= \arg\max P(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\lambda})$$



最尤系列変換法

[Toda *et al.*]

- 系列に基づく特徴量を考慮した変換処理を行う.

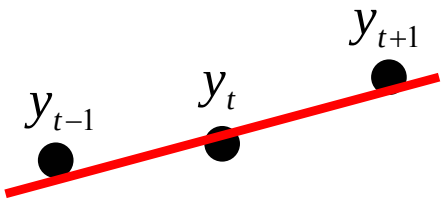
1. 動的特徴量を考慮した系列ベースの最尤推定
   - フレーム間相関を考慮した変換を行う.
   - 問題点1(時間的依存関係の無視)を解決できる.
2. 系列内変動の明示的なモデル化の導入
   - 2次モーメントを考慮した変換を行う.
   - 問題点2(過剰な平滑化)の影響を大幅に抑えることができる.

1. 動的特徴量の導入

- 学習時には，静的特徴量のみでなく動的特徴量も含んだ結合*p.d.f.*をGMMでモデル化する.

動的特徴量（微分係数）の計算

例えば $\Delta \boldsymbol{y}_t = \boldsymbol{y}_t - \boldsymbol{y}_{t-1}$ や $\Delta \boldsymbol{y}_t = \frac{1}{2}\left(\boldsymbol{y}_{t+1} - \boldsymbol{y}_{t-1}\right)$



回帰直線（$y=at+b$）の傾きa

静的・動的特徴量ベクトルに対する結合*p.d.f.*をモデル化する.

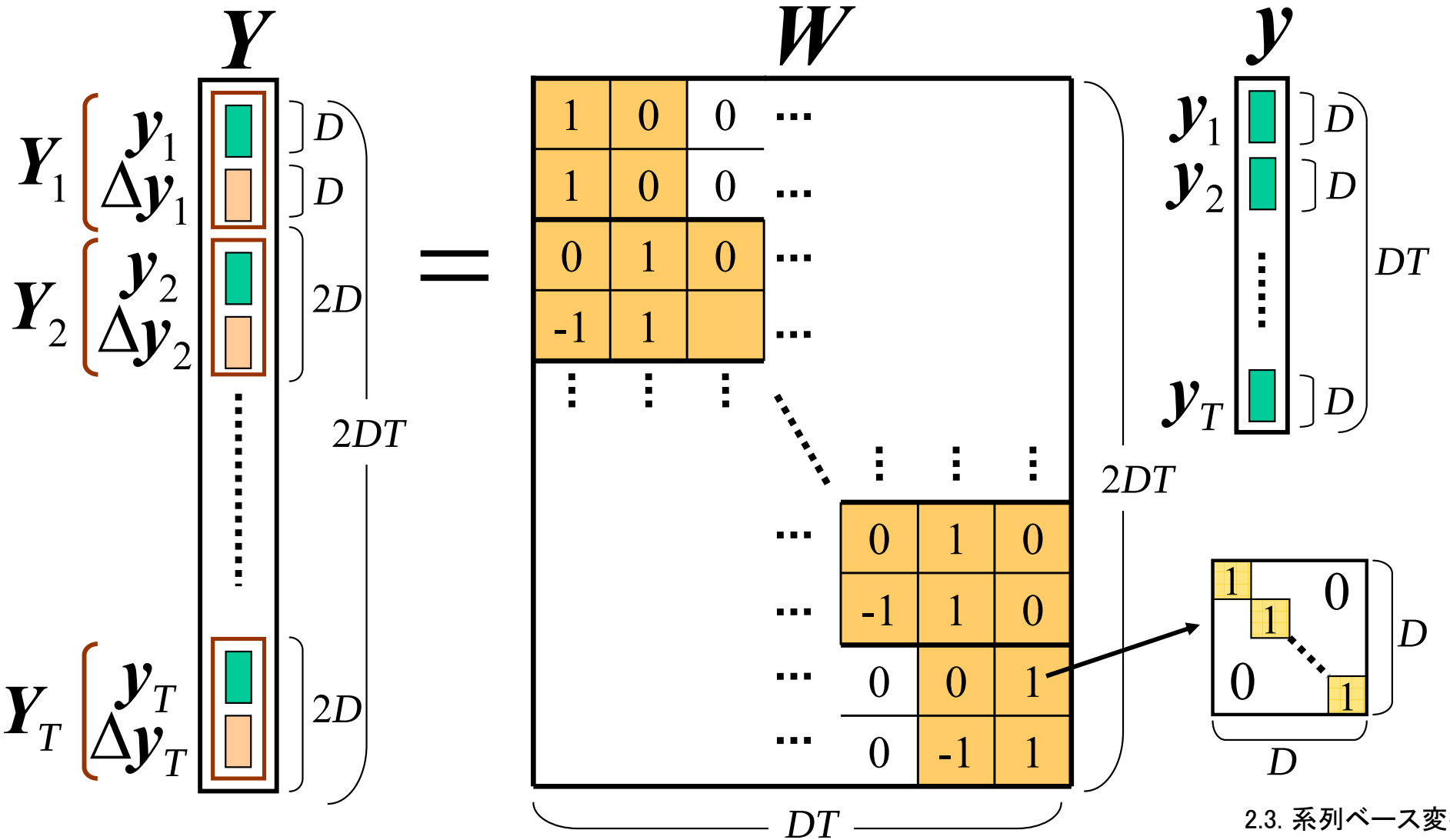
- 入力特徴量ベクトル: $\boldsymbol{X}_t = [\boldsymbol{x}_t^{\mathrm{T}} \ \ \Delta \boldsymbol{x}_t^{\mathrm{T}}]^{\mathrm{T}}$
- 出力特徴量ベクトル: $\boldsymbol{Y}_t = [\boldsymbol{y}_t^{\mathrm{T}} \ \ \Delta \boldsymbol{y}_t^{\mathrm{T}}]^{\mathrm{T}}$
- GMMによる結合*p.d.f.*のモデル化:

$$P\left(\boldsymbol{X}_t, \boldsymbol{Y}_t \mid \boldsymbol{\lambda}^{(X,Y)}\right) = \sum_{m=1}^{M} \alpha_m N\left(\begin{bmatrix} \boldsymbol{X}_t \\ \boldsymbol{Y}_t \end{bmatrix}; \begin{bmatrix} \boldsymbol{\mu}_m^{(X)} \\ \boldsymbol{\mu}_m^{(Y)} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_m^{(XX)} & \boldsymbol{\Sigma}_m^{(XY)} \\ \boldsymbol{\Sigma}_m^{(YX)} & \boldsymbol{\Sigma}_m^{(YY)} \end{bmatrix}\right)$$

- 変換時には，静的・動的特徴量間の明示的な関係を考慮して，条件付*p.d.f.*の尤度最大化に基づき入力特徴量ベクトルを変換する.

静的・動的特徴量の明示的な関係

- 静的・動的特徴量系列 $\boldsymbol{Y} = [\boldsymbol{Y}_1^{\mathrm{T}} \ \boldsymbol{Y}_2^{\mathrm{T}} \ \cdots \ \boldsymbol{Y}_T^{\mathrm{T}}]^{\mathrm{T}}$ は静的特徴量系列 $\boldsymbol{y} = [\boldsymbol{y}_1^{\mathrm{T}} \ \boldsymbol{y}_2^{\mathrm{T}} \ \cdots \ \boldsymbol{y}_T^{\mathrm{T}}]^{\mathrm{T}}$ の線形変換 $\boldsymbol{Y} = \boldsymbol{W}\boldsymbol{y}$ により表される．



動的特徴量を考慮した最尤系列変換

[Toda *et al.*]

- 静的・動的特徴量系列に対する条件付*p.d.f.*の尤度を最大化する静的特徴量系列を推定する.

制約条件: $\boldsymbol{Y} = \boldsymbol{W}\boldsymbol{y}$

$$\hat{\boldsymbol{y}} = \arg\max P(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\lambda}^{(X,Y)})$$

$$= \arg\max P(\boldsymbol{W}\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\lambda}^{(X,Y)})$$

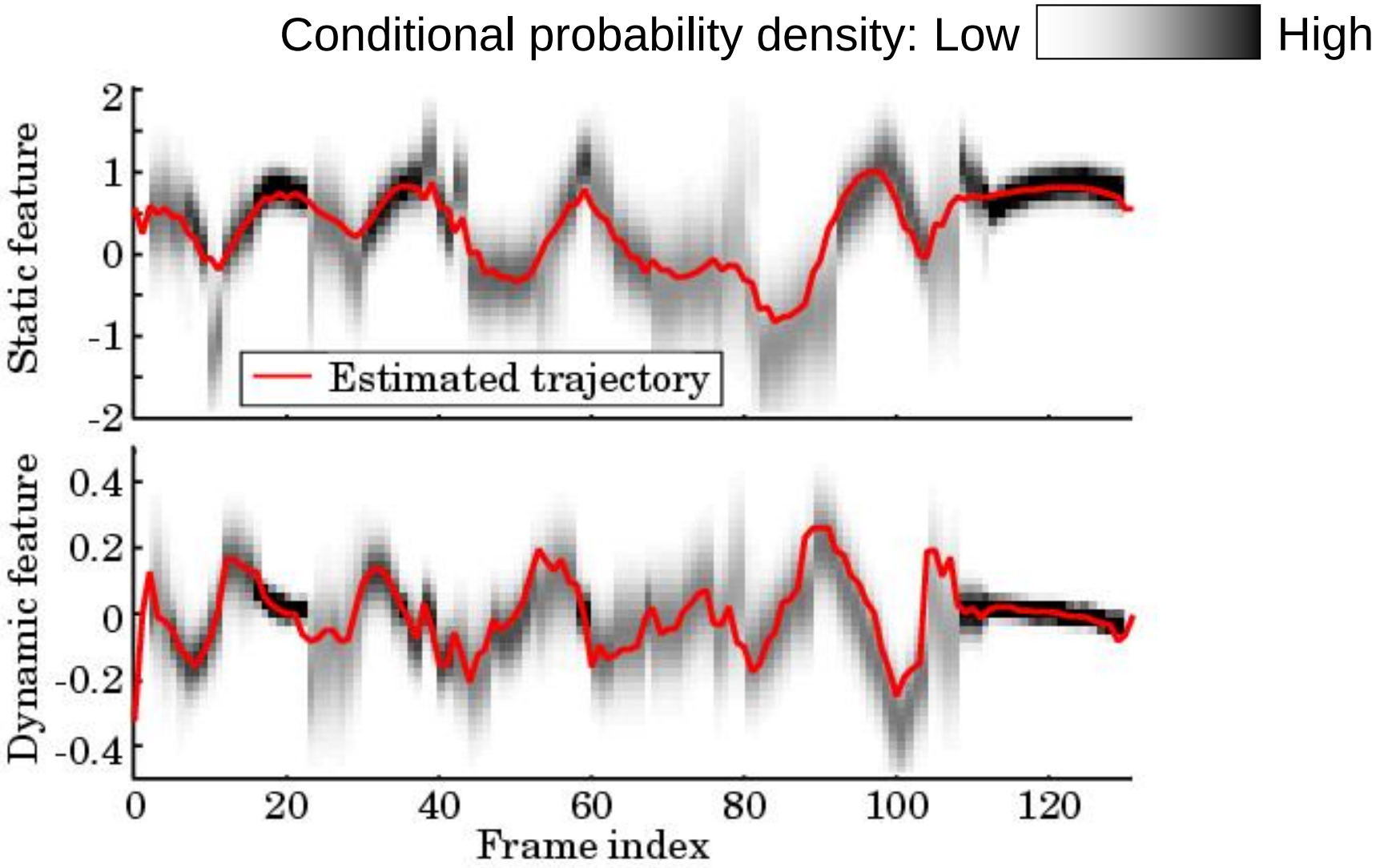
最尤な $\boldsymbol{Y}$ となる様に $\boldsymbol{y}$ を推定

静的特徴量系列に関する最大化

静的・動的特徴量系列に関する尤度

※HMM音声合成におけるパラメータ生成[Tokuda *et al.*]と同じ処理

変換特徴量系列の一例


Conditional probability density: Low
High
Static feature
2
1
0
-1
-2
Estimated trajectory
Dynamic feature
0.4
0.2
0
-0.2
-0.4
0
20
40
60
80
100
120
Frame index

系列単位の尤度関数

尤度関数

$$P\left(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\lambda}^{(X,Y)}\right) = \prod_{t=1}^{T} \sum_{m=1}^{M} P\left(m \mid \boldsymbol{X}_t, \boldsymbol{\lambda}^{(X,Y)}\right) P\left(\boldsymbol{Y}_t \mid \boldsymbol{X}_t, m, \boldsymbol{\lambda}^{(X,Y)}\right)$$

$$= \sum_{\text{all } \boldsymbol{m}} P\left(\boldsymbol{m} \mid \boldsymbol{X}, \boldsymbol{\lambda}^{(X,Y)}\right) P\left(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{m}, \boldsymbol{\lambda}^{(X,Y)}\right)$$

事後確率

$$P\left(\boldsymbol{m} \mid \boldsymbol{X}, \boldsymbol{\lambda}^{(X,Y)}\right) = \prod_{t=1}^{T} P\left(m_t \mid \boldsymbol{X}_t, \boldsymbol{\lambda}^{(X,Y)}\right)$$

分布系列 : $\boldsymbol{m} = \{m_1, m_2, \cdots m_T\}$

正規分布

$$P\left(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{m}, \boldsymbol{\lambda}^{(X,Y)}\right) = \prod_{t=1}^{T} P\left(\boldsymbol{Y}_t \mid \boldsymbol{X}_t, m_t, \boldsymbol{\lambda}^{(X,Y)}\right)$$

$$= P\left(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{m}, \boldsymbol{\lambda}^{(X,Y)}\right)$$

$$= N\left(\boldsymbol{Y}; \boldsymbol{\mu}_{\boldsymbol{m}}^{(Y|X)}, \boldsymbol{\Sigma}_{\boldsymbol{m}}^{(Y|X)}\right)$$

観測ベクトル

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \vdots \\ \boldsymbol{Y}_T \end{bmatrix}$$

平均ベクトル

$$\boldsymbol{\mu}_{\boldsymbol{m}}^{(Y|X)} = \begin{bmatrix} \boldsymbol{\mu}_{m_1}^{(Y|X)} \\ \boldsymbol{\mu}_{m_2}^{(Y|X)} \\ \vdots \\ \boldsymbol{\mu}_{m_T}^{(Y|X)} \end{bmatrix}$$

共分散行列

$$\boldsymbol{\Sigma}_{\boldsymbol{m}}^{(Y|X)} = \begin{bmatrix} \boldsymbol{\Sigma}_{m_1}^{(Y|X)} & 0 & \cdots & 0 \\ 0 & \boldsymbol{\Sigma}_{m_2}^{(Y|X)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\Sigma}_{m_T}^{(Y|X)} \end{bmatrix}$$

最尤特徴量系列の導出

- 単一分布系列 $\boldsymbol{m} = \{m_1, m_2, \cdots m_T\}$ による近似を用いる場合：

尤度関数

$$P\left(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{\lambda}^{(X,Y)}\right) = \sum_{\text{all } \boldsymbol{m}} P\left(\boldsymbol{m} \mid \boldsymbol{X}, \boldsymbol{\lambda}^{(X,Y)}\right) P\left(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{m}, \boldsymbol{\lambda}^{(X,Y)}\right)$$
$$\approx P\left(\boldsymbol{m} \mid \boldsymbol{X}, \boldsymbol{\lambda}^{(X,Y)}\right) P\left(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{m}, \boldsymbol{\lambda}^{(X,Y)}\right)$$

準最適な分布系列の決定（ハードクラスタリング）

$$\hat{\boldsymbol{m}} = \arg\max P\left(\boldsymbol{m} \mid \boldsymbol{X}, \boldsymbol{\lambda}^{(X,Y)}\right)$$

出力静的特徴量系列の推定

$$\hat{\boldsymbol{y}} = \arg\max P\left(\hat{\boldsymbol{m}} \mid \boldsymbol{X}, \boldsymbol{\lambda}^{(X,Y)}\right) P\left(\boldsymbol{Y} \mid \boldsymbol{X}, \hat{\boldsymbol{m}}, \boldsymbol{\lambda}^{(X,Y)}\right)$$
$$= \arg\max -\frac{1}{2}\left(\boldsymbol{W}\boldsymbol{y} - \boldsymbol{\mu}_{\hat{\boldsymbol{m}}}^{(Y|X)}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_{\hat{\boldsymbol{m}}}^{(Y|X)^{-1}} \left(\boldsymbol{W}\boldsymbol{y} - \boldsymbol{\mu}_{\hat{\boldsymbol{m}}}^{(Y|X)}\right)$$
$$= \left(\boldsymbol{W}^{\mathrm{T}} \boldsymbol{\Sigma}_{\hat{\boldsymbol{m}}}^{(Y|X)^{-1}} \boldsymbol{W}\right)^{-1} \boldsymbol{W}^{\mathrm{T}} \boldsymbol{\Sigma}_{\hat{\boldsymbol{m}}}^{(Y|X)^{-1}} \boldsymbol{\mu}_{\hat{\boldsymbol{m}}}^{(Y|X)}$$

※EMアルゴリズムを用いてソフトクラスタリングを行うことも可能である.

トラジェクトリモデルとしての解釈

[Zen *et al.*]

静的特徴量系列ベクトル $\boldsymbol{y}$ を確率変数としたモデルの尤度最大化

$$P(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{m}, \boldsymbol{\lambda}) = P(\boldsymbol{W}\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{m}, \boldsymbol{\lambda})$$

$$= \frac{1}{\sqrt{(2\pi)^{2DT}\left|\boldsymbol{\Sigma}_{\boldsymbol{m}}^{(Y|X)}\right|}} \exp\left(-\frac{1}{2}\left(\boldsymbol{W}\boldsymbol{y} - \boldsymbol{\mu}_{\boldsymbol{m}}^{(Y|X)}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{m}}^{(Y|X)^{-1}} \left(\boldsymbol{W}\boldsymbol{y} - \boldsymbol{\mu}_{\boldsymbol{m}}^{(Y|X)}\right)\right)$$

$$= Z_{\boldsymbol{m}} \frac{1}{\sqrt{(2\pi)^{2DT}\left|\boldsymbol{P}_{\boldsymbol{m}}\right|}} \exp\left(-\frac{1}{2}\left(\boldsymbol{y} - \bar{\boldsymbol{y}}_{\boldsymbol{m}}\right)^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{m}}^{-1} \left(\boldsymbol{y} - \bar{\boldsymbol{y}}_{\boldsymbol{m}}\right)\right)$$

$$= Z_{\boldsymbol{m}} P(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{m}, \boldsymbol{\lambda})$$

正規分布

平均：$\boldsymbol{\mu}_{\boldsymbol{m}}^{(Y|X)}$

共分散：$\boldsymbol{\Sigma}_{\boldsymbol{m}}^{(Y|X)}$

正規化項

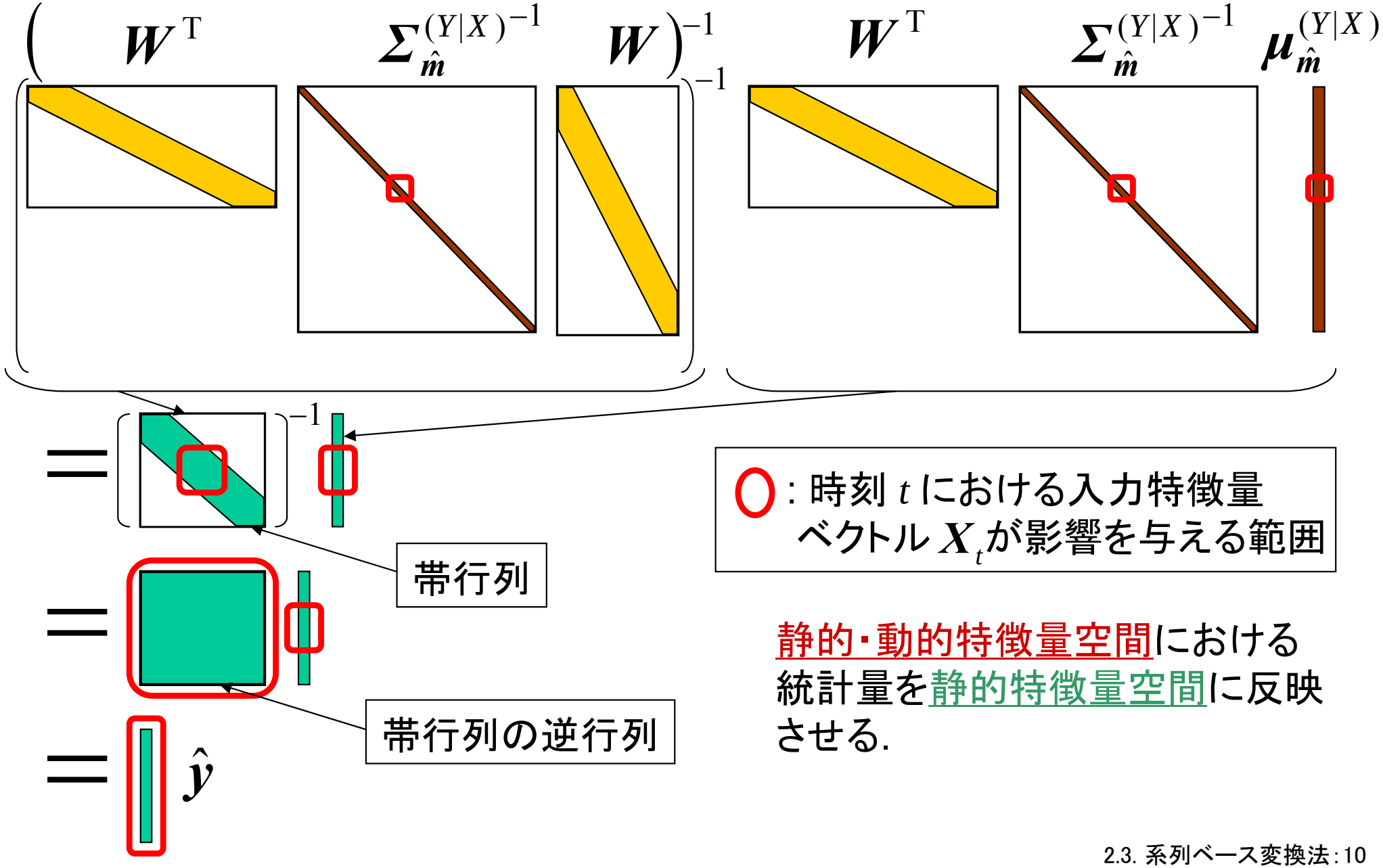
$$\frac{\sqrt{(2\pi)^{DT}\left|\boldsymbol{P}_{\boldsymbol{m}}\right|}}{\sqrt{(2\pi)^{2DT}\left|\boldsymbol{\Sigma}_{\boldsymbol{m}}^{(Y|X)}\right|}} \cdot \exp\left(-\frac{1}{2}\left(\boldsymbol{\mu}_{\boldsymbol{m}}^{(Y|X)^{\mathrm{T}}} \boldsymbol{\Sigma}_{\boldsymbol{m}}^{(Y|X)^{-1}} \boldsymbol{\mu}_{\boldsymbol{m}}^{(Y|X)} - \bar{\boldsymbol{y}}_{\boldsymbol{m}}^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{m}}^{-1} \bar{\boldsymbol{y}}_{\boldsymbol{m}}\right)\right)$$

正規分布

平均：$\bar{\boldsymbol{y}}_{\boldsymbol{m}} = \left(\boldsymbol{W}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{m}}^{(Y|X)^{-1}} \boldsymbol{W}\right)^{-1} \boldsymbol{W}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{m}}^{(Y|X)^{-1}} \boldsymbol{\mu}_{\boldsymbol{m}}^{(Y|X)}$

共分散：$\boldsymbol{P}_{\boldsymbol{m}} = \left(\boldsymbol{W}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{m}}^{(Y|X)^{-1}} \boldsymbol{W}\right)^{-1}$

最尤特徴量系列の演算



○：時刻 t における入力特徴量ベクトル X_t が影響を与える範囲

静的・動的特徴量空間における統計量を静的特徴量空間に反映させる.

コレスキー分解を用いた解法

[Tokuda *et al.*]

生成パラメータ

$TD \times TD$サイズの逆行列演算

$$\hat{\boldsymbol{y}} = \boldsymbol{R}_{\hat{m}}^{-1} \boldsymbol{r}_{\hat{m}} \quad \text{ここで } \boldsymbol{R}_{\hat{m}} = \boldsymbol{W}^{\mathrm{T}} \boldsymbol{\Sigma}_{\hat{m}}^{(Y|X)^{-1}} \boldsymbol{W} \text{ および } \boldsymbol{r}_{\hat{m}} = \boldsymbol{W}^{\mathrm{T}} \boldsymbol{\Sigma}_{\hat{m}}^{(Y|X)^{-1}} \boldsymbol{\mu}_{\hat{m}}^{(Y|X)}$$

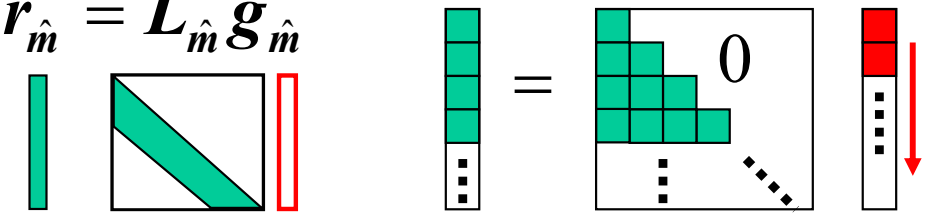
コレスキー分解

$$\boldsymbol{R}_{\hat{m}} = \boldsymbol{L}_{\hat{m}} \boldsymbol{L}_{\hat{m}}^{\mathrm{T}}$$

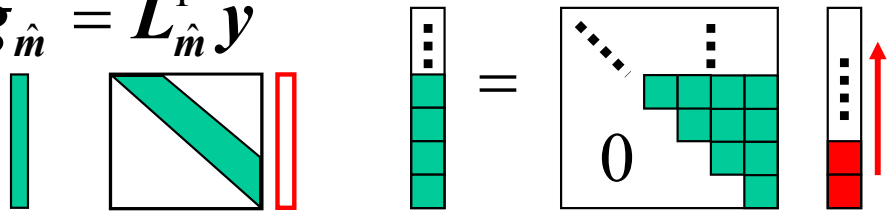
下三角行列とその転置の積に分解

$$\boldsymbol{r}_{\hat{m}} = \boldsymbol{R}_{\hat{m}} \hat{\boldsymbol{y}} = \boldsymbol{L}_{\hat{m}} \boldsymbol{L}_{\hat{m}}^{\mathrm{T}} \hat{\boldsymbol{y}} \text{ ここで } \boldsymbol{L}_{\hat{m}}^{\mathrm{T}} \hat{\boldsymbol{y}} = \boldsymbol{g}_{\hat{m}} \text{ とすると } \boldsymbol{r}_{\hat{m}} = \boldsymbol{L}_{\hat{m}} \boldsymbol{g}_{\hat{m}}$$

前進代入 $\boldsymbol{r}_{\hat{m}} = \boldsymbol{L}_{\hat{m}} \boldsymbol{g}_{\hat{m}}$



後退代入 $\boldsymbol{g}_{\hat{m}} = \boldsymbol{L}_{\hat{m}}^{\mathrm{T}} \hat{\boldsymbol{y}}$

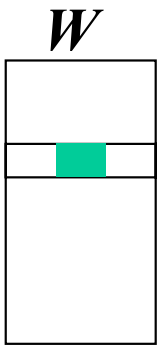


逆行列演算を必要とせず
簡単な代入処理による計算

再帰式を用いた解法

[Tokuda *et al.*]

$$\hat{\boldsymbol{y}} = \left(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{\boldsymbol{m}}}^{(Y|X)^{-1}}\boldsymbol{W}\right)^{-1}\left(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{\boldsymbol{m}}}^{(Y|X)^{-1}}\boldsymbol{\mu}_{\hat{\boldsymbol{m}}}^{(Y|X)}\right)$$



Δy_t を求める行

$\boldsymbol{w}_t = [\boldsymbol{0},\cdots,\boldsymbol{0},-\boldsymbol{I},\boldsymbol{I},\boldsymbol{0},\cdots,\boldsymbol{0}]$

$\boldsymbol{W}$ を静的特徴量を求める行と
動的特徴量を求める行に分解すると

$$= \left(\boldsymbol{\Sigma}_{\hat{\boldsymbol{m}}}^{(y|X)^{-1}} + \sum_{t=1}^{T}\boldsymbol{w}_t^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{m}_t}^{(\Delta y|X)^{-1}}\boldsymbol{w}_t\right)^{-1}\left(\boldsymbol{\Sigma}_{\hat{\boldsymbol{m}}}^{(y|X)^{-1}}\boldsymbol{\mu}_{\hat{\boldsymbol{m}}}^{(y|X)} + \sum_{t=1}^{T}\boldsymbol{w}_t^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{m}_t}^{(\Delta y|X)^{-1}}\boldsymbol{\mu}_{\hat{m}_t}^{(\Delta y|X)}\right)$$

静的特徴量に関する項　　動的特徴量に関する項

動的特徴量を考慮しない場合：

※静的・動的特徴量の相互共分散は無視

$$\hat{\boldsymbol{y}}_{(0)} = \boldsymbol{P}_{(0)}\boldsymbol{r}_{(0)} = \boldsymbol{\mu}_{\hat{\boldsymbol{m}}}^{(y|X)}$$

$$\begin{cases}\boldsymbol{P}_{(0)} = \boldsymbol{\Sigma}_{\hat{\boldsymbol{m}}}^{(y|X)} \\ \boldsymbol{r}_{(0)} = \boldsymbol{\Sigma}_{\hat{\boldsymbol{m}}}^{(y|X)^{-1}}\boldsymbol{\mu}_{\hat{\boldsymbol{m}}}^{(y|X)}\end{cases}$$

⇐ フレーム毎に独立に計算可能

フレーム t までの動的特徴量を考慮した場合：

$$\hat{\boldsymbol{y}}_{(t)} = \boldsymbol{P}_{(t)}\boldsymbol{r}_{(t)}$$

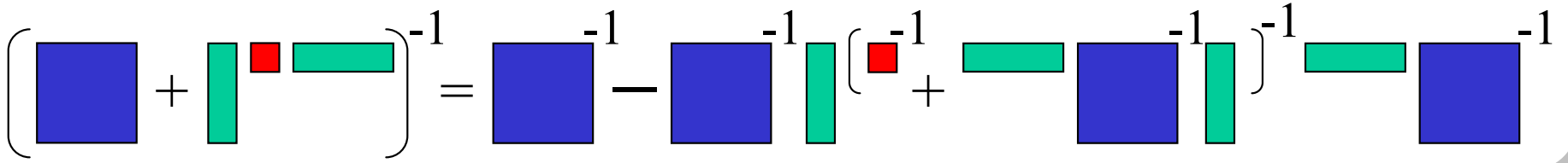
$$\begin{cases}\boldsymbol{P}_{(t)} = \left(\boldsymbol{P}_{(t-1)}^{-1} + \boldsymbol{w}_t^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{m}_t}^{(\Delta y|X)^{-1}}\boldsymbol{w}_t\right)^{-1} \\ \boldsymbol{r}_{(t)} = \boldsymbol{r}_{(t-1)} + \boldsymbol{w}_t^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{m}_t}^{(\Delta y|X)^{-1}}\boldsymbol{\mu}_{\hat{m}_t}^{(\Delta y|X)}\end{cases}$$

⇐ フレーム $t-1$ の結果を用いて再帰的に計算可能

逆行列補題の適用

行列Aとその逆行列A^{-1}が既知の時，低ランクの変動vNv^{T}が
行列Aに加わった際のその逆行列$\left(A+vNv^{\mathrm{T}}\right)^{-1}$は次式で計算可能

$$\left(A+vNv^{\mathrm{T}}\right)^{-1}=A^{-1}-A^{-1}v\left(N^{-1}+v^{\mathrm{T}}A^{-1}v\right)^{-1}v^{\mathrm{T}}A^{-1}$$



$$\boldsymbol{P}_{(t)}=\left(\boldsymbol{P}_{(t-1)}^{-1}+\boldsymbol{w}_t^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{m}_t}^{(\Delta y|X)^{-1}}\boldsymbol{w}_t\right)^{-1}=\boldsymbol{P}_{(t-1)}-\underline{\boldsymbol{P}_{(t-1)}\boldsymbol{w}_t^{\mathrm{T}}\left(\boldsymbol{\Sigma}_{\hat{m}_t}^{(\Delta y|X)}+\boldsymbol{w}_t\boldsymbol{P}_{(t-1)}\boldsymbol{w}_t^{\mathrm{T}}\right)^{-1}}\boldsymbol{w}_t\boldsymbol{P}_{(t-1)}$$

$$=\left(\boldsymbol{I}-\boldsymbol{k}_{(t)}\boldsymbol{w}_t\right)\boldsymbol{P}_{(t-1)}$$

$=\boldsymbol{k}_{(t)}$とおく.

$$\hat{\boldsymbol{y}}_{(t)}=\boldsymbol{P}_{(t)}\left(\boldsymbol{r}_{(t-1)}+\boldsymbol{w}_t^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{m}_t}^{(\Delta y|X)^{-1}}\boldsymbol{\mu}_{\hat{m}_t}^{(\Delta y|X)}\right)$$

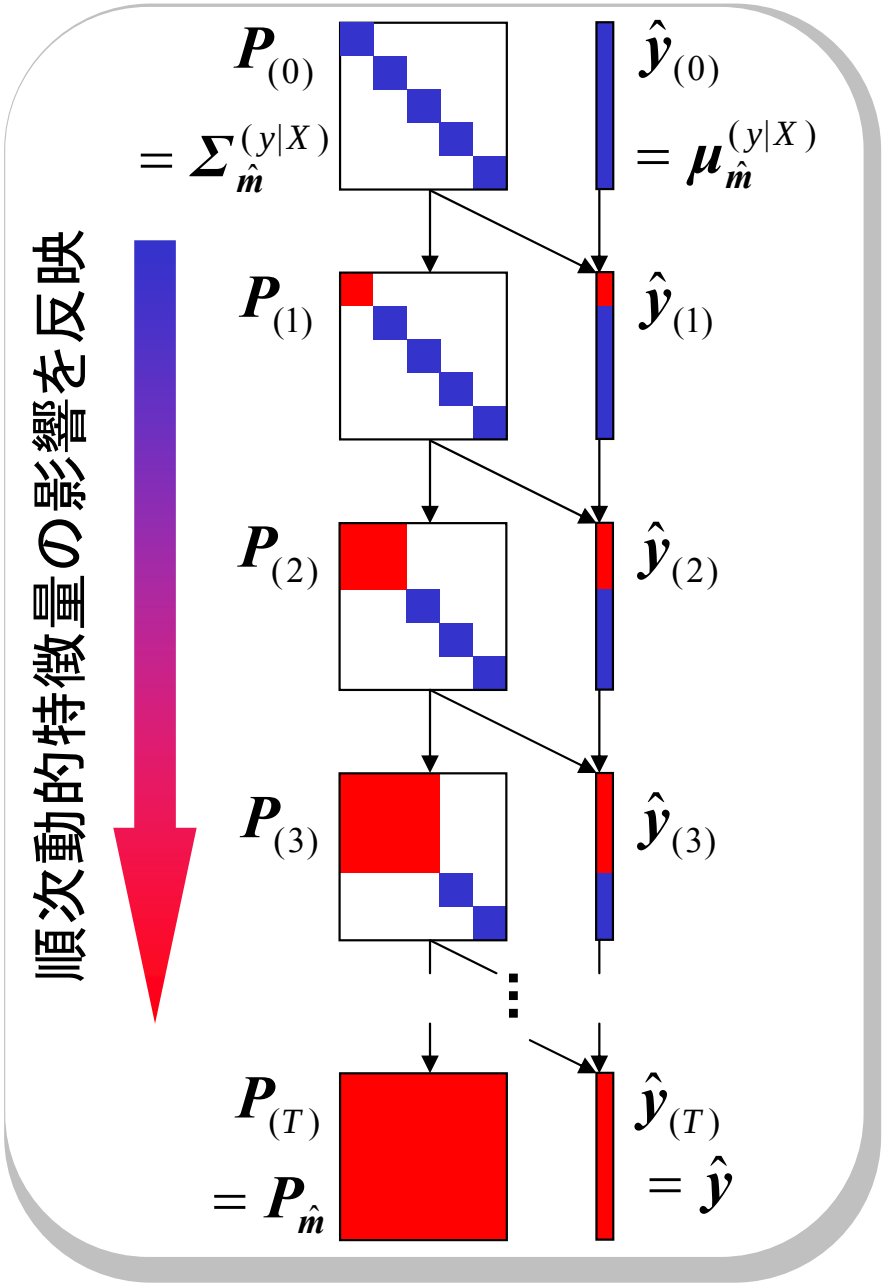
$$=\left(\boldsymbol{I}-\boldsymbol{k}_{(t)}\boldsymbol{w}_t\right)\boldsymbol{P}_{(t-1)}\left(\underline{\boldsymbol{r}_{(t-1)}}+\underline{\boldsymbol{w}_t^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{m}_t}^{(\Delta y|X)^{-1}}\boldsymbol{\mu}_{\hat{m}_t}^{(\Delta y|X)}}\right)$$

$$=\underline{\left(\boldsymbol{I}-\boldsymbol{k}_{(t)}\boldsymbol{w}_t\right)\hat{\boldsymbol{y}}_{(t-1)}}+\underline{\left(\boldsymbol{I}-\boldsymbol{k}_{(t)}\boldsymbol{w}_t\right)\boldsymbol{P}_{(t-1)}\boldsymbol{w}_t^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{m}_t}^{(\Delta y|X)^{-1}}\boldsymbol{\mu}_{\hat{m}_t}^{(\Delta y|X)}}$$

$$=\hat{\boldsymbol{y}}_{(t-1)}+\boldsymbol{k}_{(t)}\left(\boldsymbol{\mu}_{\hat{m}_t}^{(\Delta y|X)}-\boldsymbol{w}_t\,\hat{\boldsymbol{y}}_{(t-1)}\right)$$

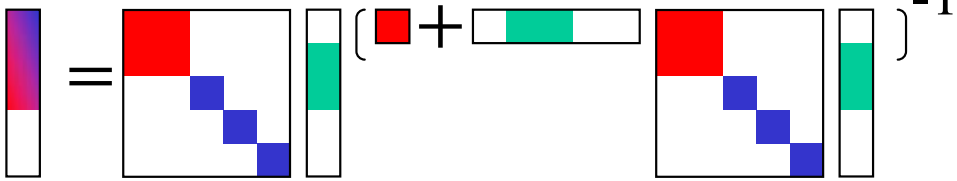
$=\boldsymbol{k}_{(t)}\boldsymbol{\mu}_{\hat{m}_t}^{(\Delta y|X)}$になる.

更新式



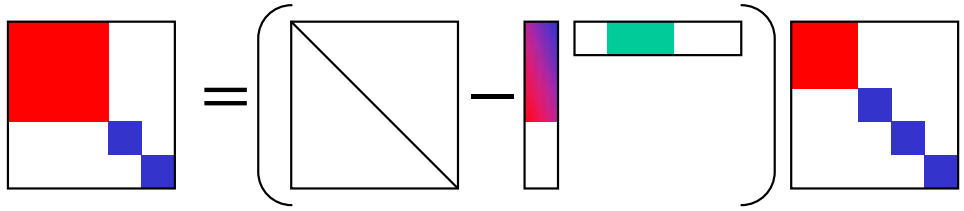
$k_{(t)}$（カルマンゲイン）の計算

$$k_{(t)} = P_{(t-1)} w_t^{\mathrm{T}} \left(\Sigma^{(\Delta y|X)}_{\hat{m}_t} + w_t P_{(t-1)} w_t^{\mathrm{T}} \right)^{-1}$$



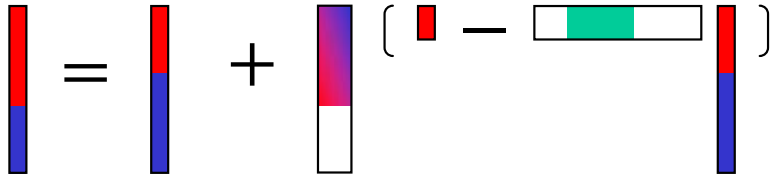
$P_{(t)}$（誤差の共分散）の更新

$$P_{(t)} = \left(I - k_{(t)} w_t \right) P_{(t-1)}$$



$\hat{y}_{(t)}$（推定状態）の更新

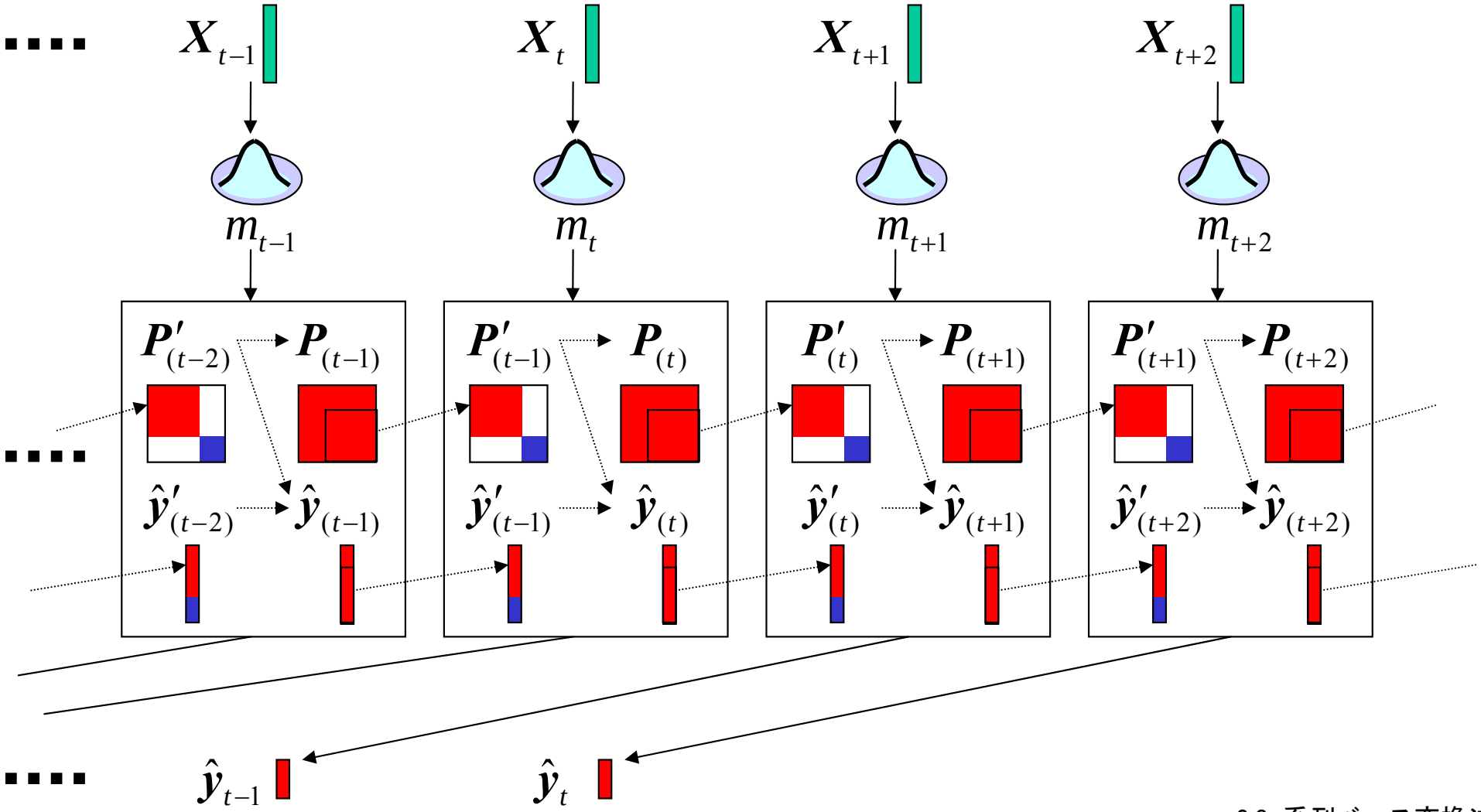
$$\hat{y}_{(t)} = \hat{y}_{(t-1)} + k_{(t)} \left(\mu^{(\Delta y|X)}_{\hat{m}_t} - w_t \hat{y}_{(t-1)} \right)$$



短遅延変換処理

[Muramatsu *et al.*]

動的特徴量が与える影響の範囲を数フレームに限定することで
短遅延変換処理が実現可能



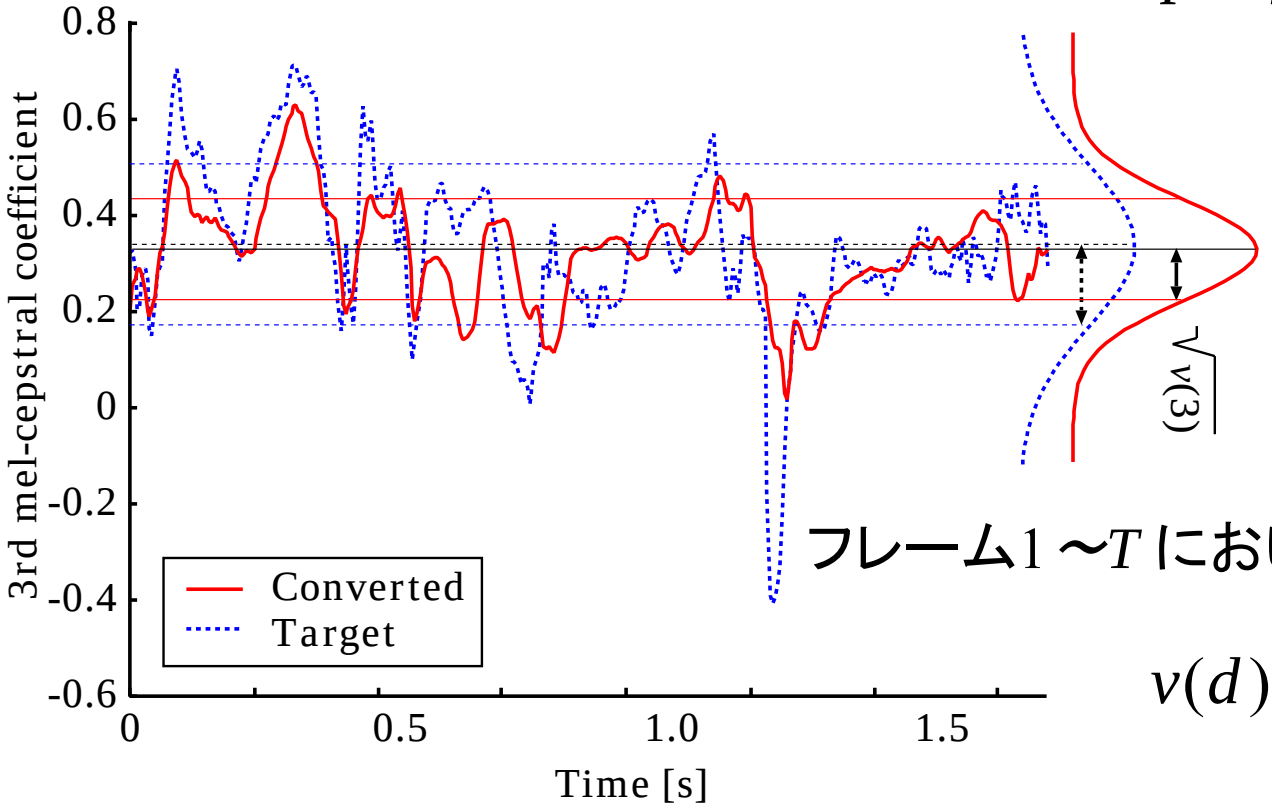
最尤系列変換法

[Toda *et al.*]

- 系列に基づく特徴量を考慮した変換処理を行う.
  1. 動的特徴量を考慮した系列ベースの最尤推定
     - フレーム間相関を考慮した変換を行う.
     - 問題点1(時間的依存関係の無視)を解決できる.
  2. 系列内変動の明示的なモデル化の導入
     - 2次モーメントを考慮した変換を行う.
     - 問題点2(過剰な平滑化)の影響を大幅に抑えることができる.

2. 系列内変動モデルの導入

- 系列内（例えば一発話）に含まれる全フレームにわたって計算される分散（Global Variance: GV）の*p.d.f.*をモデル化する.



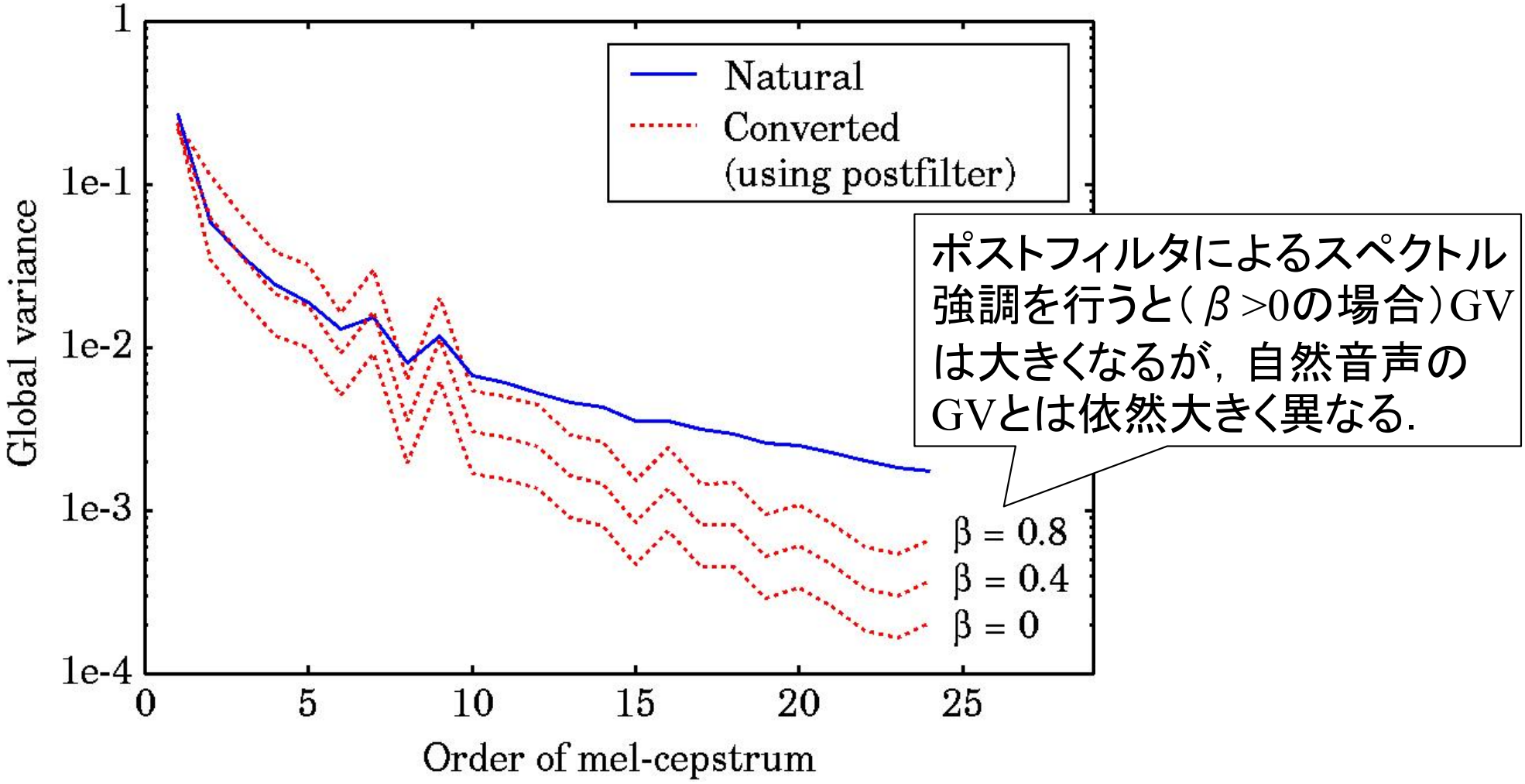
フレーム1～Tにおいて計算されるd次元目のGV:

$$v(d) = \frac{1}{T}\sum_{t=1}^{T}\left(y_t^{(d)} - \frac{1}{T}\sum_{\tau=1}^{T} y_\tau^{(d)} \right)^2$$

- 変換時には静的特徴量とGV間の明示的な関係を考慮して，条件付*p.d.f.*及びGVの*p.d.f.*の尤度最大化に基づき入力特徴量ベクトルを変換する.

GVの統計的特徴

- 一般に，統計的変換処理で得られる変換特徴量系列のGVは減少する.



ポストフィルタによるスペクトル強調を行うと（$\beta > 0$の場合）GVは大きくなるが，自然音声のGVとは依然大きく異なる.

GVを考慮した最尤系列変換法

[Toda *et al.*]

- GVの確率密度も考慮した尤度関数を用いる.

$$L(\boldsymbol{y}) = \log\left\{ P(\boldsymbol{W}\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{\lambda})^{\omega} \cdot P(\boldsymbol{v}(\boldsymbol{y}) \mid \boldsymbol{\lambda}_v) \right\}$$

静的・動的特徴量系列に関する尤度

静的特徴量系列のGV $\boldsymbol{v}(\boldsymbol{y})$ に関する尤度

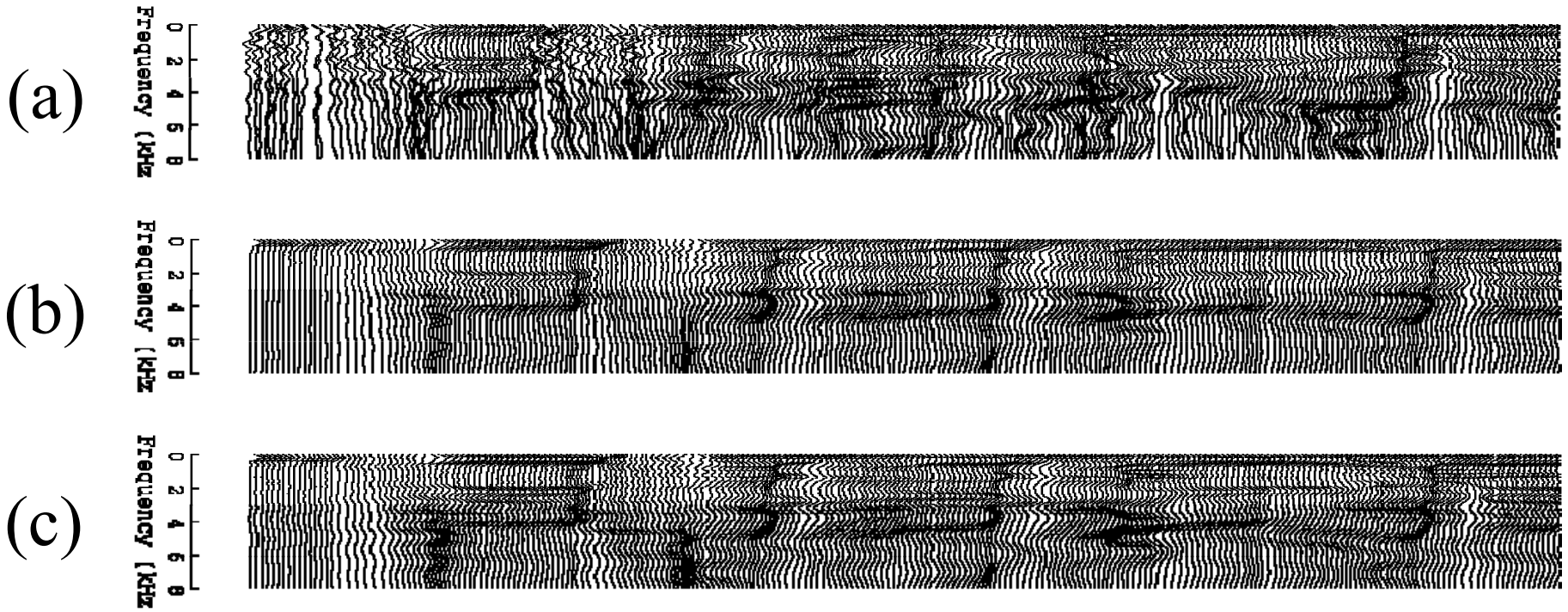
ω : 2つの尤度のバランスを調節する重み

- 勾配法により尤度最大化を行う.

$$\hat{\boldsymbol{y}} = \arg\max L(\boldsymbol{y})$$

最急降下法: $\boldsymbol{y}^{(i+1)-th} = \boldsymbol{y}^{(i)-th} + \alpha \cdot \left. \frac{\partial L(\boldsymbol{y})}{\partial \boldsymbol{y}} \right|_{\boldsymbol{y}=\boldsymbol{y}^{(i)-th}}$

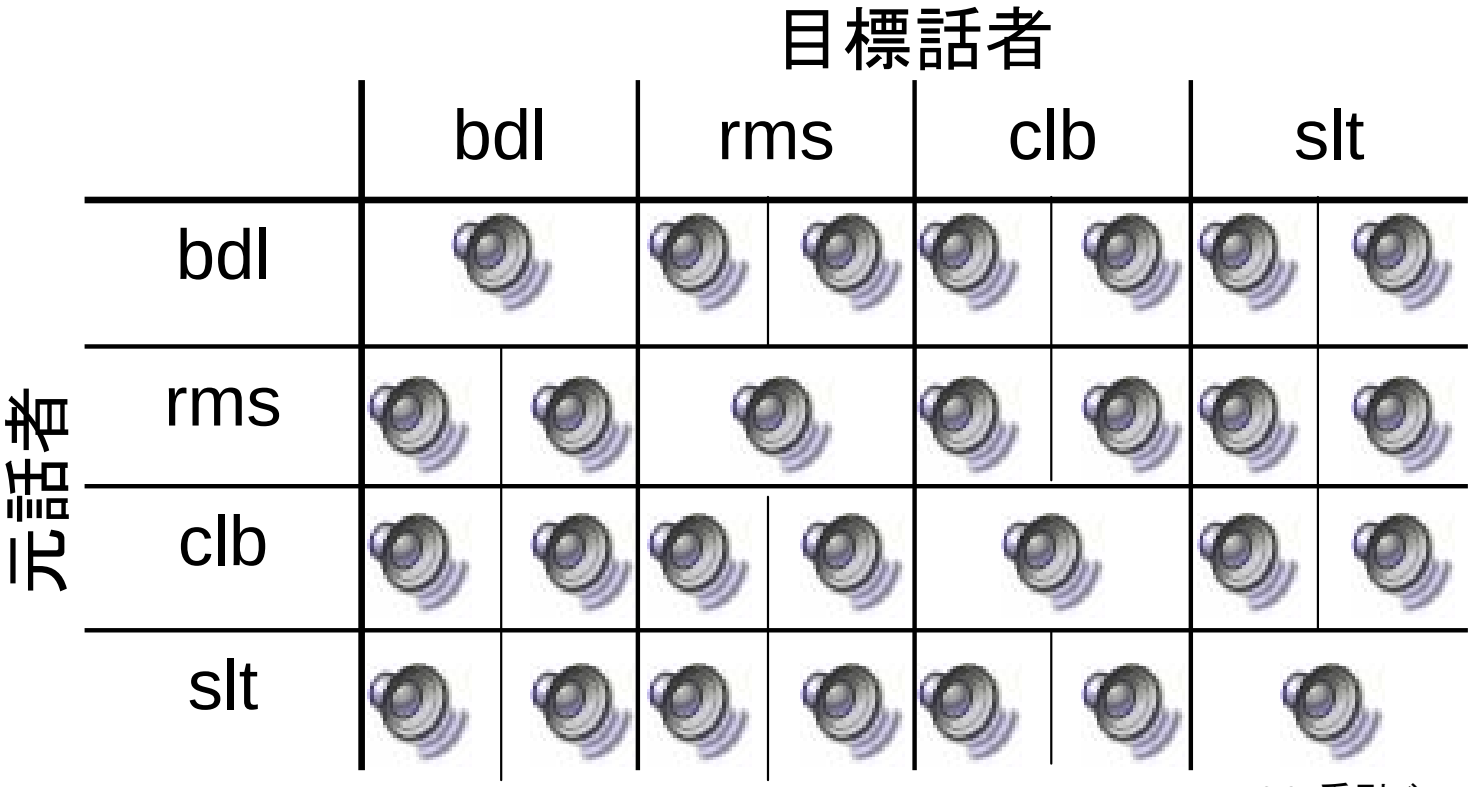
GVを考慮する効果



(a) 目標スペクトル
(b) 変換スペクトル(GV未使用)
(c) 変換スペクトル(GV使用)

話者変換音声サンプル

- 話者：4名（男：bdl, 男：rms, 女：clb, 女：slt）
- 学習：50文対（完全自動学習）
- 変換法
  - 左側：最小平均自乗誤差推定法（フレームベース変換法）
  - 右側：GVありの最尤系列変換法



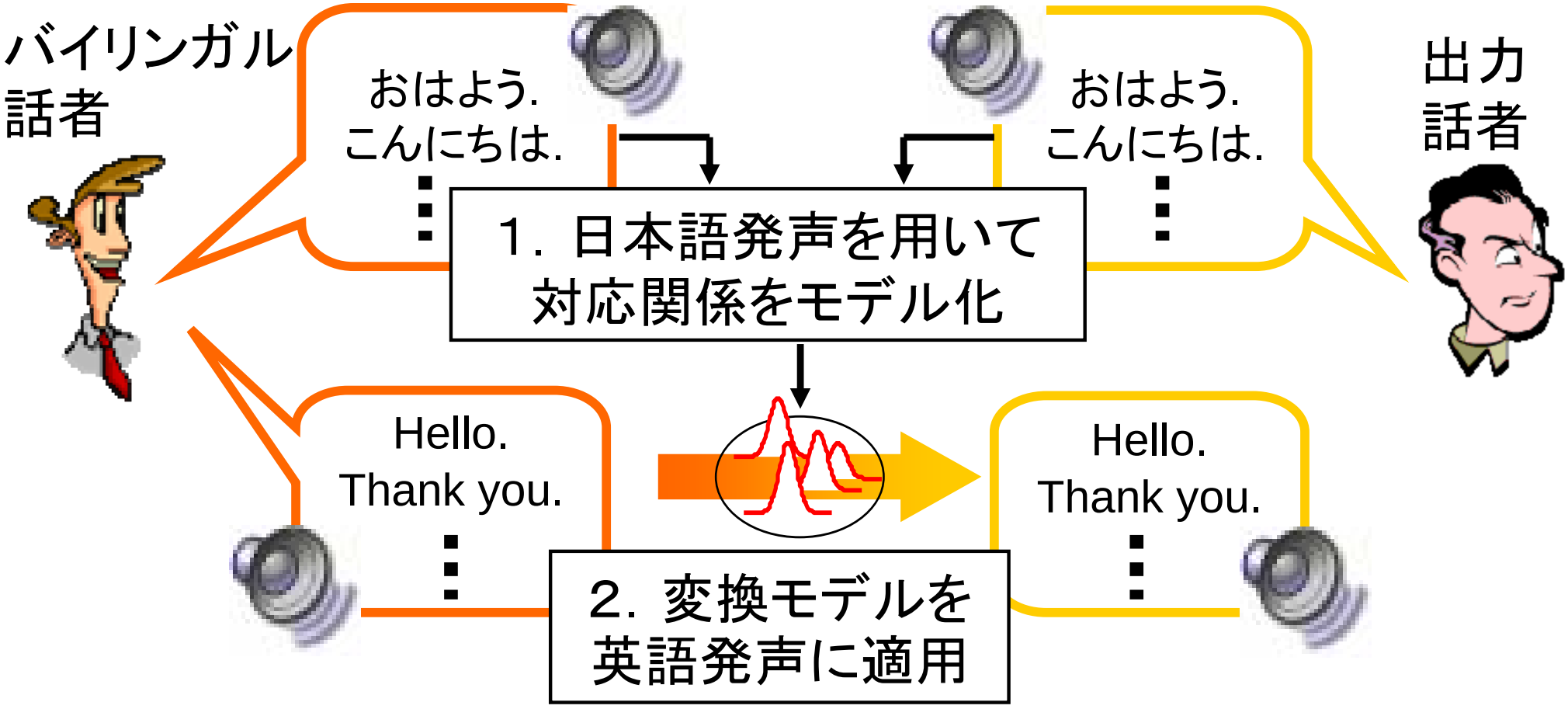
内容

1. 音声変換のしくみ
2. 統計的手法による声質変換
   2.1. 基本的な枠組み
   2.2. フレームベース変換法
   2.3. 系列ベース変換法
3. 応用例

1. 異なる言語間における話者変換

[Abe *et al.*]

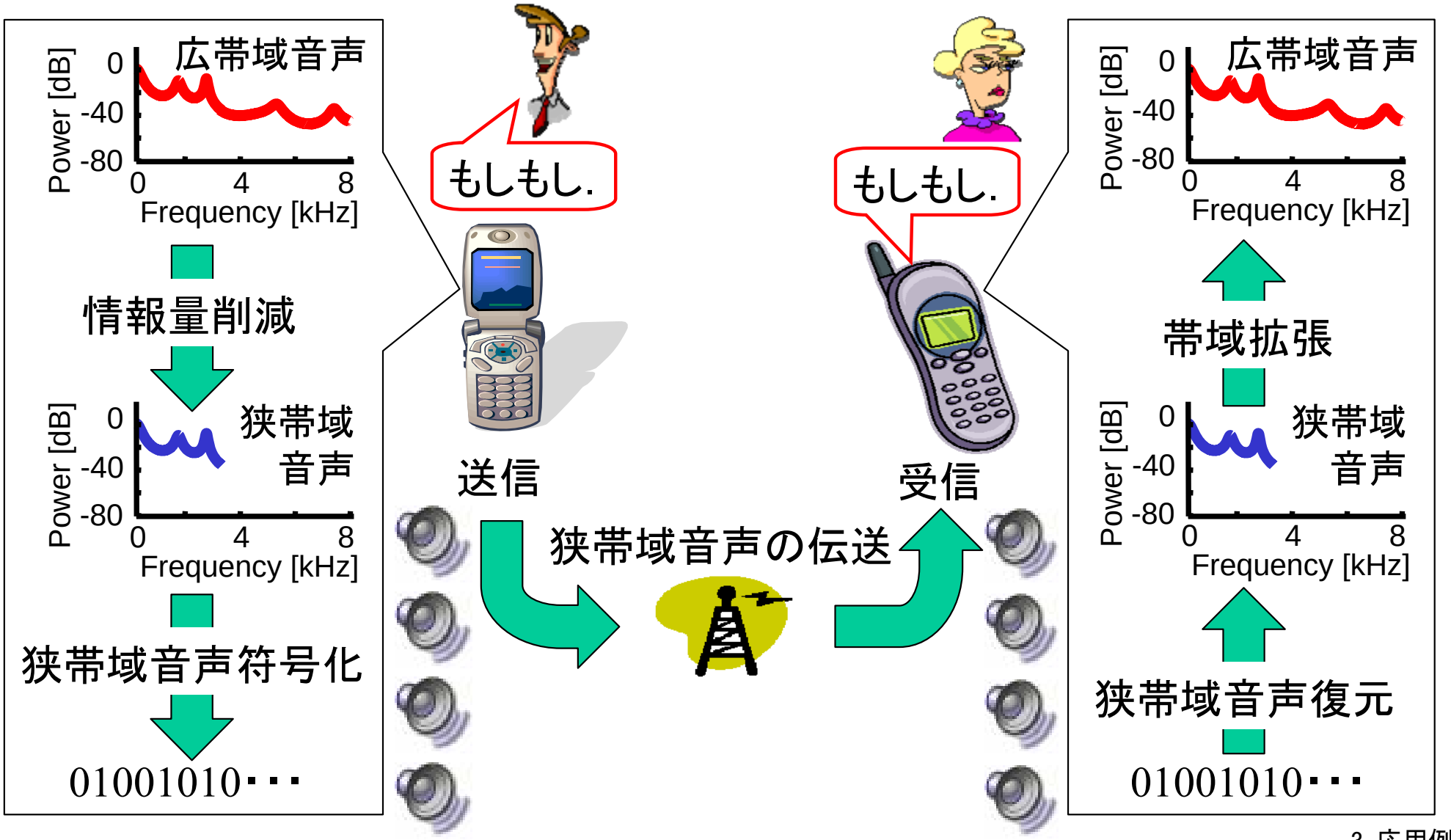
- 自身の声で他の言語を発声することができる.
- 自動音声翻訳，映画の吹き替え，外国語学習に有用である.



2. 携帯電話音声の帯域拡張

[Jax and Vary]

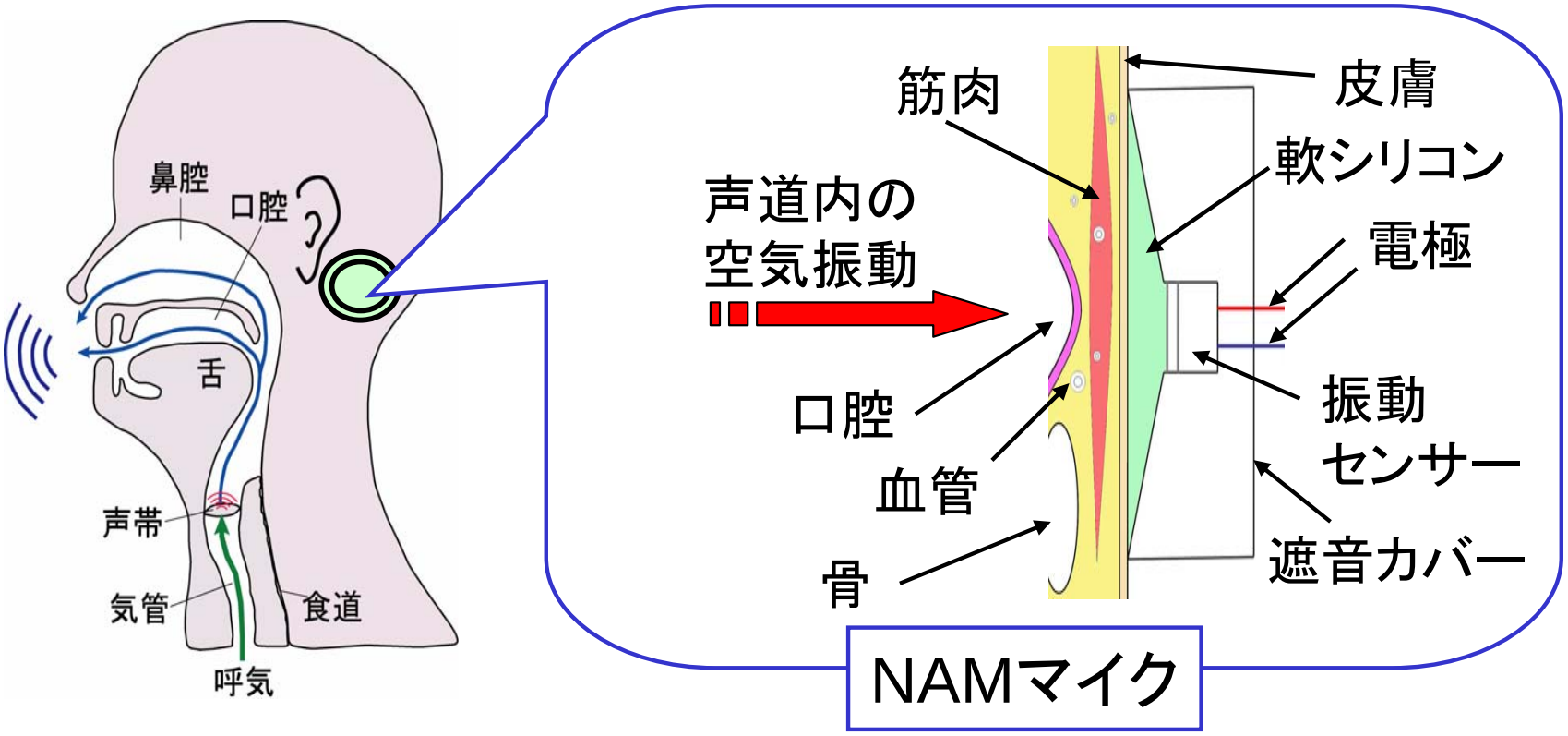
- 従来の伝送量で高品質な広帯域音声を受聴できる.



3. 肉伝導音声コミュニケーション

[中島 他]

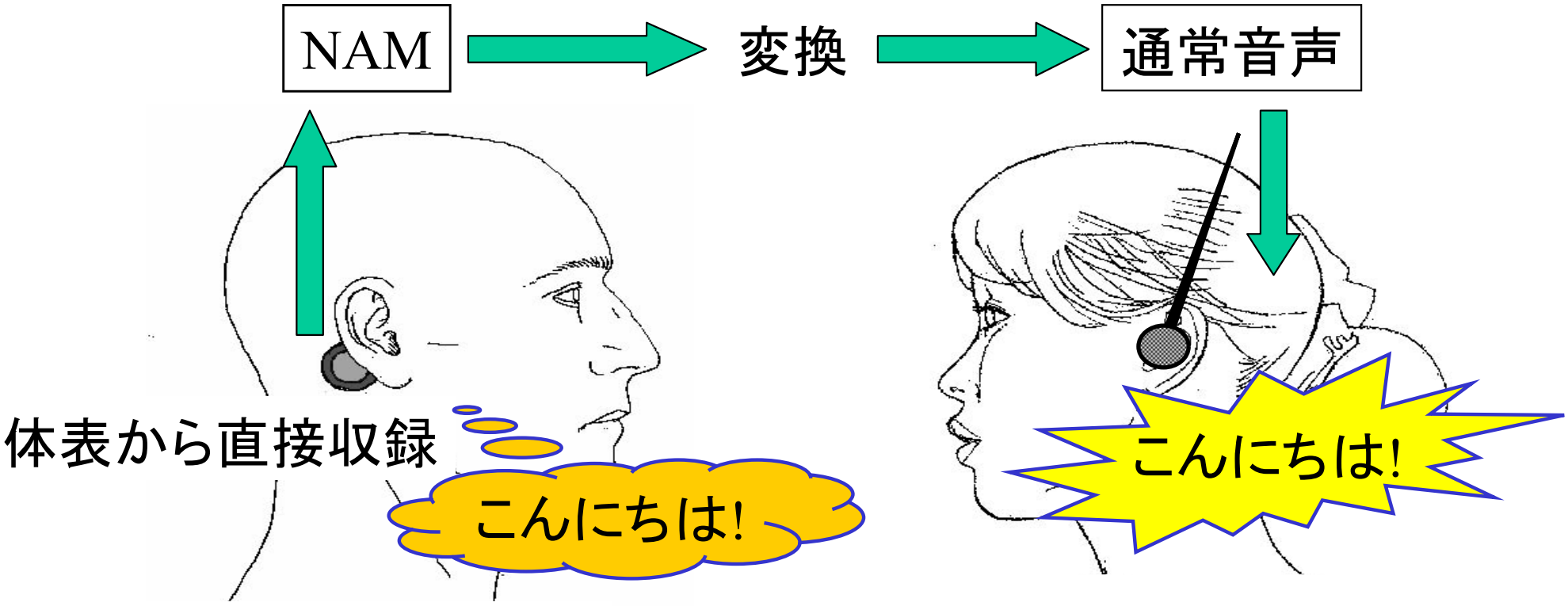
- NAMマイクを用いることで，軟組織を伝わる音（肉伝導音）を体表から収録し，それを音声コミュニケーションに用いる．



※非可聴つぶやき（NAM: Non-Audible Murmur）：
周りに聞こえないほど小さなつぶやき声のようなものであり，
呼気が調音されることで生成される無声音である．

応用例:無音声電話

- 第三者には聞こえない音声コミュニケーションを実現する.



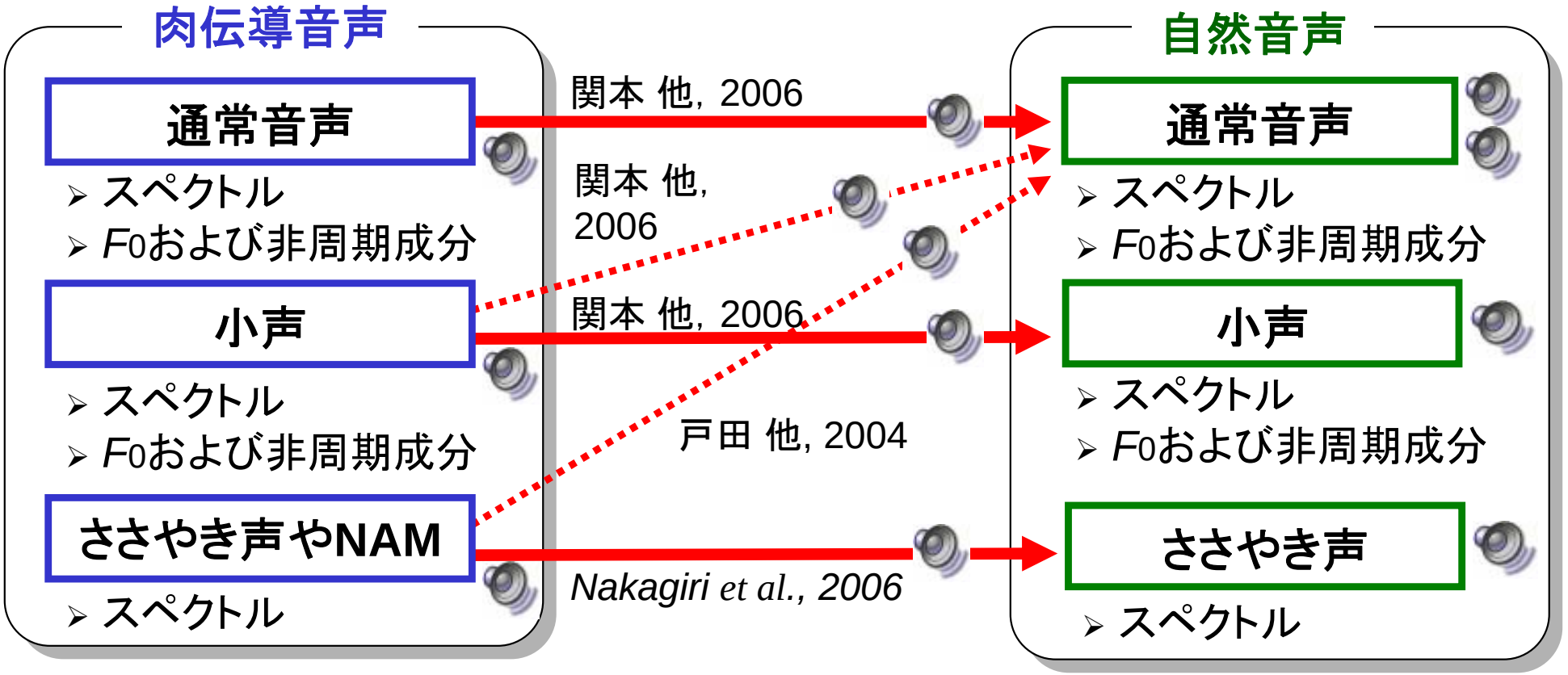
非可聴つぶやき(NAM)による発声

通常音声による受聴

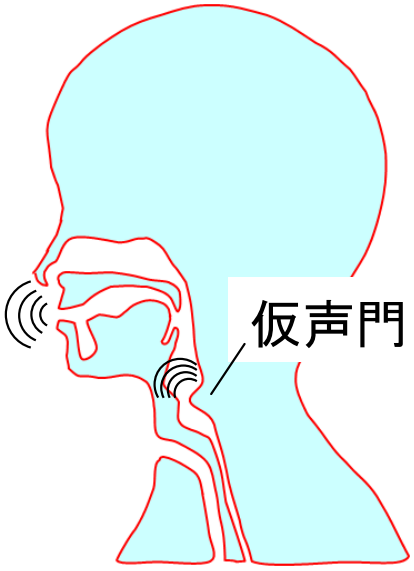
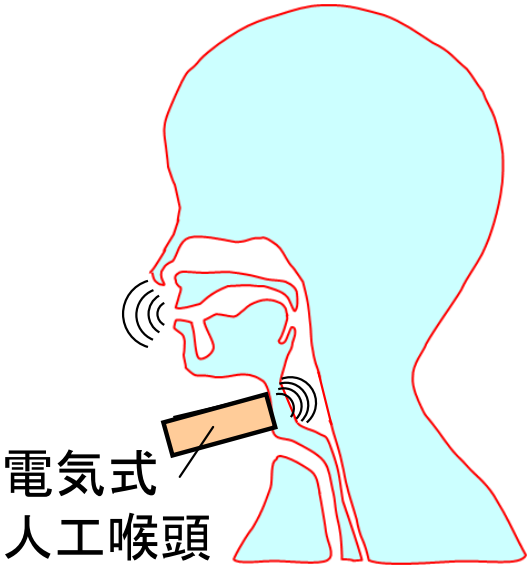
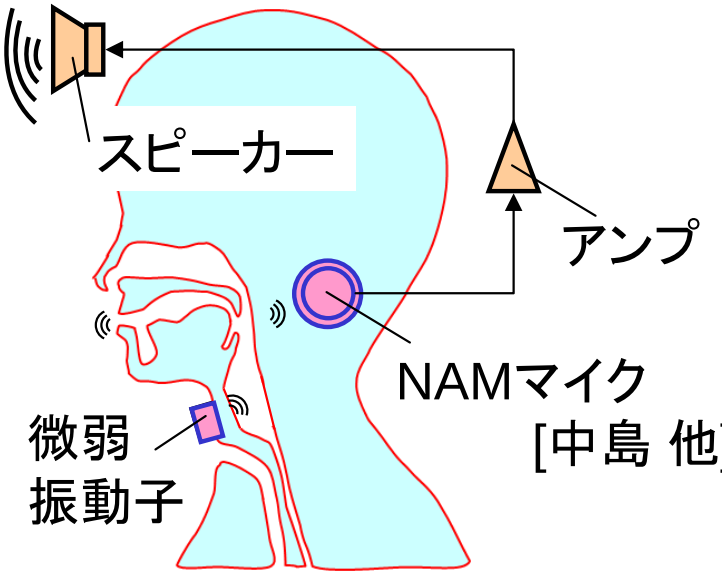
肉伝導音声の変換法

[Toda *et al.*]

- NAMマイクは様々な肉伝導音声を収録することができ，各種肉伝導音声は様々な用途に使用できる.



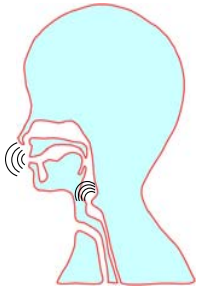
4. 発声障害者補助

	食道発声	電気式人工喉頭を用いた発声 [橋場 他]	微弱振動子及びNAMマイクを用いた発声 [中村 他]
			
補助器具	不要	必要	必要
習得	困難	容易	比較的容易
伝達される音信号	無喉頭音声	無喉頭音声及び人工喉頭の音源信号	スピーカーから提示される無喉頭音声
無喉頭音声	食道音声	電気音声	肉伝導微弱電気音声

無喉頭音声の変換法

食道音声の変換

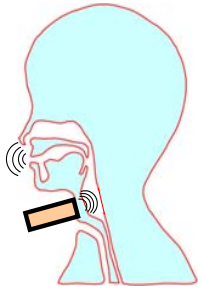
[Doi *et al.*, 2009－2010]



電気音声の変換

[Nakamura *et al.*, 2009－2010]

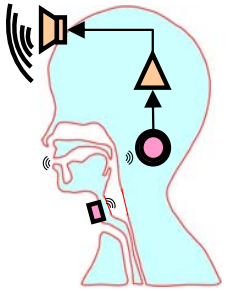
[Doi *et al.*, 2010]



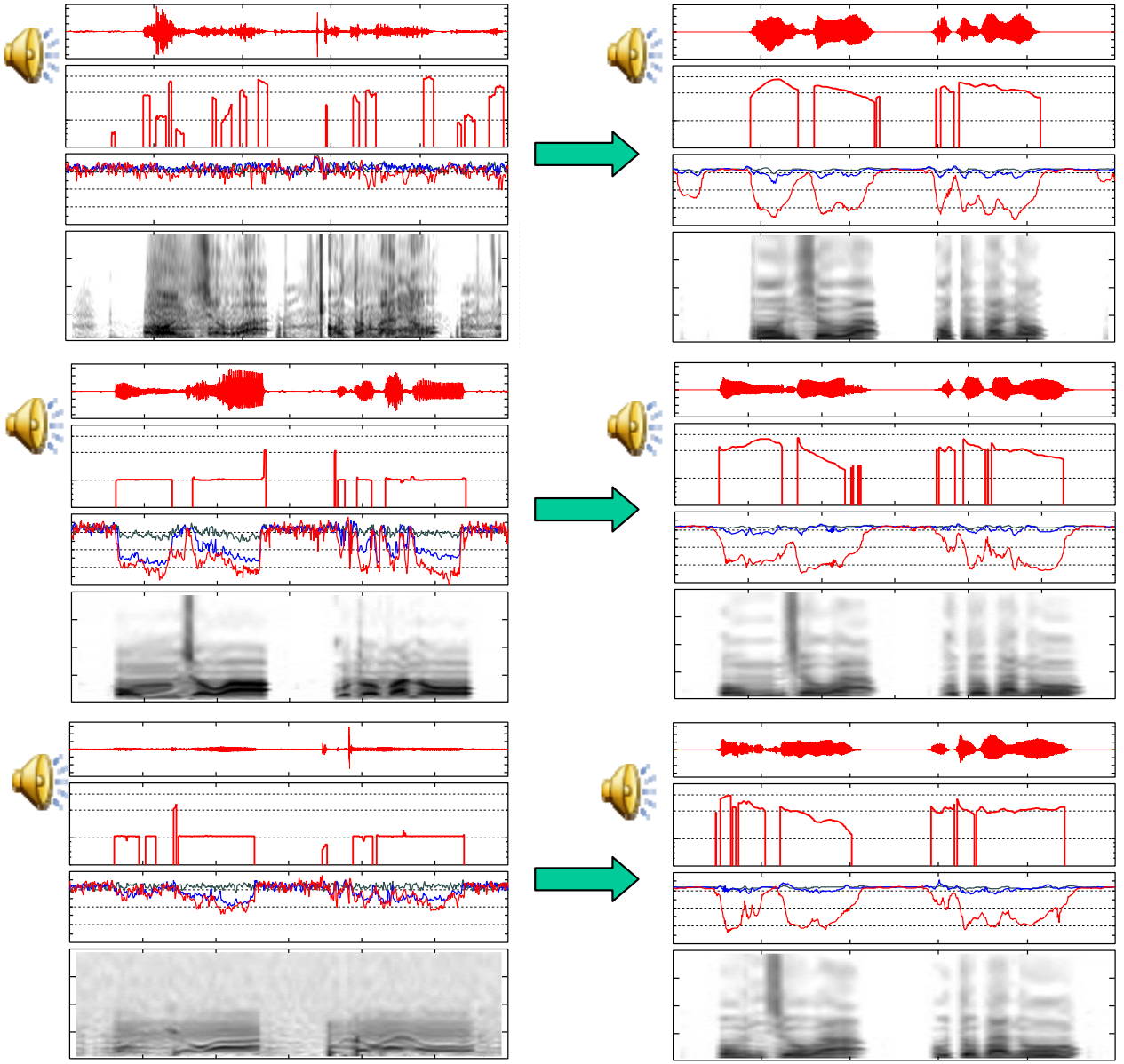
肉伝導微弱電気音声の変換

[Nakamura *et al.*, 2007－2010]

[Doi *et al.*, 2010]



目標話者：



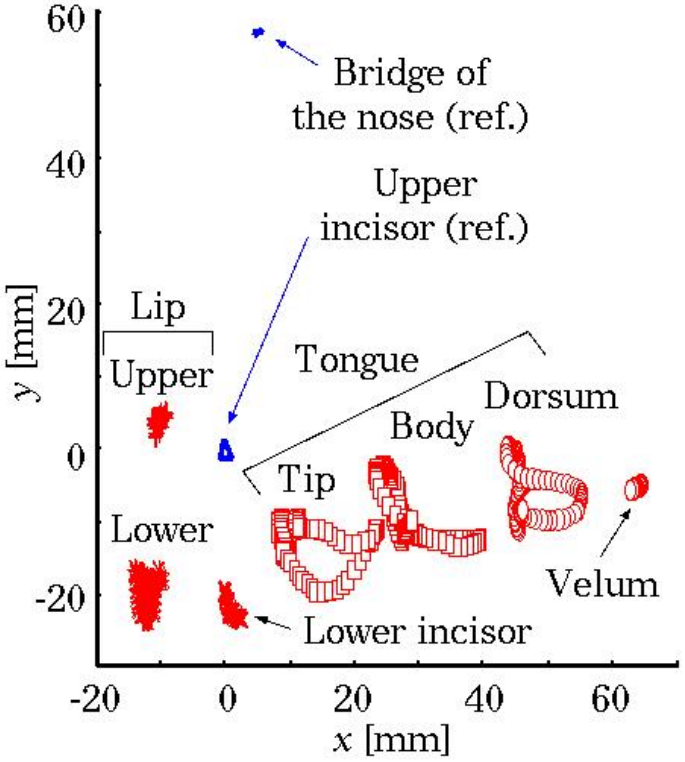
5. 調音運動制御による音声変換

[Toda *et al.*]

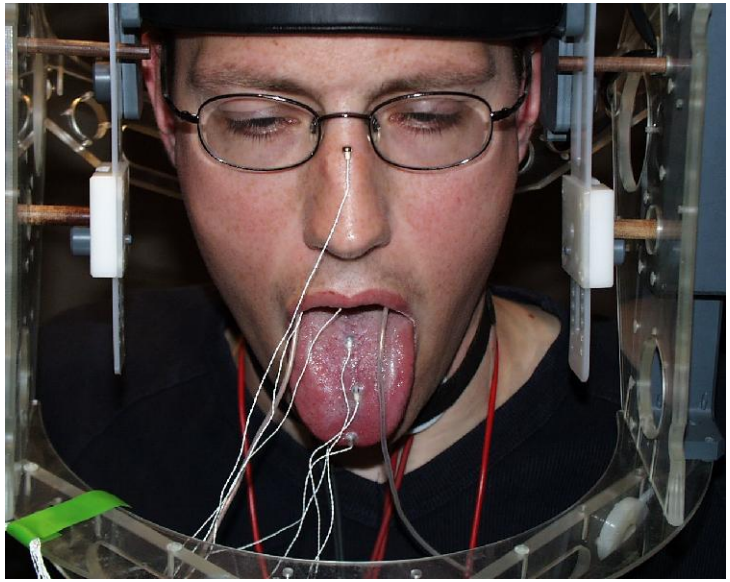
- 調音運動制御による音声変換を実現する.

手順1. 調音運動と音声信号を同時に収録する.

手順2. 調音運動と音声特徴量の対応関係をモデル化する.



Electromagnetic articulograph (EMA)データ

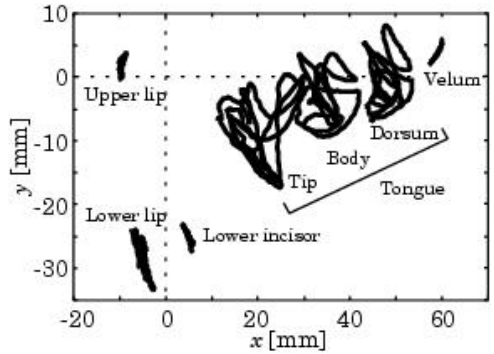


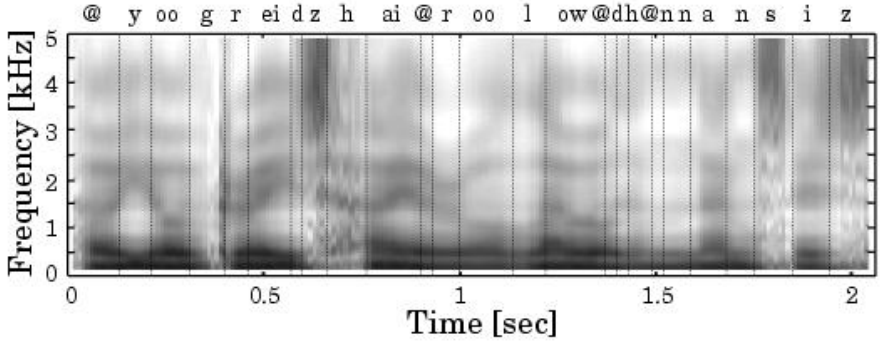
MOCHAデータベース
Edinburgh大から公開

http://www.cstr.ed.ac.uk/research/projects/artic/mocha.html

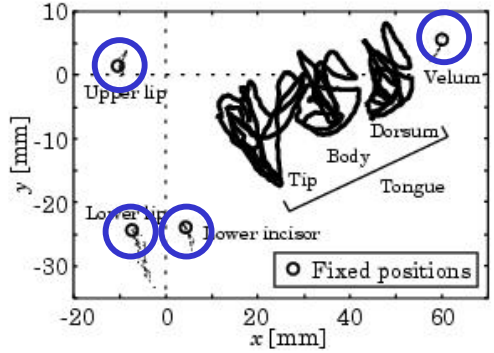
調音運動制御による音声変換の一例

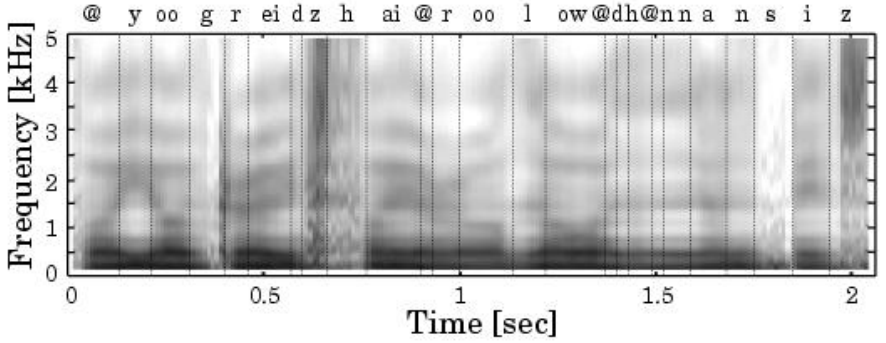
1. 自然な調音運動からの合成音



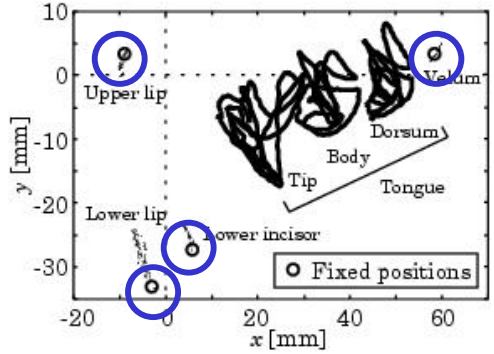


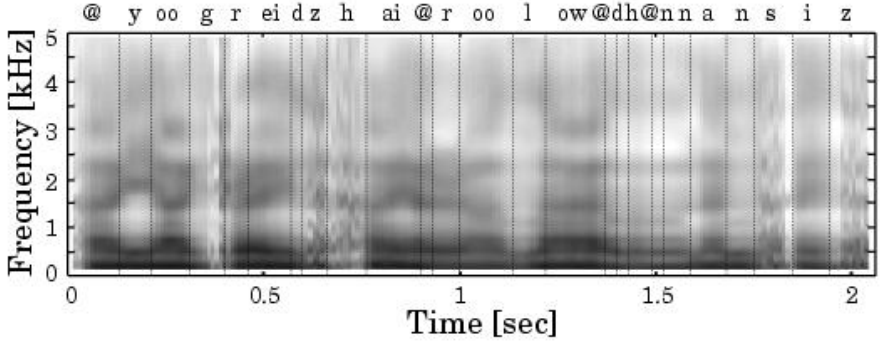
2. 口をすぼめたままの合成音





3. 口を開けたままの合成音





参考文献

- M. Abe, S. Nakamura, K. Shikano, H. Kuwabara, “Voice conversion through vector quantization,” *J. Acoust. Soc. Jpn. (E)*, vol. 11, no. 2, pp. 71–76, 1990.
- 中村哲，鹿野清宏，“ファジィベクトル量子化を用いたスペクトログラムの正規化,” *日本音響学会誌*, vol. 45, no. 2, pp. 107－114, 1989.
- H. Matsumoto and Y. Yamashita, “Unsupervised speaker adaptation from short utterances based on a minimized fuzzy objective function,” *J. Acoust. Soc. Jpn. (E)*, vol. 14, no. 5, pp. 353－361, 1993.
- H. Valbret, E. Moulines, and J. P. Tubach, “Voice transformation using PSOLA technique,” *Speech Communication*, vol. 11, no. 2－3, pp. 175－187, 1992.
- Y. Stylianou, O. Cappe, E. Moulines, “Continuous probabilistic transform for voice conversion,” *IEEE Trans. Speech Audio Process.*, vol. 6, no. 2, pp. 131–142, 1998.
- A. Kain, M. W. Macon, “Spectral voice conversion for text-to-speech synthesis,” *Proc. ICASSP*, Seattle, USA, pp. 285–288, May 1998.
- T. Toda, A.W. Black, K. Tokuda, “Voice conversion based on maximum likelihood estimation of spectral parameter trajectory,” *IEEE Trans. Audio, Speech and Language Process.*, vol. 15, no. 8, pp. 2222–2235, 2007.
- Heiga Zen, Keiichi Tokuda, Alan W. Black, “Statistical parametric speech synthesis,” *Speech Communication*, vol.51, no.11, pp.1039–1154, 2009.
- K. Tokuda, T. Yoshimura, T. Masuko, T. Kobayashi, and T. Kitamura, “Speech parameter generation algorithms for HMM-based speech synthesis,” *Proc. ICASSP*, pp. 1315–1318, Istanbul, Turkey, June 2000.
- H. Zen, K. Tokuda, and T. Kitamura, “Reformulating the HMM as a trajetory model by imposing explicit relationships between static and dynamic feature vector sequences,” *Computer Speech and Language*, vol. 21, no. 1, pp. 153–173, 2007.
- 徳田恵一, 益子貴史, 小林隆夫, 今井 聖, “動的特徴を用いた HMMからの音声パラメータ生成アルゴリズム,” 日本音響学会誌, vol.53, no.3, pp.192–200, Mar. 1997.
- T. Muramatsu, Y. Ohtani, T. Toda, H. Saruwatari, K. Shikano, “Low-delay voice conversion based on maximum likelihood estimation of spectral parameter trajectory,” *Proc. INTERSPEECH*, pp. 1076–1079, Brisbane, Australia, Sep. 2008.

参考文献

- M. Abe, K. Shikano, and H. Kuwabara, “Statistical analysis of bilingual speaker’s speech for cross-language voice conversion,” *J. Acoust. Soc. Am.*, vol. 90, no. 1, pp. 76–82, 1991.
- M. Mashimo, T. Toda, H. Kawanami, K. Shikano, N. Campbell, “Cross-language voice conversion evaluation using bilingual databases,” *IPSJ Journal*, vol. 43, no. 7, pp. 2177–2185, 2002.
- P. Jax, P. Vary, “On artificial bandwidth extension of telephone speech,” *Signal Processing*, vol. 83, pp. 1707–1719, 2003.
- 中島淑貴，柏岡秀紀，ニック キャンベル，鹿野清宏，“非可聴つぶやき認識，” *電子情報通信学会論文誌*, vol. J87-D-II, no. 9, pp. 1757–1764, 2004.
- T. Toda, K. Nakamura, H. Sekimoto, K. Shikano, “Voice conversion for various types of body transmitted speech,” *Proc. ICASSP*, pp. 3601–3604, Taipei, Taiwan, Apr. 2009.
- M. Nakagiri, T. Toda, H. Kashioka, and K. Shikano, “Improving body transmitted unvoiced speech with statistical voice conversion,” *Proc. INTERSPEECH*, pp. 2270—2273, Pittsburgh, USA, Sep. 2006.
- 中村圭吾，戸田智基，猿渡洋，鹿野清宏，“肉伝導人工音声の変換に基づく喉頭全摘出者のための音声コミュニケーション支援システム，” *電子情報通信学会論文誌*, vol. J90-D, no. 3, pp. 780–787, 2007.
- H. Doi, K. Nakamura, T. Toda, H. Saruwatari, and K. Shikano, “Esophageal speech enhancement based on statistical voice conversion with Gaussian mixture models,” *IEICE Trans. on Inf. and Syst.*, vol. E93-D, no. 9, pp. 2472—2482, 2010.
- T. Toda, A.W. Black, K. Tokuda, “Mapping between articulatory movements and acoustic spectrum with a Gaussian mixture model, ” *Speech Communication*, vol. 50, no. 3, pp. 215–227, Mar. 2008.