

音声処理と部分空間法

Subspace method in speech processing

神戸大学 システム情報学研究科
情報科学専攻

有木康雄

問題設定

電話がかかってきて話をしている場合、

(1) 誰からの電話なのか 話者情報

(2) どのような内容で 音韻情報

(3) 先方はどんな感情を持っているのか 感性情報

(4) 雑音がかなり重畳していても 雑音

人は容易に、これらの情報を聞き分けることができる。

研究目的

音声信号から、各情報と雑音を分離抽出し、認識する

➤ 話者認識・話者照合

➤ 話者適応

➤ 感情認識

➤ 音声認識

➤ 音声強調・特徴抽出



• 大語彙連続音声認識

• 音声情報検索

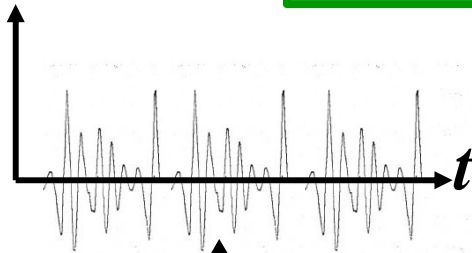
• 音声対話

音声生成モデル

$$S(\omega) = G(\omega) \cdot H(\omega) \cdot R(\omega)$$

音声信号

$S(\omega)$

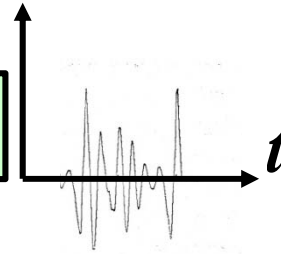


放射 $R(\omega)$

声道共振

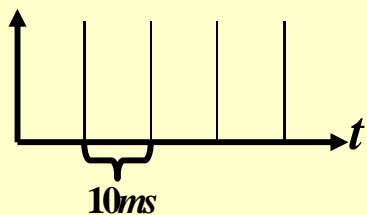
$H(\omega)$

インパルス応答



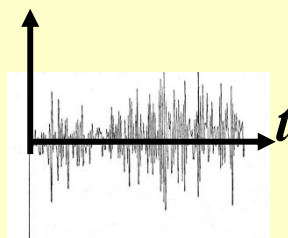
駆動音源 $G(\omega)$

有声音

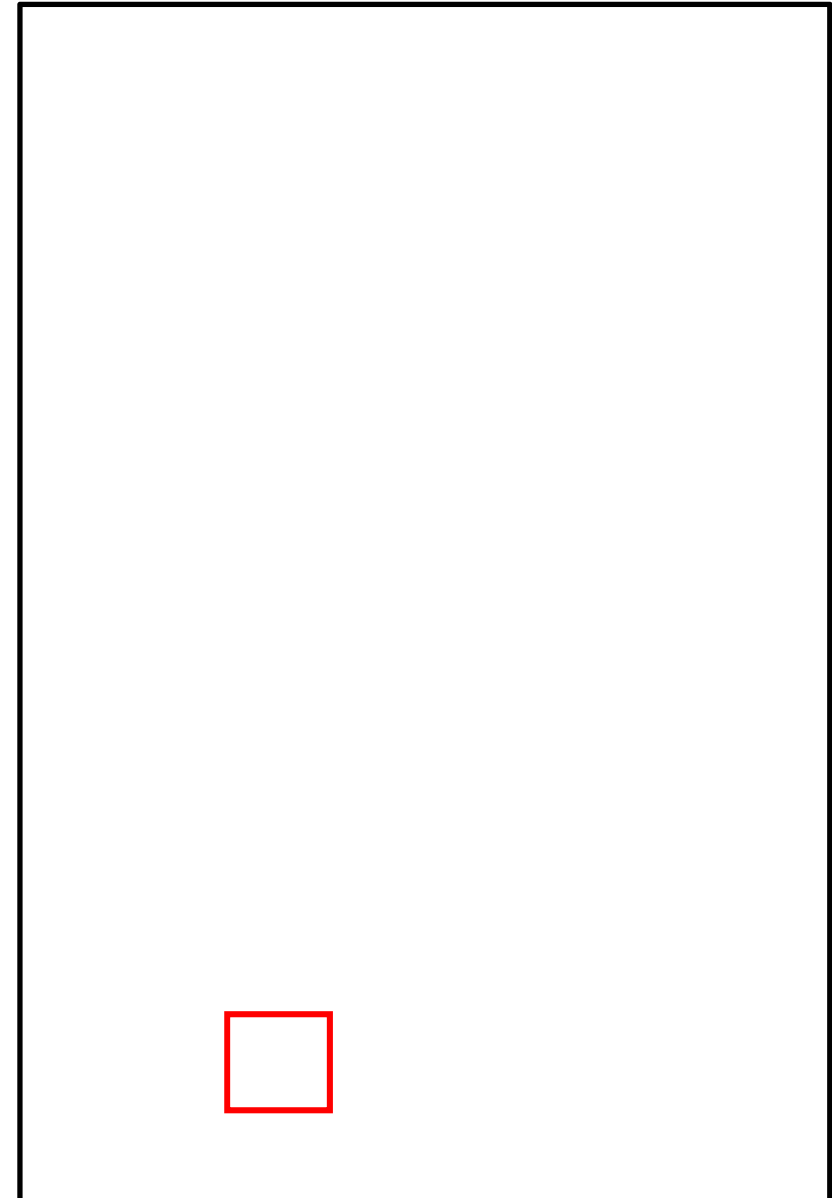


インパルス列

無声音



雑音



音声器官の部位と名称

声道での共振周波数

波の波長 : λ (m)

声道長 : $\frac{2k-1}{4} \cdot \lambda = 0.17$ (m)

波の周波数 : f (Hz)

波の速度 : $v = \lambda \cdot f = 340$ (m/sec)

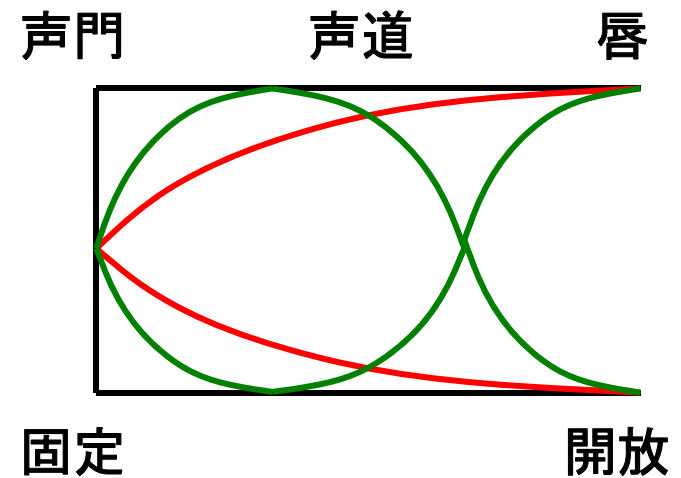
$$\frac{2k-1}{4} \cdot \frac{340}{f} = 0.17$$

$$\therefore f = \frac{340}{4 \times 0.17} \cdot (2k-1) = 500 \cdot (2k-1) \text{ (Hz)}$$

$k=1$ 第1フォルマント 500Hz
(第1共振周波数)

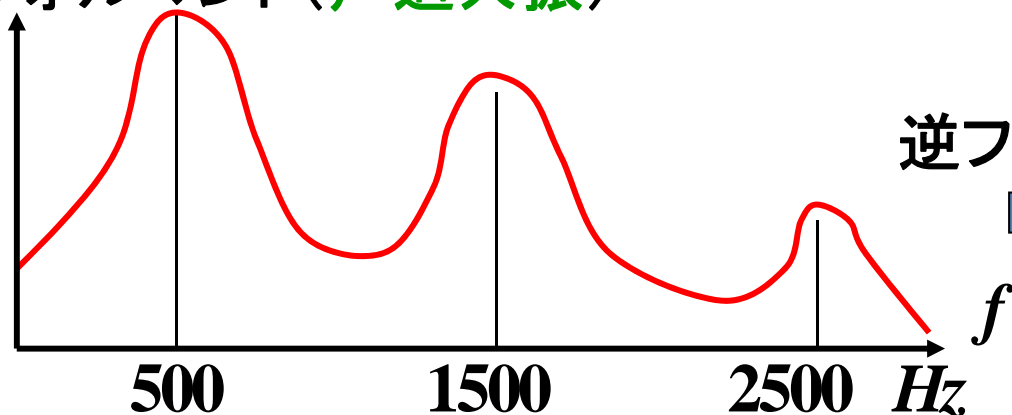
$k=2$ 第2フォルマント 1500Hz

$k=3$ 第3フォルマント 2500Hz



音声信号の特徴

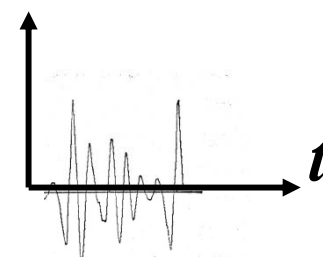
フォルマント(声道共振)



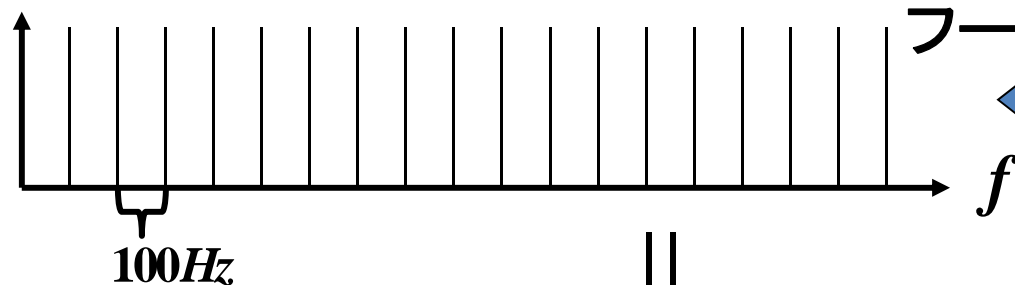
逆フーリエ変換



インパルス応答



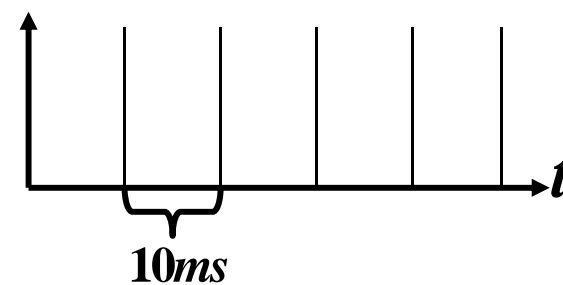
基本周波数(声帯振動) ×



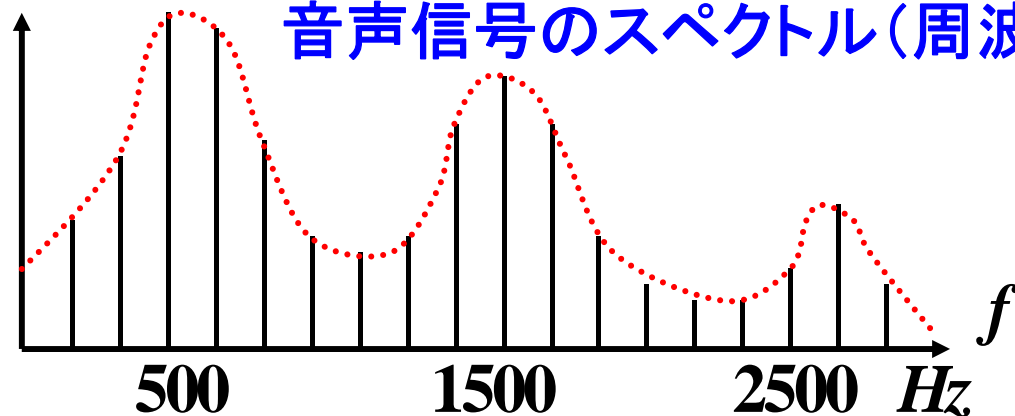
フーリエ変換



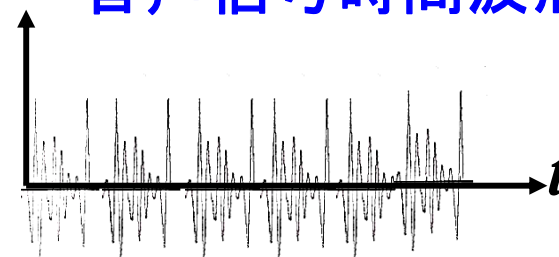
声帯振動



音声信号のスペクトル(周波数特性)



音声信号時間波形



音声信号の特徴

時間 →

音声波形

基本周波数

2500Hz

1500Hz

500Hz

声道共振(スペクトル)

音声信号の特徴

音声を特徴づけているのは、声道共振（スペクトル） $H(\omega)$ である

声帯振動と声道共振の畳み込みとして生成される音声から、どのように声道共振のみを取り出すか。

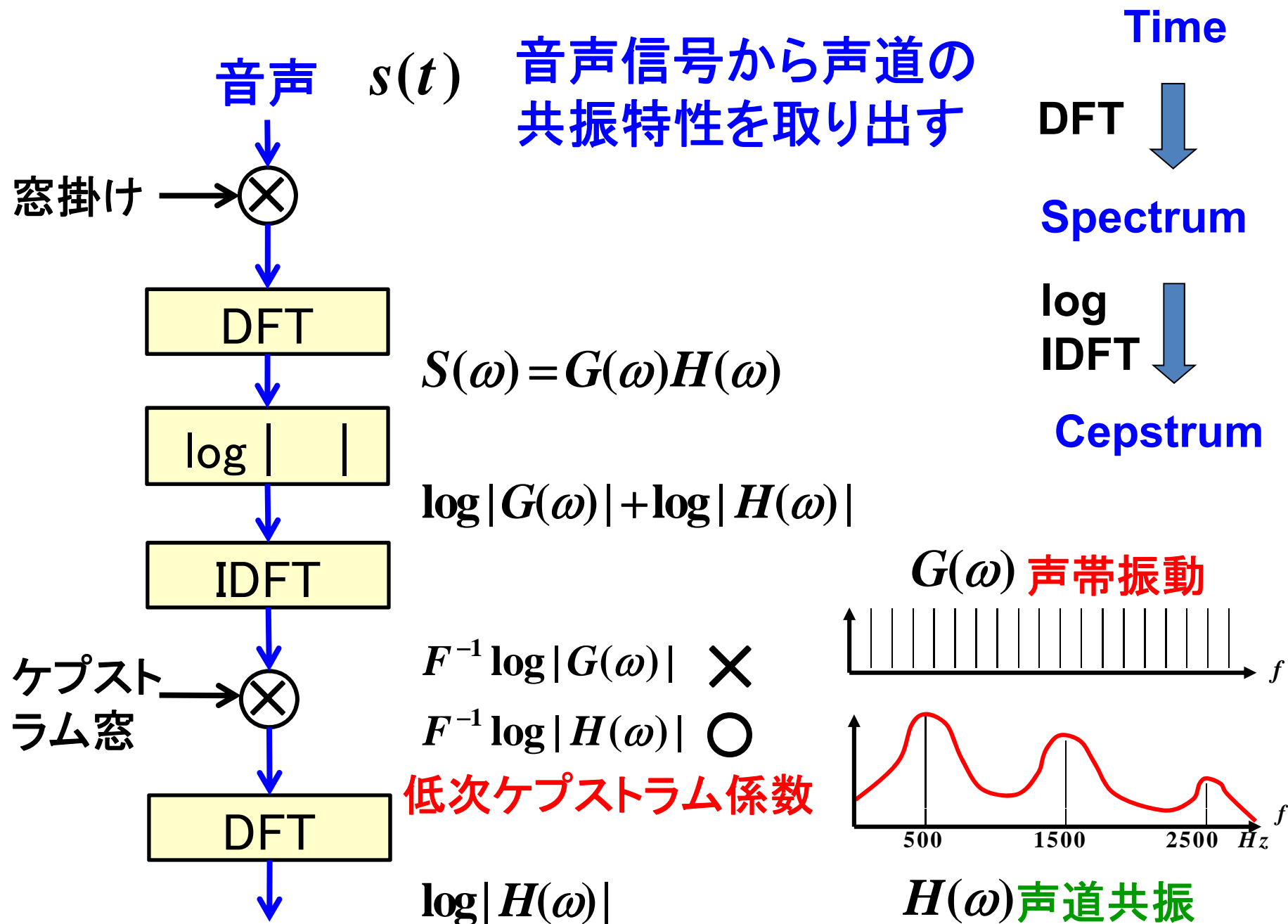
$$S(\omega) = G(\omega) \cdot H(\omega) \cdot R(\omega)$$

女性

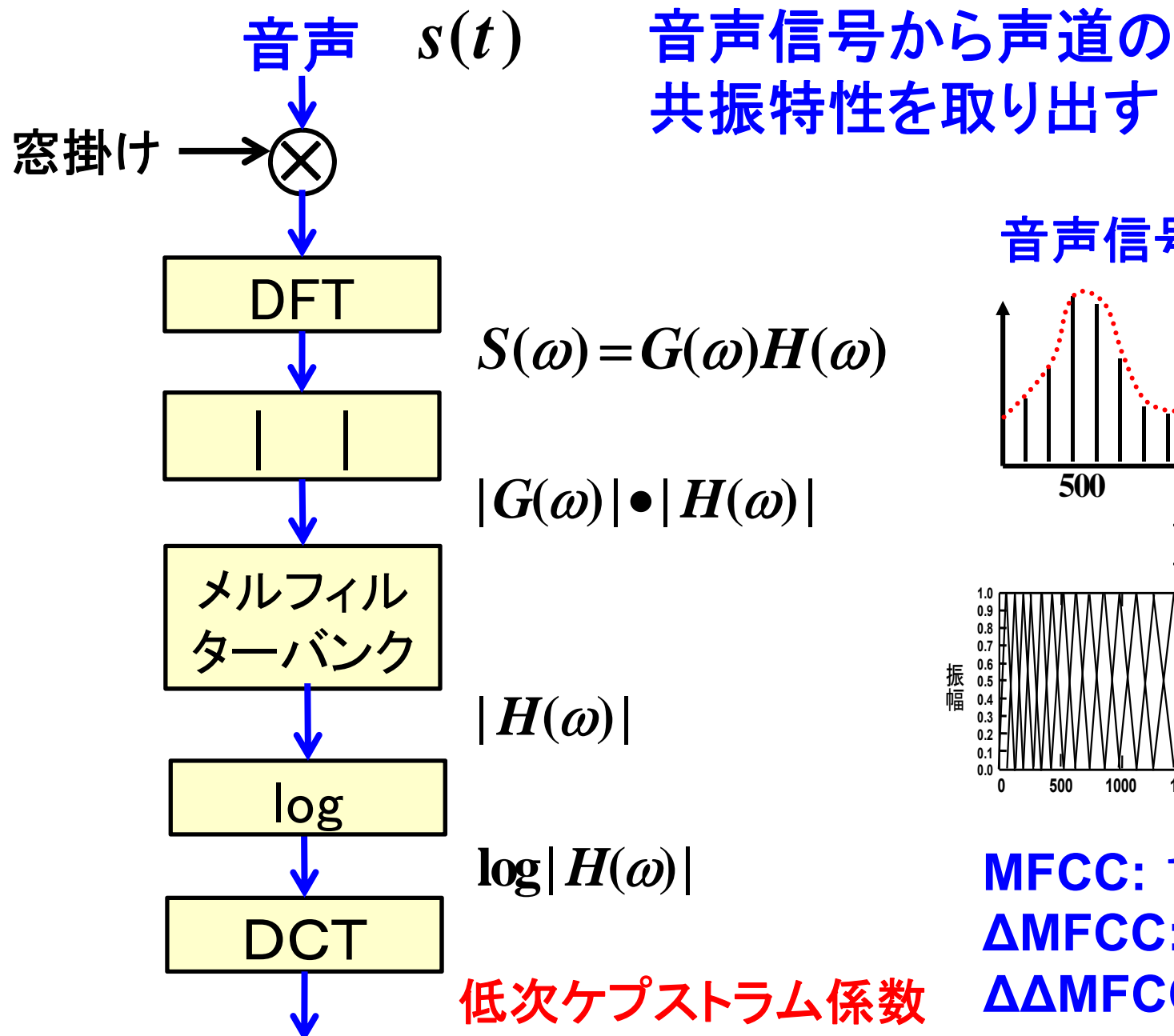
男性

声道長の違い
声道の形の違い

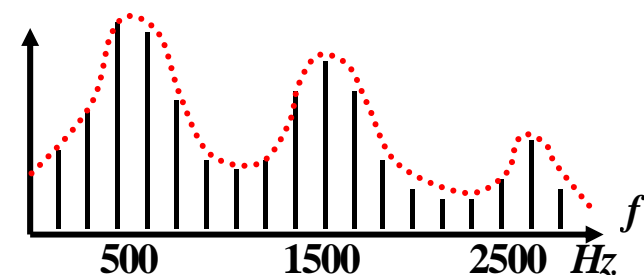
音声信号の特徴 Cepstrum(ケプストラム)



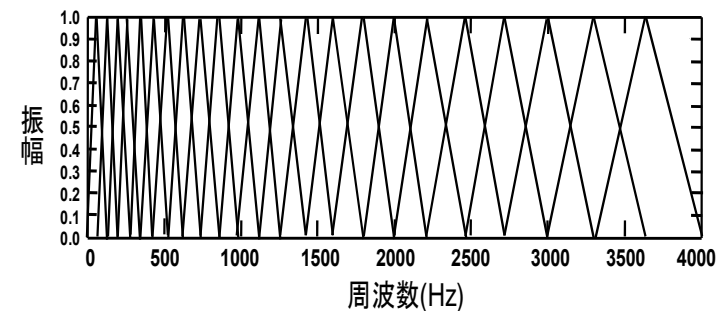
音声信号の特徴 MFCC(メル周波数ケプストラム係数)



音声信号のスペクトル



\times



MFCC: 12次元

Δ MFCC: 12次元

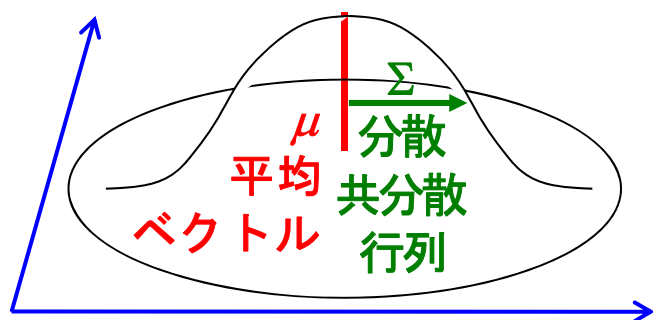
$\Delta\Delta$ MFCC: 12次元

話者認識・話者照合

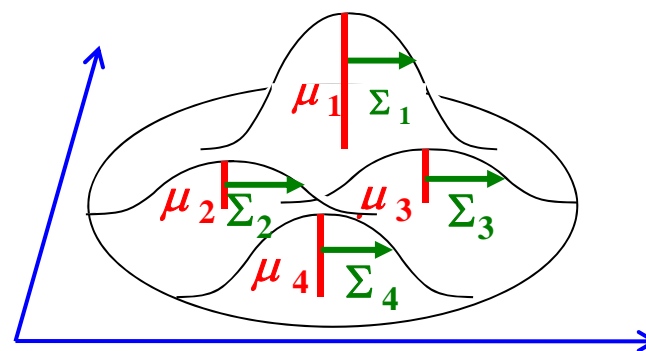
(1) 部分空間法以外

1. GMM (Gaussian Mixture Model: 混合正規分布) [1] 1995

Gaussian Model



GMM



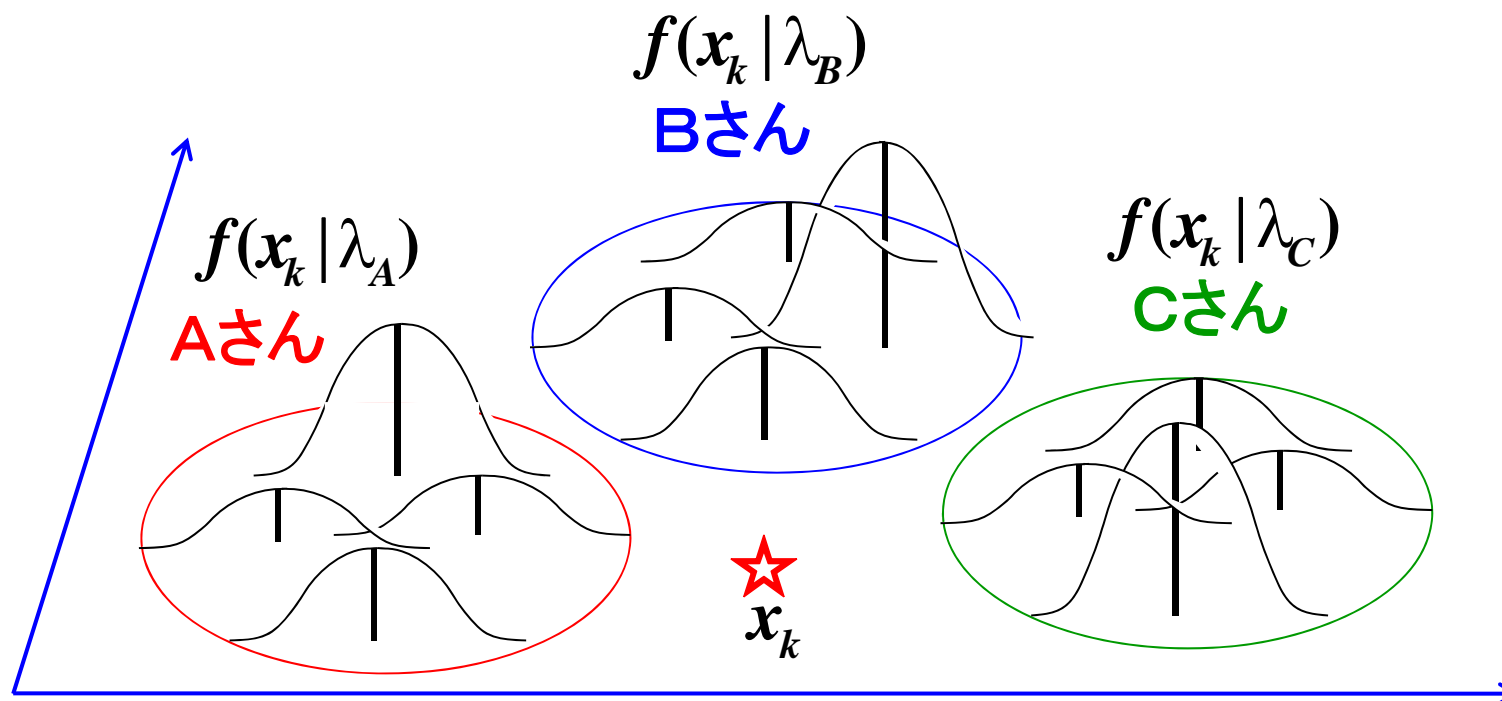
$$f(x_k) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (x_k - \mu)^T \Sigma^{-1} (x_k - \mu) \right]$$
$$f(x_k | \lambda) = \sum_i \omega_i f_i(x_k)$$

μ_i : 平均ベクトル

Σ_i : 分散共分散行列

ω_i : 混合係数

これらの推定にはEM
(Expectation-Maximization)
アルゴリズムを実行する



話者照合 $\frac{f(x_k | \lambda_r)}{f_{UBG}(x_k | \lambda)} \geq \theta_1$

話者認識 $\hat{r} = \arg \max_r f(x_k | \lambda_r)$

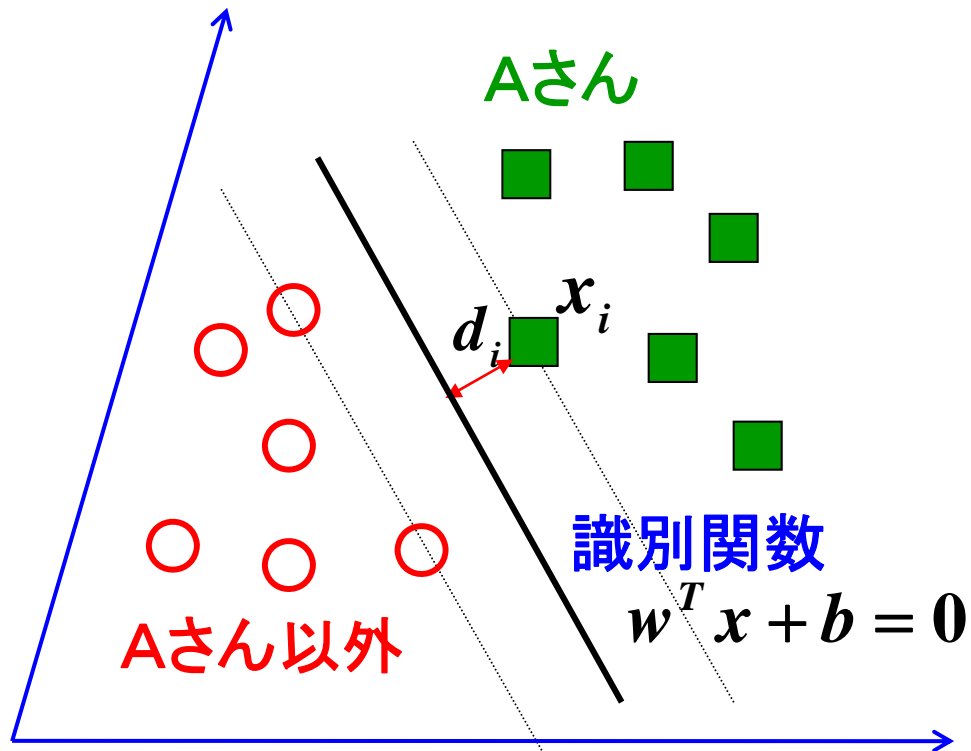
$$\frac{f(x_k | \lambda_{\hat{r}})}{f_{UBG}(x_k | \lambda)} \geq \theta_2$$

事後確率

$$f(r | x_k) = \frac{f(x_k | r)f(r)}{f(x_k)}$$

$$= \frac{f(x_k | \lambda_r)}{f_{UBG}(x_k | \lambda)}$$

2. SVM [2] 1996 テストデータに対する汎化能力が高い



マージン:

学習データと超平面との最小距離

$$\min_{i=1,\dots,n} d_i = \min_{i=1,\dots,n} \frac{|w^T x_i + b|}{\|w\|}$$

マージン最大化: w と b の決定

$$\max_{w,b} \min_{i=1,\dots,n} \frac{|w^T x_i + b|}{\|w\|}$$

目的関数: 最小化

$$g(w, b) = \frac{1}{2} \|w\|^2$$

制約: $\min_{i=1,\dots,n} |w^T x_i + b| = 1$

$$y_i (w^T x_i + b) \geq 1$$

$$y_i = \{+1, -1\}$$

ラグランジェ関数:

$$L_a(w, b, \lambda) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \lambda_i \{y_i (w^T x_i + b) - 1\}$$

w と b に関して最小化、 λ に関して最大化
制約条件付き非線形最適化問題として解く

$$w^* = \sum_{i=1}^n \lambda_i^* y_i x_i \quad b^* = y_s - w^{*T} x_s$$

x_s : サポートベクター

線形識別関数 $f(x)$:

$$f(x) = w^{*T} x + b^* = \sum_{i=1}^n \lambda_i^* y_i x_i^T x + b^*$$

線形識別関数 $f(x)$:

$$f(x) = w^{*T} x + b^* = \sum_{i=1}^n \lambda_i^* y_i x_i^T x + b^*$$

非線形識別関数：非線形変換関数 $\phi(x)$

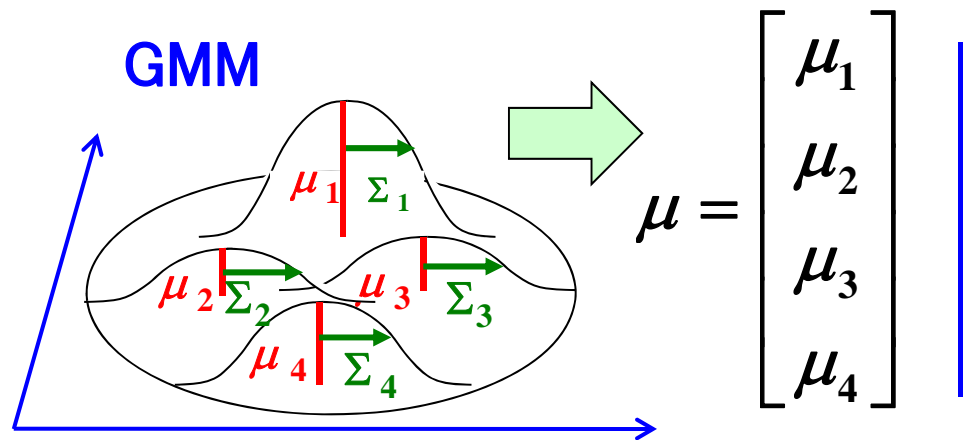
$$\begin{aligned} f(\phi(x)) &= w^{*T} \phi(x) + b^* = \sum_{i=1}^n \lambda_i^* y_i \phi(x_i)^T \phi(x) + b^* \\ &= \sum_{i=1}^n \lambda_i y_i K(x_i, x) + b^* > 0 \Rightarrow \text{Aさん} \end{aligned}$$

カーネル関数： $K(x_i, x) = \phi(x_i)^T \phi(x)$

$K(x, y) = (1 + x^T y)^p$ p 次多項式関数

$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2p^2}\right)$ RBF

3. GMM-SVM [3] 2006

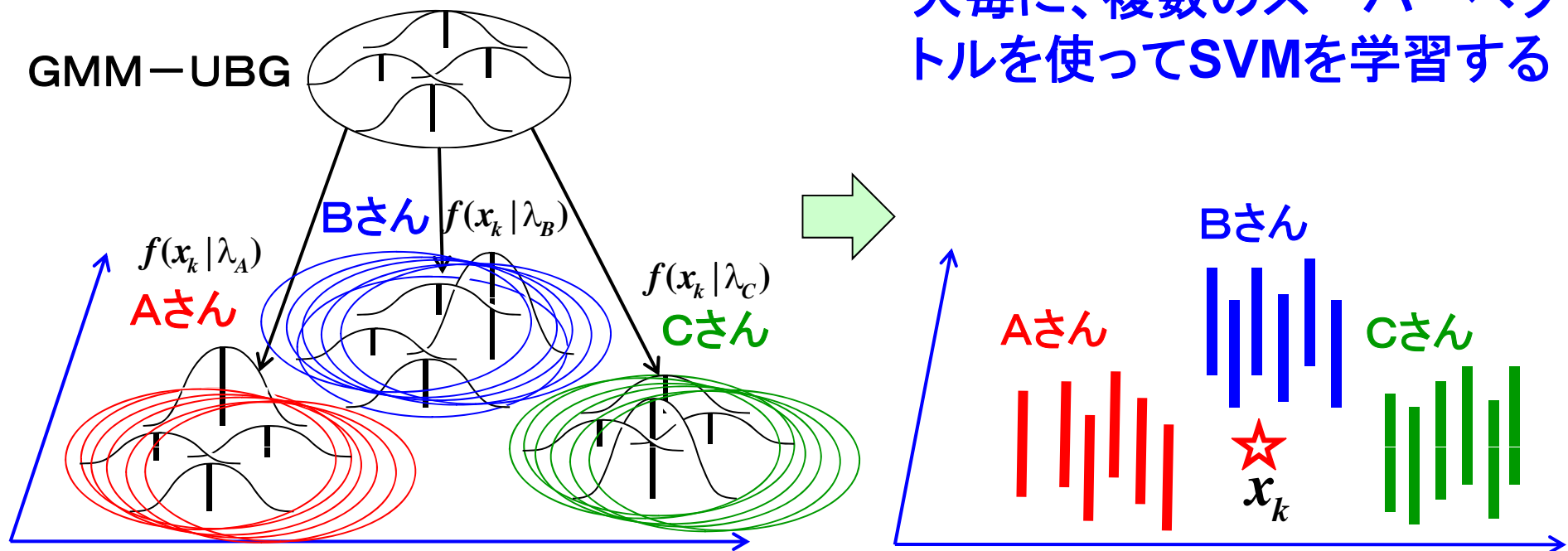


GMMの平均ベクトルを並べ
スーパーベクトル μ を作る

スーパーベクトル μ を使って
SVMを学習し、認識する

人(s)の発話(k)毎に、
GMM-UBGを適応させる

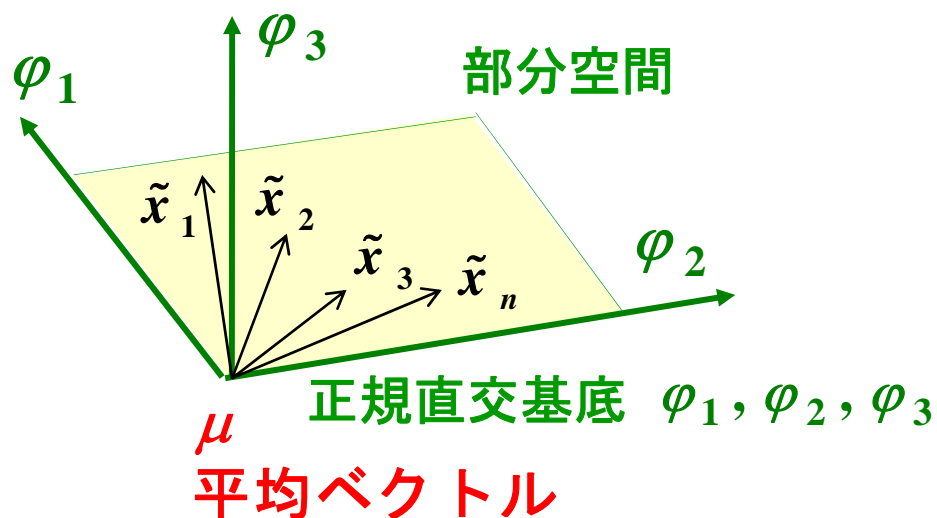
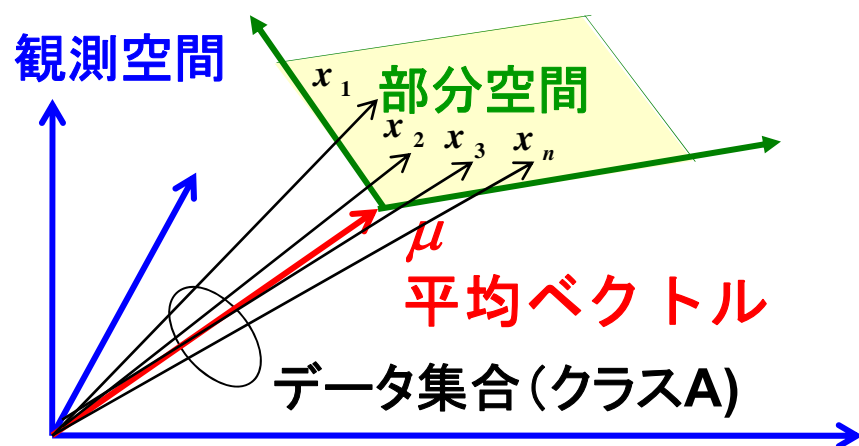
人毎に、複数のスーパーベク
トルを使ってSVMを学習する



話者認識・話者照合

(2) 部分空間法

4. 部分空間法[4] 1994



データ集合だけを表現する
小さな空間＝部分空間

学習:

データ集合からどのようにして、
部分空間の正規直交基底を
求めるか

認識:

テストデータと部分空間の距離
をどのように求め、クラス識別
を行うのか

データ集合からどのようにして、部分空間の正規直交基底を求めるか

学習データ: $X = (x_1, \dots, x_n)$

テストデータ: $Y = (y_1, \dots, y_n)$

平均引算後: $\tilde{X} = (\tilde{x}_1, \dots, \tilde{x}_n)$

平均引算後: $\tilde{Y} = (\tilde{y}_1, \dots, \tilde{y}_n)$

\tilde{X} の特異値分解: $\tilde{X} = V\Sigma^{1/2}U^T \Rightarrow V = \tilde{X}U\Sigma^{-1/2}, V^T = \Sigma^{-1/2}U^T \tilde{X}^T$

分散共分散行列: $C = \tilde{X}\tilde{X}^T = V\Sigma V^T \Rightarrow$

内積行列: $G = \tilde{X}^T \tilde{X} = U\Sigma U^T$

V は、分散最大基準で求めた学習データ X の正規直交基底。固有値の大きい固有ベクトルで構成する。(主成分分析)

(1) テストデータ \tilde{y}_k を部分空間に射影して射影ベクトル $V^T \tilde{y}_k$ を求める

① 分散共分散行列 C を固有値分解して V を求め、 $V^T \tilde{y}_k$ を求める

② 内積行列 G を固有値分解して U と Σ を求め、次式で求める

$$V^T \tilde{y}_k = \Sigma^{-1/2} U^T \tilde{X}^T \tilde{y}_k$$

非線形化：非線形変換関数 $\phi(x)$ [5-8] 1998,1999, 2000

学習データ: $X = (\phi(x_1), \dots, \phi(x_n))$

テストデータ: $Y = (\phi(y_1), \dots, \phi(y_n))$

平均引算後: $\tilde{X} = (\tilde{\phi}(\tilde{x}_1), \dots, \tilde{\phi}(\tilde{x}_n))$

平均引算後: $\tilde{Y} = (\tilde{\phi}(y_1), \dots, \tilde{\phi}(y_n))$

\tilde{X} の特異値分解: $\tilde{X} = V \Sigma^{1/2} U^T \rightarrow V = \tilde{X} U \Sigma^{-1/2}, V^T = \Sigma^{-1/2} U^T \tilde{X}^T$

分散共分散行列: $C = \tilde{X} \tilde{X}^T = V \Sigma V^T \rightarrow$

内積行列: $G = \tilde{X}^T \tilde{X} = U \Sigma U^T$

内積行列: $H = \tilde{X}^T \tilde{Y}$

V は、分散最大基準で求めた
学習データ X の正規直交基底
(カーネル主成分分析)

(1) テストデータ \tilde{y}_k を部分空間に射影して射影ベクトル $V^T \tilde{y}_k$ を求める

~~① 分散共分散行列 C を固有値分解して V を求め、 $V^T \tilde{y}_k$ を求める~~

② 内積行列 G を固有値分解して U と Σ を求め、次式で求める

$$V^T \tilde{y}_k = \Sigma^{-1/2} U^T \tilde{X}^T \tilde{y}_k = \Sigma^{-1/2} U^T H_{\cdot k} \quad (H_{\cdot k} = \tilde{X}^T \tilde{y}_k \text{ で } H = \tilde{X}^T \tilde{Y} \text{ の第 } k \text{ 列})$$

内積行列 G と H は、それぞれカーネル行列 $X^T X$ 、 $X^T Y$ より求められる

テストデータと部分空間の距離をどのように求め、クラス識別を行うか

テストデータ \tilde{y}_k は、それを最もよく説明できる部分空間に属する

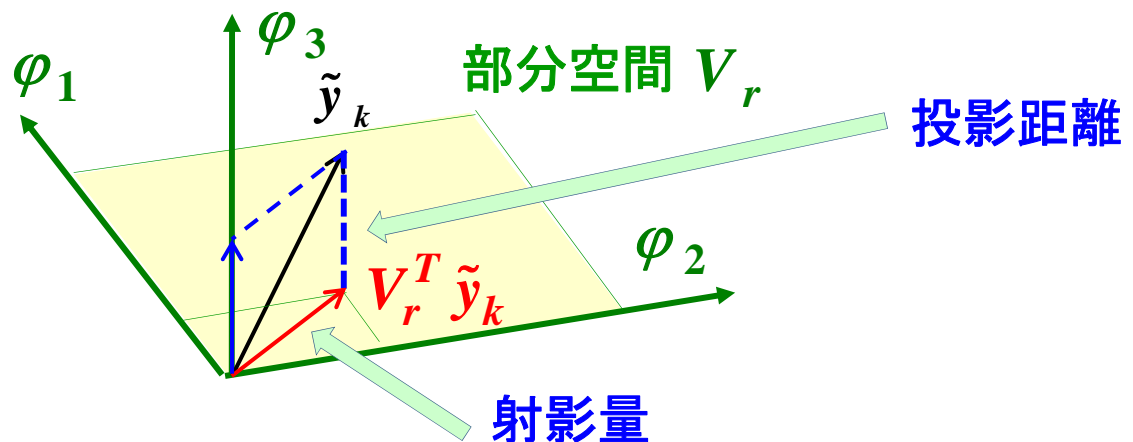
テストデータ \tilde{y}_k の部分空間 V_r への射影量 $\|\mathbf{V}_r^T \tilde{y}_k\|^2$ が最大

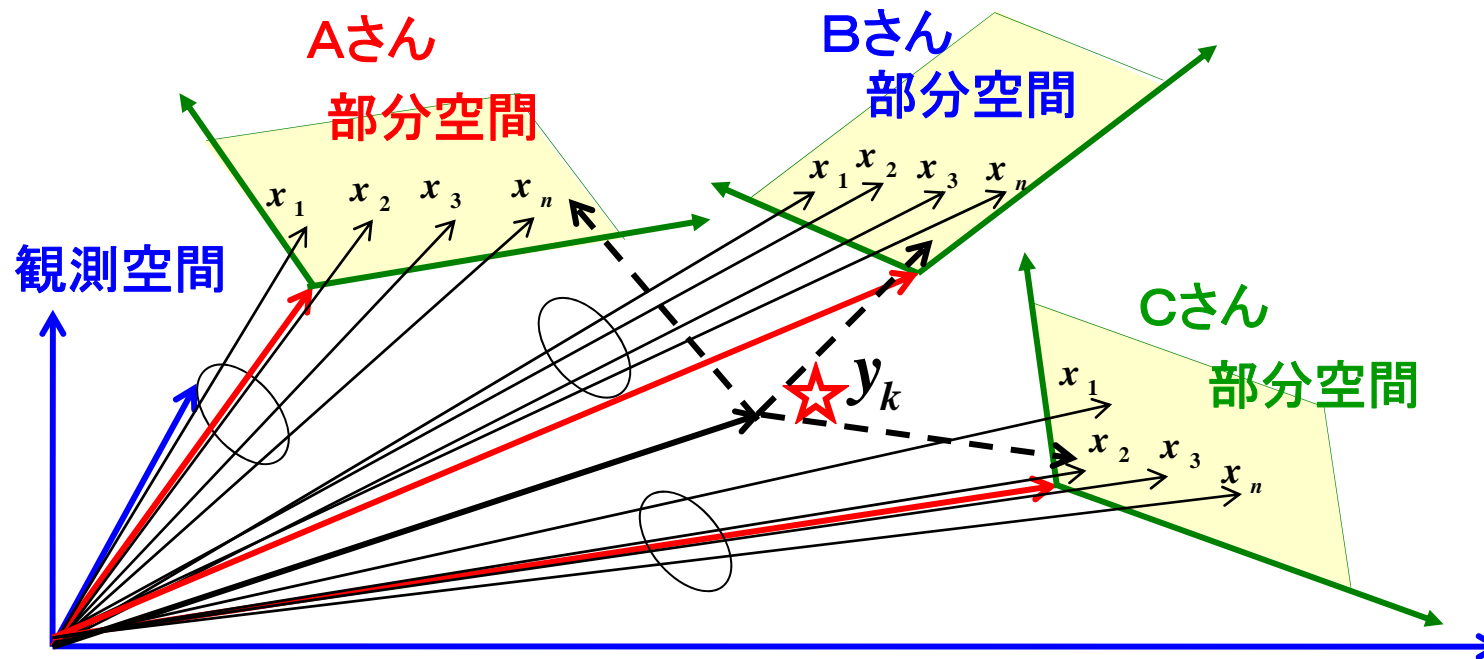
$$\therefore \hat{r} = \arg \max_r \|\mathbf{V}_r^T \tilde{y}_k\|^2$$

テストデータ \tilde{y}_k と部分空間 V_r との距離 $\|\tilde{y}_k\|^2 - \|\mathbf{V}_r^T \tilde{y}_k\|^2$ が最小

$$\therefore \hat{r} = \arg \min_r \left(\|\tilde{y}_k\|^2 - \|\mathbf{V}_r^T \tilde{y}_k\|^2 \right) \quad \text{この距離のことを投影距離}$$

$\|\tilde{y}_k\|^2$ は $F = \tilde{Y}^T \tilde{Y}$ の
第 k 行 k 列の値 $F_{k k}$





(1) 学習: データ集合から各クラスの分散共分散行列 C を固有値分解して、部分空間の正規直交基底を求め、テストデータを射影して射影ベクトルを求める。

非線形の場合には、内積行列 G を固有値分解して U と Σ を求め、テストデータの部分空間への射影ベクトルを求める。

(2) 認識: テストデータの射影ベクトルより、射影量が最大となるクラスに分類する。あるいは、投影距離が最小となるクラスに分類する。

平均を引く(分散共分散行列): 主成分分析、 平均を引かない(相関行列): CLAFIC (Class-Featuring Information Compression: Watanabe, 1967)

投影距離とマハラノビス距離

事後確率: $p(\omega | x) = \frac{p(x | \omega)p(\omega)}{P(x)}$ ベイズ識別

正規分布: $p(x | \omega) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right]$

識別関数: $g(x) = (x - \mu)^T \Sigma^{-1} (x - \mu) + \log |\Sigma| - 2 \log P(\omega)$

$P(\omega)$ の定数化

(1) ベイズ識別: $g(x) = \sum_{i=1}^n \frac{(x - \mu, v_i)^2}{\lambda_i} + \sum_{i=1}^n \log \lambda_i$ $\Sigma = V \Lambda V^T = \sum_{i=1}^n \lambda_i v_i v_i^T$

高次固有値の定数化

(2) 疑似ベイズ: $g(x) = \sum_{i=1}^k \frac{(x - \mu, v_i)^2}{\lambda_i} + \sum_{i=k+1}^n \frac{(x - \mu, v_i)^2}{\delta} + \sum_{i=1}^k \log \lambda_i + \sum_{i=k+1}^n \log \delta$

固有値項 $\log|\Sigma|$ の除去

(3) マハラノビス距離: $g(x) = (x - \mu)^T \Sigma^{-1} (x - \mu) = \sum_{i=1}^n \frac{(x - \mu, v_i)^2}{\lambda_i}$

$$(3) \text{ マハラノビス距離: } g(x) = (x - \mu)^T \Sigma^{-1} (x - \mu) = \sum_{i=1}^n \frac{(x - \mu, v_i)^2}{\lambda_i}$$

高次固有値の定数化

$$(4) \text{ 疑似マハラノビス距離: } g(x) = \sum_{i=1}^k \frac{(x - \mu, v_i)^2}{\lambda_i} + \sum_{i=k+1}^n \frac{(x - \mu, v_i)^2}{\delta}$$

$$h(x) = \delta g(x) = \sum_{i=1}^k \frac{\delta (x - \mu, v_i)^2}{\lambda_i} + \sum_{i=k+1}^n (x - \mu, v_i)^2$$

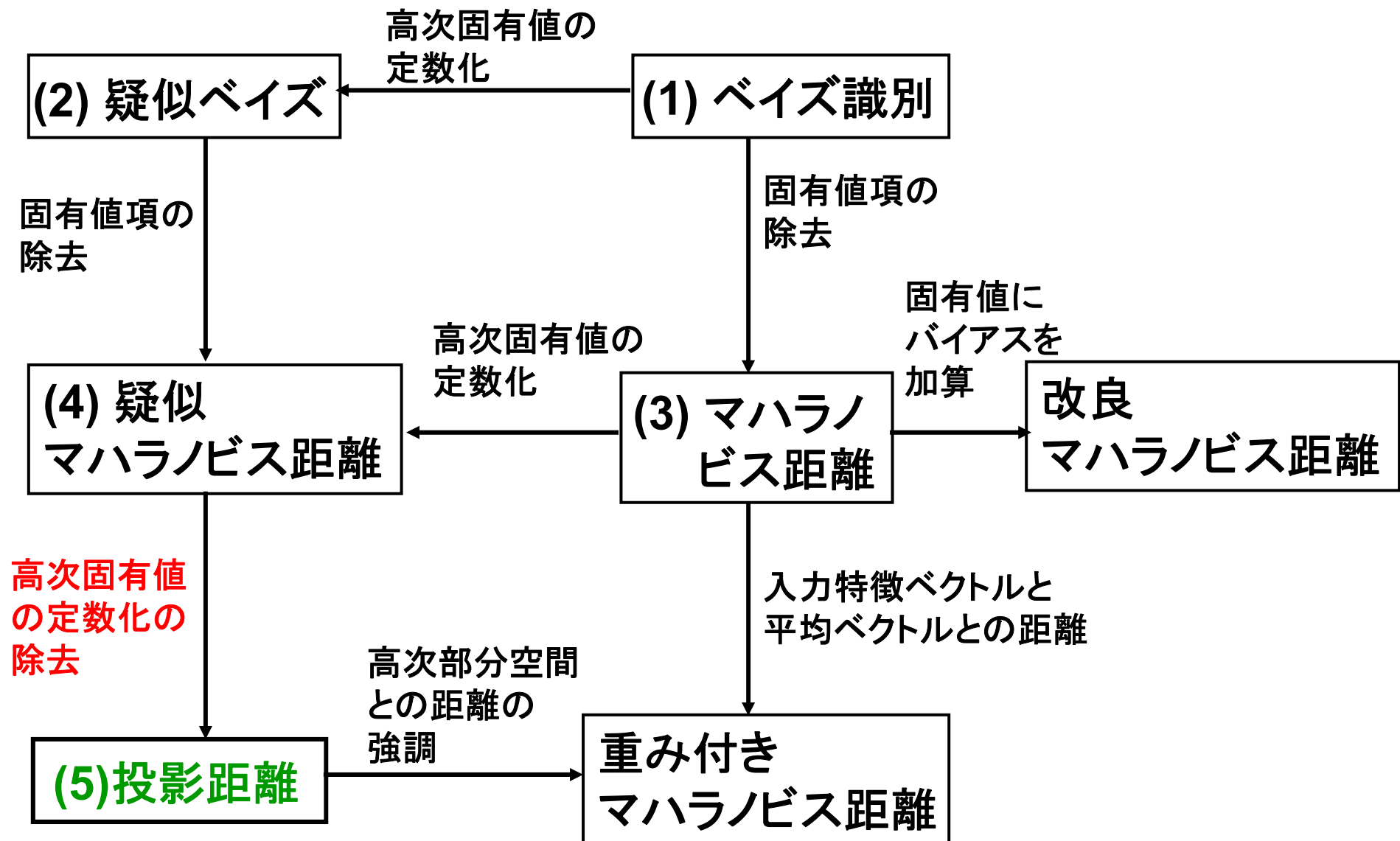
$$= \|x - \mu\|^2 - \sum_{i=1}^k (x - \mu, v_i)^2 + \sum_{i=1}^k \frac{\delta (x - \mu, v_i)^2}{\lambda_i}$$

$$= \|x - \mu\|^2 - \sum_{i=1}^k \left(1 - \frac{\delta}{\lambda_i}\right) (x - \mu, v_i)^2$$

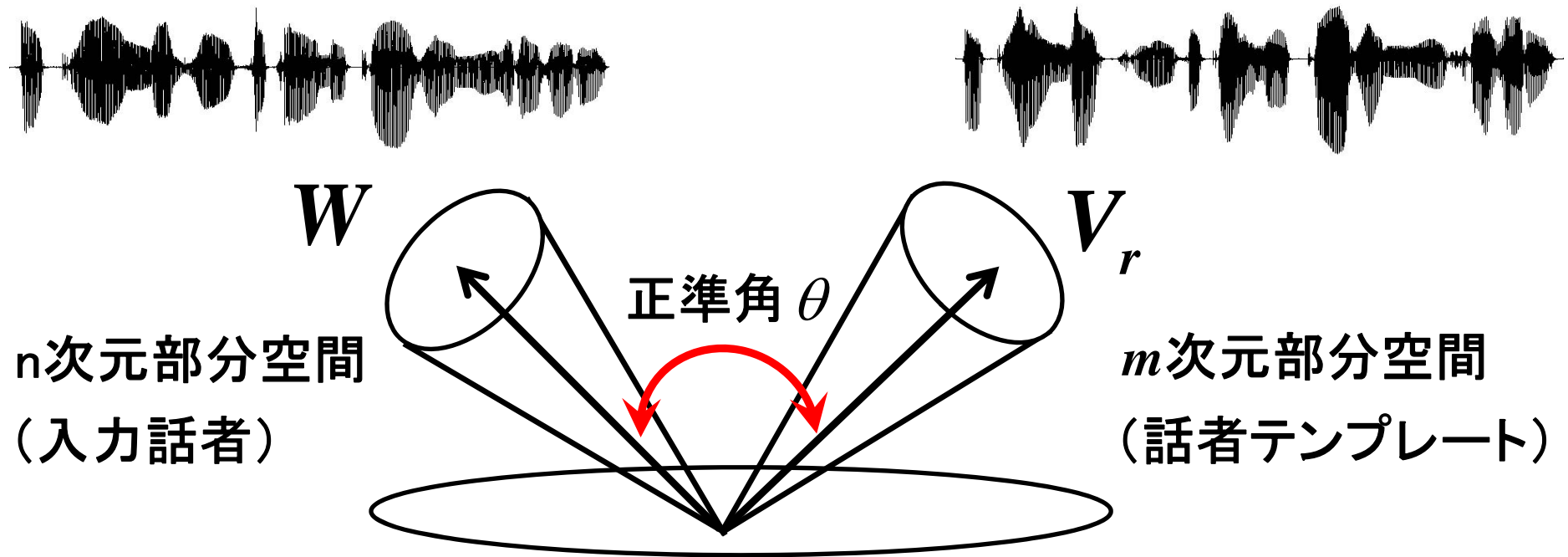
高次固有値の定数化の除去 ($\delta \ll \lambda$)

$$(5) \text{ 投影距離: } g(x) = \|x - \mu\|^2 - \sum_{i=1}^k (x - \mu, v_i)^2 = \sum_{i=k+1}^n (x - \mu, v_i)^2$$

投影距離の位置づけ： マハラノビス距離の次元削減



5. 核非線形相互部分空間法[9-11] 1985, 2001, 2008



$$\cos^2 \theta = \max_{w \in W, v \in V_r, \|w\| \neq 0, \|v\| \neq 0} \frac{|(w, v)|^2}{\|w\|^2 \|v\|^2}$$

アルゴリズム

(1) 学習時系列データから各話者の部分空間 V_r (m 次元)を構成。

$$V_r = (v_1^r, \dots, v_m^r)$$

入力時系列データから入力話者の部分空間 W (n 次元)を構成。

$$W = (w_1, \dots, w_n)$$

(2) 次の行列 X を求める。

$$X = (x_{ij}) \quad x_{ij} = \sum_{k=1}^m (w_i, v_k^r)(v_k^r, w_j)$$

(3) 行列 X の固有値を求める。

$$Xc = \lambda c$$

最大固有値 $\lambda_{\max} = \cos^2 \theta$ は、第1正準角。第1正準角から第 p 正準角までの平均を、2つの部分空間の類似度とする。最大の平均正準角を持つ話者に認識する。

非線形化した核非線形相互部分空間法が提案されている [11]

話者認識・話者照合

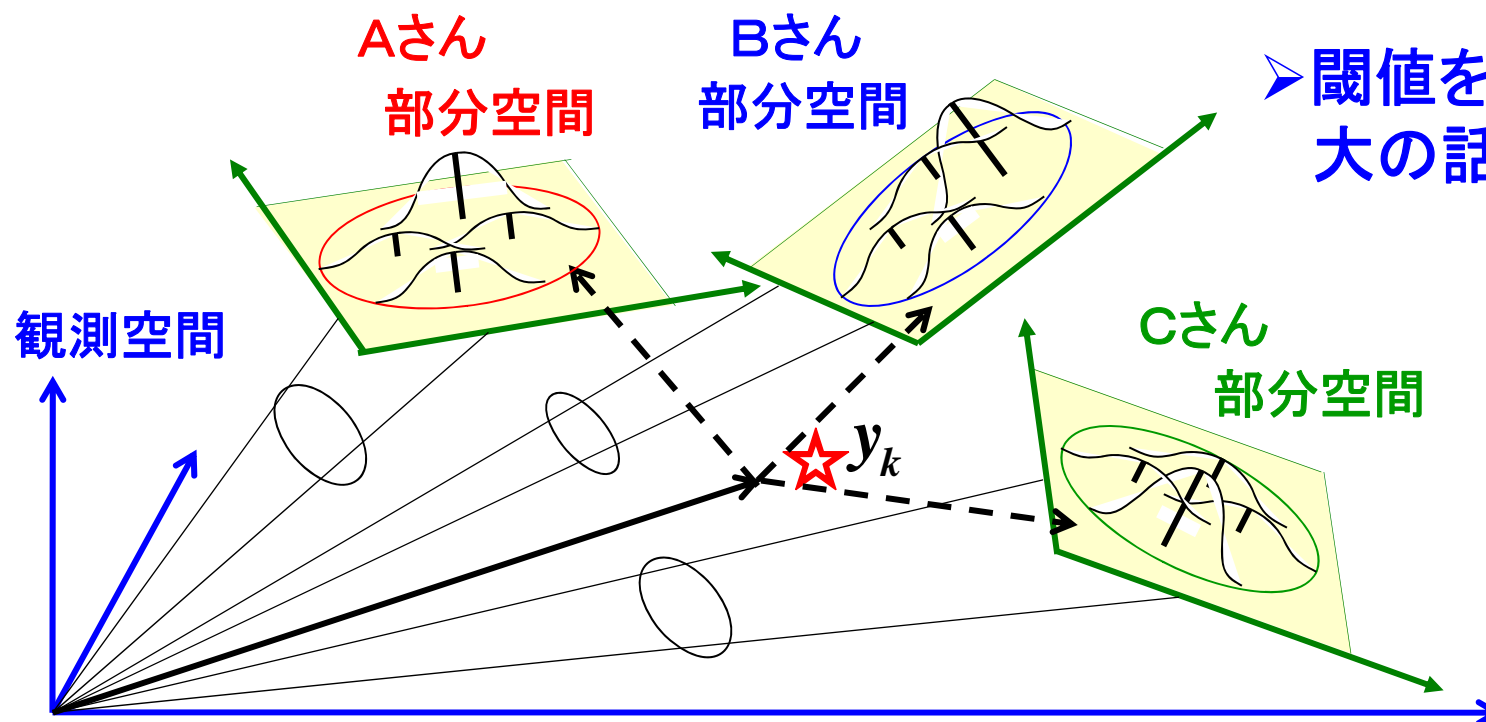
(3) 部分空間法＋確率

6. 部分空間GMM[12] 2007

➤ 学習データを基に、各話者の部分空間を構成し、部分空間内でGMMを学習する。

➤ テストデータを各話者の部分空間に射影して射影ベクトルを求める。各話者のGMMを使って尤度を計算する。

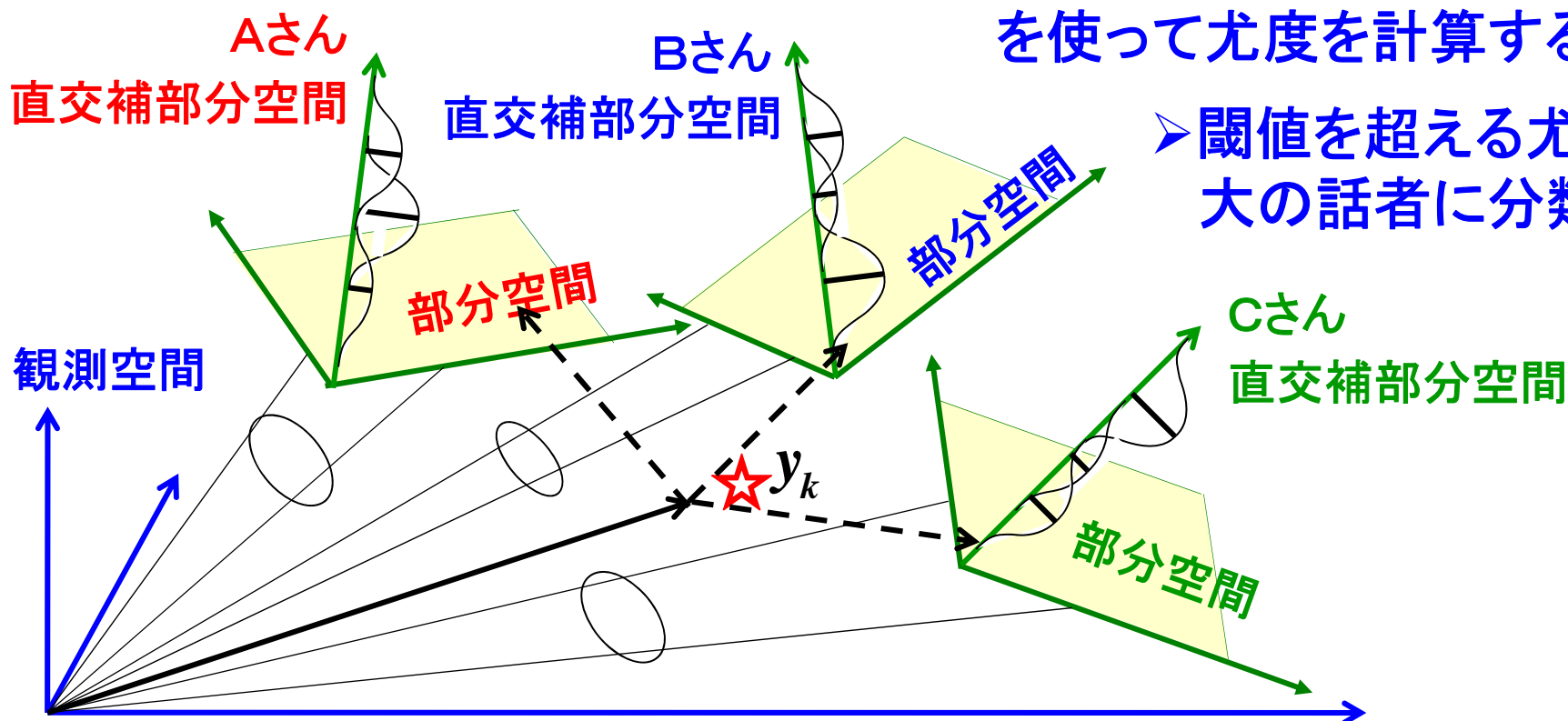
➤ 閾値を超える尤度最大の話者に分類する。



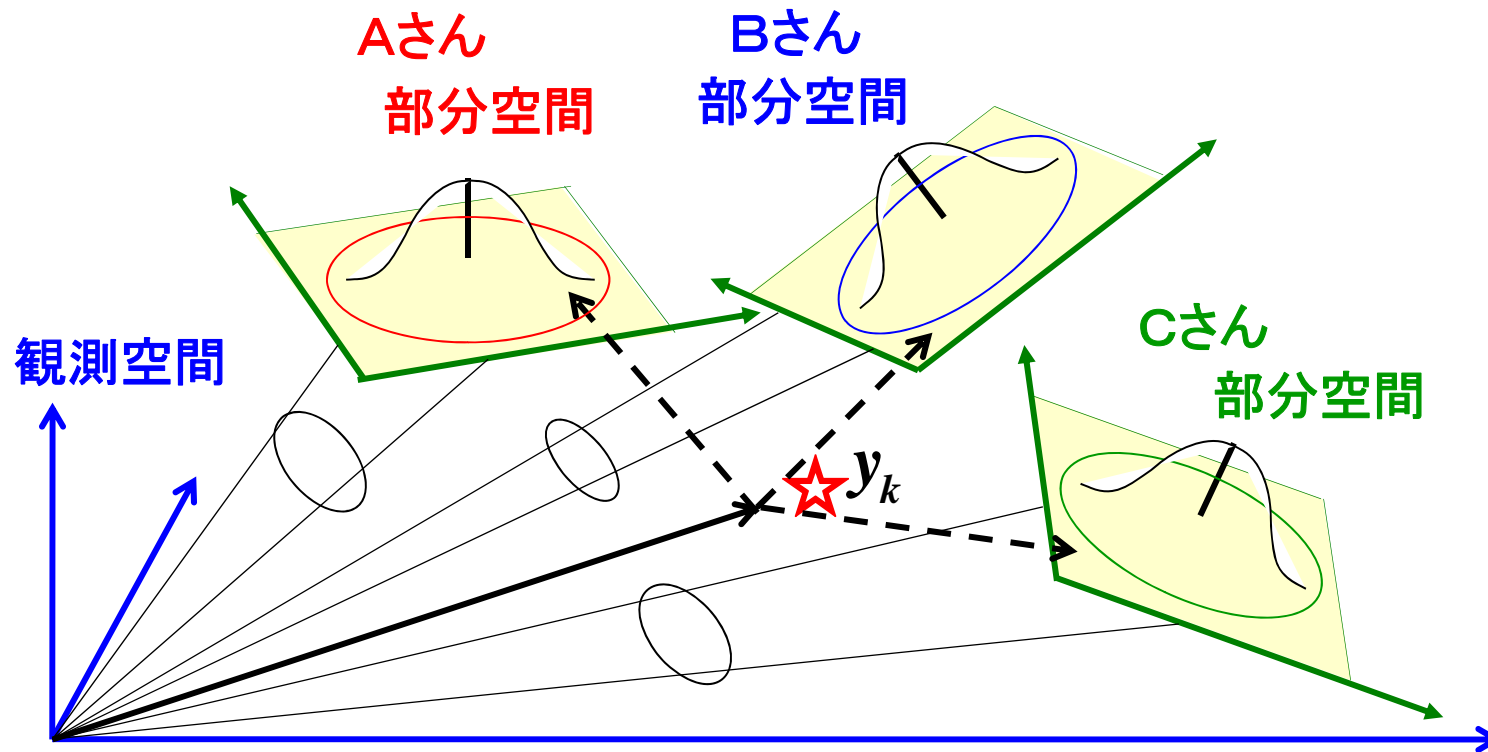
7. 直交補部分空間GMM[13] 2002

変動が大きい音韻性は部分空間に局在し、変動が小さい話者性は、直交補空間に局在する
と考える。

- 学習データを基に、各話者の部分空間を構成し、直交補部分空間内でGMMを学習する。
- テストデータを各話者の直交補部分空間に射影して射影ベクトルを求める。各話者のGMMを使って尤度を計算する。
- 閾値を超える尤度最大の話者に分類する。



8. LDA部分空間＋正規分布対数尤度 [14] 1988

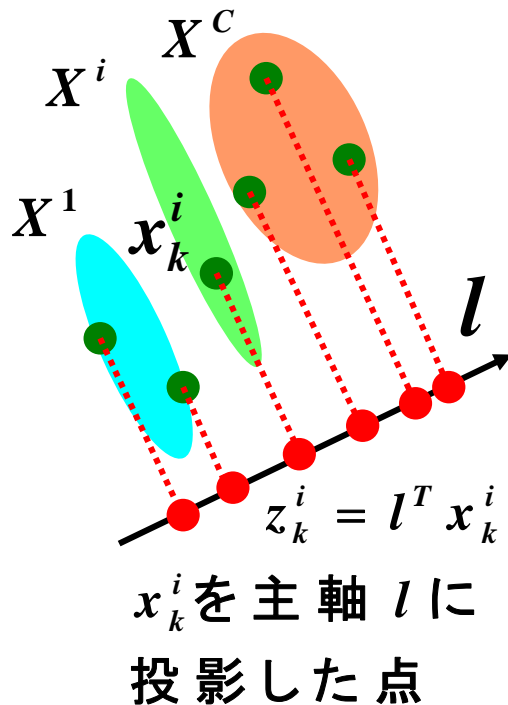


- (1) 部分空間をPCAではなく、LDAで構成する。
- (2) 確率モデルは単峰正規分布であり、スコアは対数尤度を用いる。

LDA

データ集合 $\{X^i\}$ の1つの主軸を l とする. $\{X^i\}$ の全データをこの主軸に投影して異なるクラス分離度が最大となる主軸 l を求める.

$$\text{分離度: } J = \tilde{\sigma}_B / \tilde{\sigma}_W$$



$$\tilde{\sigma}_W = \sum_i P(w^i) \tilde{\sigma}^i \quad \text{投影後の平均クラス内分散}$$

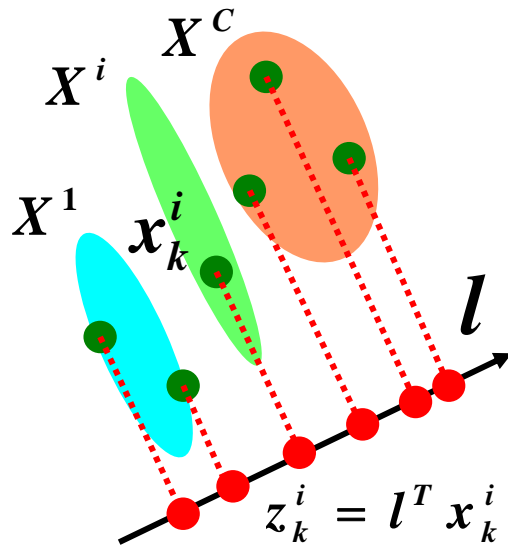
$$\tilde{\sigma}^i = 1/n_i \sum_k (z_k^i - \bar{z}^i)(z_k^i - \bar{z}^i)^T \quad \begin{array}{l} \text{投影後の} \\ \text{各クラス内分散} \end{array}$$

$$= l^T \left\{ \underbrace{1/n_i \sum_k (x_k^i - \bar{x}^i)(x_k^i - \bar{x}^i)^T}_{\Sigma^i} \right\} l = l^T \Sigma^i l$$

$$\tilde{\sigma}_W = \sum_i P(w^i) l^T \Sigma^i l = l^T \underbrace{\sum_i P(w^i) \Sigma^i}_\Sigma l = l^T \Sigma_W l$$

LDA

データ集合 $\{X^i\}$ の1つの主軸を l とする. $\{X^i\}$ の全データをこの主軸に投影して異なるクラスの間隔度が最大となる主軸 l を求める.



x_k^i を主軸 l に
投影した点

$$\begin{aligned} \bar{z}^i &= l^T \bar{x}^i \\ \bar{z} &= l^T \bar{x} \end{aligned}$$

$$\text{分離度: } J = \tilde{\sigma}_B / \tilde{\sigma}_W = l^T \Sigma_B l / l^T \Sigma_W l$$

$$\tilde{\sigma}_W = \sum_i P(w^i) l^T \Sigma^i l = l^T \Sigma_W l$$

$$\tilde{\sigma}_B = \sum_i P(w^i) (\bar{z}^i - \bar{z})(\bar{z}^i - \bar{z})^T \quad \text{投影後の}$$

クラス間分散

$$= l^T \left\{ \sum_i P(w^i) (\bar{x}^i - \bar{x})(\bar{x}^i - \bar{x})^T \right\} l$$

$$= l^T \Sigma_B l$$

最大となる A を求める

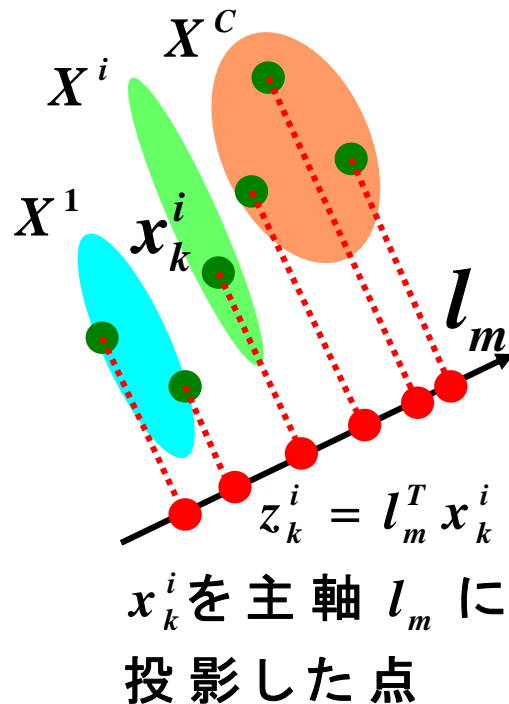
複数次軸 $l \Rightarrow A$

$$\text{分離度: } J = \text{tr}\{A^T \Sigma_B A\} / \text{tr}\{A^T \Sigma_W A\}$$

LDA部分空間

データ集合 X^m の1つの主軸を l_m とする. $\{X^i\}$ の全データをこの主軸に投影して異なるクラス分離度が最大となる主軸 l_m を求める.

$$\text{分離度: } J = \tilde{\sigma}_{Bm} / \tilde{\sigma}_W$$



$$\tilde{\sigma}_W = \sum_i P(w^i) \tilde{\sigma}^i \quad \text{投影後の平均クラス内分散}$$

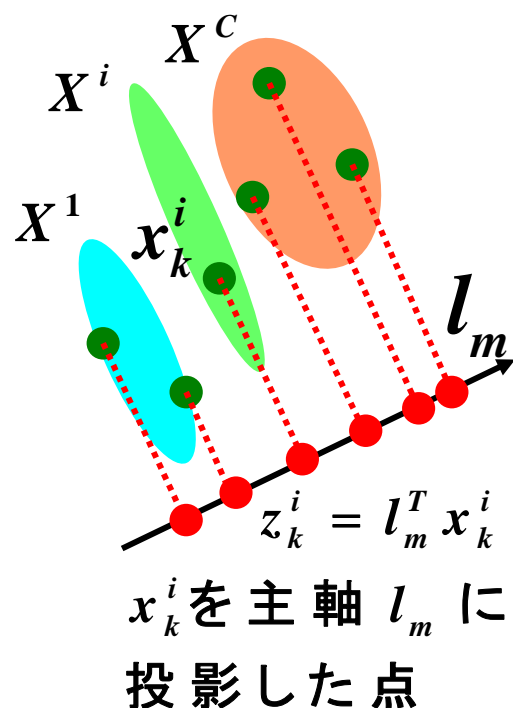
$$\tilde{\sigma}^i = 1/n_i \sum_k (z_k^i - \bar{z}^i)(z_k^i - \bar{z}^i)^T \quad \begin{array}{l} \text{投影後の} \\ \text{各クラス内分散} \end{array}$$

$$= l_m^T \{ \underbrace{1/n_i \sum_k (x_k^i - \bar{x}^i)(x_k^i - \bar{x}^i)^T}_{\Sigma^i} \} l_m = l_m^T \Sigma^i l_m$$

$$\tilde{\sigma}_W = \sum_i P(w^i) l_m^T \Sigma^i l_m = l_m^T \underbrace{\sum_i P(w^i) \Sigma^i}_{\Sigma_W} l_m = l_m^T \Sigma_W l_m$$

LDA部分空間

データ集合 X^m の1つの主軸を l_m とする. $\{X^i\}$ の全データをこの主軸に投影して異なるクラス分離度が最大となる主軸 l_m を求める.



$$\bar{z}^i = l_m^T \bar{x}^i$$

複数軸 $l_m \Rightarrow A_m$

$$\text{分離度: } J = \tilde{\sigma}_{Bm} / \tilde{\sigma}_W = l_m^T \Sigma_{Bm} l_m / l_m^T \Sigma_W l_m$$

$$\tilde{\sigma}_W = \sum_i P(w^i) l_m^T \Sigma^i l_m = l_m^T \Sigma_W l_m$$

$$\tilde{\sigma}_{Bm} = \sum_i P(w^i) (\bar{z}^i - \bar{z}^m) (\bar{z}^i - \bar{z}^m)^T \quad \text{投影後のクラス間分散}$$

$$= l_m^T \left\{ \sum_i P(w^i) (\bar{x}^i - \bar{x}^m) (\bar{x}^i - \bar{x}^m)^T \right\} l_m$$

$$= l_m^T \Sigma_{Bm} l_m$$

最大となる A_m を求める

$$\text{分離度: } J = \text{tr}\{A_m^T \Sigma_{Bm} A_m\} / \text{tr}\{A_m^T \Sigma_W A_m\}$$

話者認識・話者照合

(4) 部分空間法による変動成分の吸収

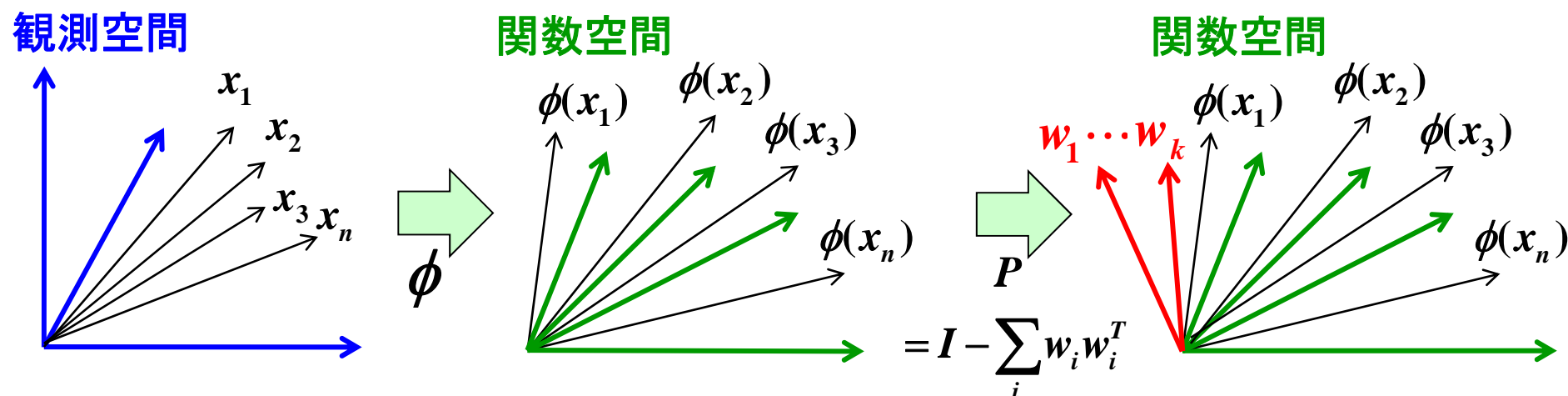
9. NAP(Nuisance Attribute Projection)

回線特性の差[15,16]

感情[17]

時期差、個人差

2004, 2005, 2007



$$\delta = \sum_{ij} W_{ij} \| P(\phi(x_i) - \phi(x_j)) \|^2$$

が最小となる X を決定する

$$W_{ij} = \begin{cases} 1 & x_i \text{の回線} \neq x_j \text{の回線} \\ 0 & \text{それ以外} \end{cases}$$

$P\phi(x_i)$ は、 $\phi(x_i)$ から $\sum_i w_i w_i^T$ への射影成分を取り除いた射影ベクトル

$$X = (w_1, \dots, w_k) \quad XX^T = \sum_i w_i w_i^T$$

$$P = I - XX^T$$

$$W_{ij} = \begin{cases} 1 & x_i \text{の回線} \neq x_j \text{の回線} \\ 0 & \text{それ以外} \end{cases}$$

$$\delta = \sum_{ij} W_{ij} \| P(\phi(x_i) - \phi(x_j)) \|$$

が最小となる X を決定する

$$AZ(W)A^T X = X\Lambda$$

$$Z(W) = \text{diag}(W \mathbf{1}) - W$$

$$KZ(W)KV = KV\Lambda$$

$$Z(W)KV = V\Lambda$$

$Z(W)K$ を固有値分解して V を求める

$$A = (\phi(x_1), \dots, \phi(x_n))$$

$$X = (w_1, \dots, w_k)$$

$$XX^T = \sum_i w_i w_i^T$$

$$P = I - XX^T$$

$$X = AV$$

$$K = A^T A$$

回線の違いを吸収した空間
においてカーネル行列を求める

$$K' = (PA)^T (PA)$$

$$= A^T (I - XX^T)^T (I - XX^T) A$$

$$= K - (KV)(KV)^T$$

カーネルSVMで話者認識を実行

話者認識・話者照合

(5) 確率主成分分析と因子分析

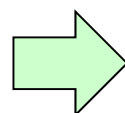
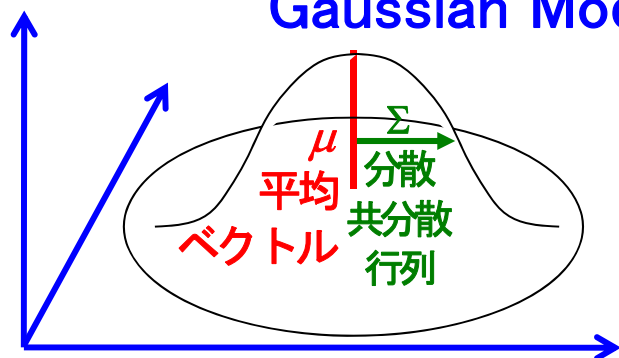
9. 確率主成分分析 [18] 1999

確率主成分分析

主成分分析
(1つの線形部分空間)

観測空間

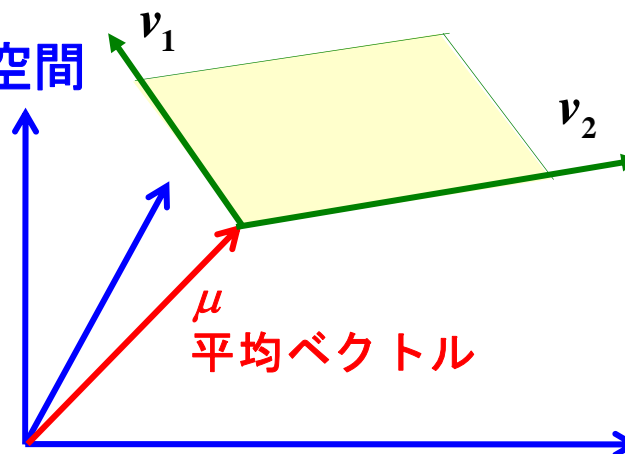
Gaussian Model



固有値分解

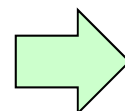
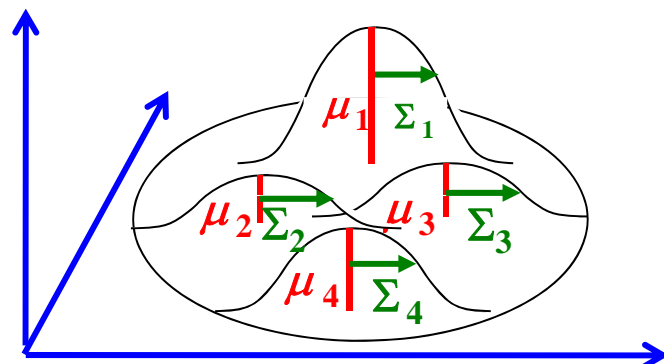
$$\Sigma = V \Lambda V^T$$

観測空間



観測空間

GMM

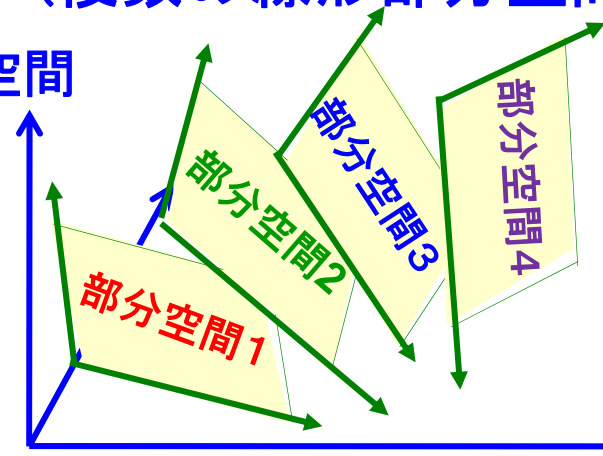


固有値分解

$$\Sigma = V \Lambda V^T$$

観測空間

混合主成分分析
(複数の線形部分空間)



確率主成分分析

$$x = W \cdot f + \mu + \varepsilon$$

観測ベクトル(d次元)
因子負荷行列
潜在変数ベクトル
平均ベクトル
ノイズベクトル

9次元(因子)
 $N(0, \mathbf{I})$
 $N(0, \mathbf{I})$
 $N(0, \sigma^2 \mathbf{I})$

$$p(x | f) = (2\pi\sigma^2)^{-\frac{d}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \|x - Wf - \mu\|^2 \right\}$$

$$p(f) = (2\pi)^{-\frac{q}{2}} \exp \left\{ -\frac{1}{2} f^T f \right\}$$

$$p(x) = \int p(x | f) p(f) df$$

$$= (2\pi)^{-\frac{d}{2}} |\mathbf{C}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \mathbf{C}^{-1} (x - \mu) \right\}$$

$$\mathbf{C} = \sigma^2 \mathbf{I} + \mathbf{W}\mathbf{W}^T$$

対数尤度:

$$\frac{\partial L}{\partial \mu} = 0$$

$$\frac{\partial L}{\partial W} = 0$$

対数尤度:

$$L = \sum_{k=1}^N \ln \{ p(x_k) \} = -\frac{N}{2} \left\{ d \ln(2\pi) + \ln |\mathbf{C}| + \text{tr}(\mathbf{C}^{-1} \mathbf{S}) \right\}$$

$$\mathbf{S} = \frac{1}{N} \sum_{k=1}^N (x_k - \mu)(x_k - \mu)^T$$

確率主成分分析

$$x = W \cdot f + \mu + \varepsilon$$

観測ベクトル(d次元)
 因子負荷行列
 潜在変数ベクトル
 平均ベクトル
 ノイズベクトル
 $N(0, \sigma^2 \mathbf{I})$
 $N(0, \mathbf{I})$

$$x - \mu = W^q \cdot f^q + \varepsilon = W^d \cdot f^d$$

$$\varepsilon = W^{d-q} \cdot f^{d-q} \sim N(0, \sigma^2 \mathbf{I})$$

直交補部分空間における
平均分散

対数尤度:

$$L = -\frac{N}{2} \left\{ d \ln(2\pi) + \ln |C| + \text{tr}(C^{-1}S) \right\}$$

$$S = \frac{1}{N} \sum_{k=1}^N (x_k - \mu)(x_k - \mu)^T$$

$$\frac{\partial L}{\partial \mu} = 0 \text{より、} \hat{\mu} = \frac{1}{N} \sum_{k=1}^N x_k$$

$$\frac{\partial L}{\partial W} = 0 \text{より、} \hat{W} = V_q (\Lambda_q - \sigma^2 I)^{\frac{1}{2}} R$$

$S = V \Lambda V^T$ であり、 V_q と Λ_q は固有値の
 大きい固有ベクトル行列と固有値行列
 R は任意の直交(回転)行列

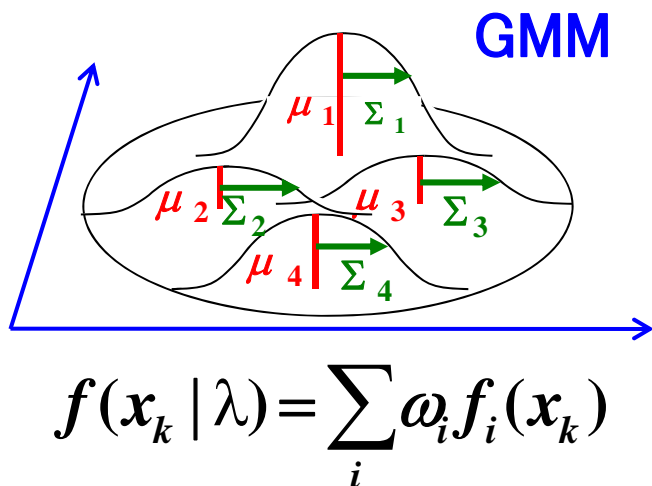
$$\hat{\sigma}^2 = \frac{1}{d-q} \sum_{j=q+1}^d \lambda_j$$

混合確率主成分分析 (複数の線形部分空間)

$$x = W_m \cdot f_m + \mu_m + \varepsilon_m$$

観測ベクトル(d次元)
因子負荷行列
q次元(因子)
潜在変数ベクトル
平均ベクトル
ノイズベクトル

$N(0, \sigma_m^2 I)$
 $N(0, I)$



$$p(x) = \sum_{m=1}^M \pi_m p(x | m) \quad \pi_m \geq 0, \quad \sum_{i=1}^m \pi_m = 1$$

$$p(x | m) = \int p(x | m, f) p(f | m) df$$

$$= (2\pi)^{-\frac{d}{2}} |C_m|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x - \mu_m)^T C_m^{-1} (x - \mu_m) \right\}$$

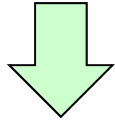
$$C_m = \sigma_m^2 I + W_m W_m^T$$

対数尤度:

$$L = \sum_{k=1}^N \ln \{ p(x_k) \} = \sum_{k=1}^N \ln \left\{ \sum_{m=1}^M \pi_m p(x_k | m) \right\}$$

GMMと同じく、EMアルゴリズムで
 $\{\pi_m, \mu_m, W_m, \sigma_m^2\}$ を推定する

確率主成分分析

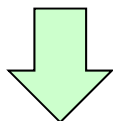


因子分析

$$x = W \cdot f + \mu + \varepsilon$$

観測ベクトル	因子負荷行列	(因子)	潜在変数	平均ベクトル	ノイズ
				$N(0, \sigma^2 I)$	
				$N(0, I)$	$N(0, \Psi)$

混合確率主成分分析



混合因子分析

$$x = W_m \cdot f_m + \mu_m + \varepsilon_m$$

変動要因の吸収 [19, 20] 2007, 2008

因子 f として、回線特性や時期差、個人性を取り、複数因子で観測データを表現する。例えば、

$$x = W \cdot f + U \cdot g + \mu + \varepsilon$$

と表し、 W, U, μ, ε を学習する。

入力データ x が与えられると、 x を分解し、 $U \cdot g$ が回線特性であれば、これを削除して比較する。

変動要因の吸収 [21, 22] 2002, 2003

データが混合因子分析の制約に基づいて発生しているものと考え、パラメータを推定する。これを基に、テストデータの発声確率を話者ごとに求め、話者認識を行う。

参考文献

- [1] D. A. Reynolds, R. C. Rose : “Robust text-independent speaker identification using Gaussian mixture speaker models”, IEEE Trans. on speech and audio processing, Volume 3, Issue 1, pp.72–83, 1995.
- [2] M.Schmidt, H.Gish: “Speaker identification via support vector classifiers”, Proc. of ICASSP, pp.105–108, 1996.
- [3] W. M. Campbell, D. E. Sturim, D. A. Reynolds, and A. Solomonoff: “SVM based speaker verification using a GMM supervector kernel and NAP variability compensation,” in Proc. of ICASSP, pp. 97–100, 2006.
- [4] Y.Ariki and K.Doi: “Speaker Recognition based on Subspace Method”, ICSLP ‘94, pp.1859–1862, 1994.
- [5] B.Scholkopf, A.Smola, and K.-R.Muller, “Nonlinear Component Analysis as a Kernel Eigenvalue Problem,” Neural Computation, Vol.10, pp.1299–1319,1998.
- [6] 津田宏治: “ヒルベルト空間における部分空間,” 電子情報通信学会論文誌, Vol.82-D-II, No.4, pp.592–599, 1999.
- [7] 前田英作, 村瀬洋: “カーネル非線形部分空間法によるパターン認識,” 電子情報通信学会論文誌, Vol.82-D-II, No.4, pp.600–612, 1999.
- [8] 浜崎武, 野田秀樹, 河口英二: “ヒルベルト空間で部分空間法を用いた話者識別,” 電子情報通信学会技報, PRMU研究会, pp.57–62, 2000.
- [9] 前田 賢一 渡辺 貞一: “局所的構造を導入したパターン・マッチング法,” 電子情報通信学会論文誌, Vol.J68-D, No.3, pp.345–352, 1985.
- [10] 市野将嗣, 坂野鋭, 小松尚久: “話者認識における核非線形相互部分空間法の適用と有効性に関する一考察, 画像の認識・理解シンポジウム(MIRU2008)サテライトワークショップ部分空間法研究会Subspace2008,2008.

- [11]坂野鋭, 武川直樹, 中村太一: “核非線形相互部分空間法による物体認識,” 電子情報通信学会論文誌, Vol.J84-D-II, No.8, pp.1549-1556, 2001.
- [12] C. L. Ying and A. BJ Teoh: “Speaker verification using probabilistic 2D CLAFIC,” IEICE Electronics Express, Vol.4, No.5, pp.179-184, 2007.
- [13]西田昌史, 有木康雄: “音韻性を抑えた話者空間への射影による話者認識”, 電子情報通信学会論文誌, Vol.85-D-II, No.4, pp.554-562, 2002.
- [14]J.B.Attili, M.Savic and J.P.Campbell: “A TMs32020-based real time, text-independent, automatic speaker verification system,” Proc. of ICASSP, pp.599-602, 1988.
- [15] A.Solomonoff, W.M.Campbell, I.Boardman: “Advances in Channel Compensation for SVM Speaker Recognition,” ICASSP, pp.18-23, 2005.
- [16] A.Solomonoff, C.Quillen, W.M.Campbell: “Channel Compensation for SVM Speaker Recognition,” In Proc. Odyssey: The Speaker and Language Recognition Workshop, ISCA, pp.41-44, 2004.
- [17]H.Bao, M.Xu, T.F.Zheng: “Emotion Attribute Projection for Speaker Recognition on Emotional Speech,” Interspeech, pp.758-761, 2007.
- [18] M.E. Tipping and C.M. Bishop, “Mixtures of probabilistic principal component analyzers,” Neural Computation, vol.11, no.2, pp.443-482, 1999.
- [19] D.Matrouf, N. Scheffer, B.Fauve, J-F. Bonastre: “A straightforward and efficient implementation of the factor analysis model for speaker verification,” In INTERSPEECH-2007, 1242-1245, 2007.
- [20] P.Kenny, P.Ouellet, N.Dehak, V.Gupta, and P.Dumouchel: “A Study of Inter-Speaker variability in Speaker Verification, IEEE Transactions on Audio, Speech and Language Processing, Vol.16, 5, pp.980-988, 2008.

[21] P.Ding, Y. Liu, B. Xu: “Factor Analyzed Gaussian Mixture Models for Speaker Identification,” In Proc. ICSLP, pp.1341–1344, 2002.

[22]山本 啓善, 南角 吉彦, 宮島 千代美, 徳田 恵一, 北村 正, “混合因子分析に基づく話者識別モデルのパラメータ共有構造” 電子情報通信学会技術研究報告, vol.103, no.519, pp.91–96, Dec. 2003.