Vol.2011-SLP-87 No.3
2011/7/21

条件付きエントロピー最小化基準に基づくマルチカーネル学習を用いた発話スタイル変動に頑健な話者照合

小川哲司[†1]　日野英逸[†2]
村田昇[†2]　小林哲則[†3]

話者内変動に頑健な話者照合システムについて検討を行った．発話スタイルや発話時期の違いなどの影響で，同一話者の音声であっても音響的な変動が生じる．このような音響変動は，一般的に話者照合システムの性能を劣化させることが知られている．この問題を解決するため，条件付きエントロピー最小化という，同一クラスのデータを密集させ，かつ異なるクラスのデータを互いに遠ざける性質を持つ最適化基準を用いてマルチカーネル学習を行い，話者照合システムを構築することを試みた．話者照合実験の結果，提案システムは，従来のマージン最大化に基づき構築したシステムと比較して，発話スタイル変動に起因する話者クラス内での音響特徴変動に対して頑健な性能を与えた．

Speaker verification system robust to speaking style variation using multiple kernel learning based on conditional entropy minimization

Tetsuji Ogawa,[†1] Hideitsu Hino,[†2] Noboru Murata[†2]
and Tetsunori Kobayashi[†3]

We developed a new speaker verification system that is robust to intra-speaker variation. There is a strong likelihood that intra-speaker variations will occur due to changes in speaking styles, the periods when an individual speaks, and so on. It is well known that such variation generally degrades the performance of speaker verification systems. To solve this problem, we applied multiple kernel learning based on conditional entropy minimization, which impose the data to be compactly aggregated for each class and ensure that the different classes were far apart from each other, to speaker verification. Experimental results showed that the proposed speaker verification system achieved a robust performance to intra-speaker variation derived from changes in the speaking styles compared to the conventional maximum margin-based system.

1. はじめに

話者照合システムでは，サポートベクタマシン (support vector machine; SVM) に代表されるカーネル法に基づく手法が多く用いられている[1),2)]．特に近年，複数のカーネル関数の凸結合を用いるマルチカーネル学習 (multiple kernel learning; MKL)[3),4)] が話者照合に用いられている[5)]．我々も，条件付きエントロピー最小化に基づくマルチカーネル学習 (MKL based on conditional entropy minimization; MCEM)[6)] を話者認識システムに適用し[7)]，マージン最大化に基づく MKL を用いたシステムの性能を上回ることを明らかにした．

発話スタイルや発話時期の違いにより，同一話者の音声であっても音響的な変動が生じるが，そのような音響変動は話者照合システムの性能に悪影響を及ぼすことが知られている[8),9)]．文献[5)] をはじめ多くの MKL では，マージン最大化に基づいて判別境界が構築されている．これらの方法では，クラス間でのデータの散らばりに焦点が当てられており，クラス内でのデータの散らばりについては陽に考慮されていない．それに対し，MCEM では，同一クラスのデータは特徴空間において密集し，異なるクラスのデータは離れるような最適化が行われる．したがって，MCEM を用いて構築した話者照合システムは，従来のマージン最大化に基づくシステムと比較して，同一話者内の特徴量の変動に対する頑健性を向上できる可能性がある．

本研究では，MCEM に基づき構築されたシステムと従来のマージン最大化に基づき構築されたシステムを比較することで，発話スタイルの変動に起因する話者内変動に対する MCEM の頑健性について調査を行う．本稿では，通常発声と Lombard 発声から成る多様な発話スタイルを含む音声を用いて話者モデルを学習することを考える．Lombard 効果は，騒音下で発話された音声に生じる音響変動現象であり，Lombard 効果を含む音声の音響特性は，通常発声音声の特性とは著しく異なることが知られている．例えば，Lombard 効果を含む音声では，通常発声の音声に対して音圧や音高の上昇，フォルマントのシフトなどが観測される．つまり，同一話者の音声であっても通常発声と Lombard 発声が含まれる場合

†1 早稲田大学 高等研究所
Waseda Institute for Advanced Study
†2 早稲田大学 先進理工学部 電気・情報生命工学科
Department of Electrical Engineering and Bioscience, Waseda University
†3 早稲田大学 基幹理工学部 情報理工学科
Department of Computer Science, Waseda University

は，音響的な変動を陽に含んでいると言える．したがって，このようなデータを用いて，話者照合システムにおける話者内変動に対する頑健性を調査することができる．

本稿の構成は以下の通りである．まず，2 において，MKL について簡単に述べる．3 では，本研究の基礎となる MCEM の最適化基準と最適化アルゴリズムについて述べる．さらに，4 では，MCEM に基づく話者照合システムを，発話スタイルの変動に対する頑健性に焦点を当てて評価を行う．最後に，5 においてまとめを述べる．

2. マルチカーネル学習 (MKL)

本章では，MKL について概観する．観測データ $D = \{\boldsymbol{x}_i\}_{i=1}^N$，およびそれらのクラスラベル $\{y_i\}_{i=1}^N$，$y_i \in \{\pm 1\}$ が与えられたとき，カーネル法における判別関数は，$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \Phi(\boldsymbol{x}_i), \Phi(\boldsymbol{x}_j) \rangle$ なるカーネル関数を用いて以下のように書ける．

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i k(\boldsymbol{x}, \boldsymbol{x}_i) \tag{1}$$

ここで，Φ は，入力空間から特徴空間への写像を表す．このとき，典型的なカーネルマシンの学習は，下記のように表される．

$$\min_{\boldsymbol{\alpha}} \left[\sum_{i=1}^{N} \mathcal{L}\left(y_i, \sum_{j=1}^{N} \alpha_j k(\boldsymbol{x}_i, \boldsymbol{x}_j) \right) + ||\boldsymbol{\alpha}||_K^2 \right] \tag{2}$$

ここで，$\mathcal{L}(\cdot)$ は損失関数であり，SVM では，hinge 損失 $\mathcal{L}(y,t) = \max(0, 1-yt)$ が用いられる．また，カーネル行列は，$[K]_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ のように，カーネル関数を要素とする行列である．MKL では，カーネル行列 K を，S 個のカーネル行列の凸結合 $\sum_{s=1}^{S} \boldsymbol{\beta}_s K_s$ で置き換える．ここで，結合係数 $\boldsymbol{\beta}_s$ は，$\sum_{s=1}^{S} \boldsymbol{\beta}_s = 1$ および $\boldsymbol{\beta}_s \geq 0$ を満たす．MKL では，以下の式 (3) のように，$\boldsymbol{\alpha}$ と $\boldsymbol{\beta} = \{\beta_s\}_{s=1}^S$ の双方に対して損失関数を最適化する必要がある．

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left[\sum_{i=1}^{N} \mathcal{L}\left(y_i, \sum_{j=1}^{N} \alpha_j \sum_{s=1}^{S} \beta_s [K_s]_{ij} \right) + ||\boldsymbol{\alpha}||_K^2 \right] \tag{3}$$

従来の MKL の多くは，マージン最大化基準に基づいて $\boldsymbol{\alpha}$ と $\boldsymbol{\beta}$ を最適化するアプローチを採用している．最近では，Do らが，SVM の汎化誤差がマージンのみならず特徴空間において全ての学習データを包含する最小の球のラディウス (半径) にも依存すること，および，カーネル関数を結合することで特徴空間におけるデータの分布も変化することを考慮して，マージンとラディウスの双方を最適化する MKL の枠組みである RMKL[4] を提案した．RMKL は，計算コストの面で SimpleMKL[10] と同等に効率的であり，SDP-MKL[3] よりも高精度である．したがって，本研究では，比較対象とするマージン最大化に基づく MKL として，RMKL を採用する．

3. 条件付きエントロピー最小化に基づくマルチカーネル学習

本章では，条件付きエントロピー最小化に基づくマルチカーネル学習 (Multiple kernel learning based on Conditional Entropy Minimization; MCEM)[6] について概説する．

3.1 条件付きエントロピー最小化基準

MCEM は，以下のように定式化される．

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left[H\Big(f(\boldsymbol{X}; \boldsymbol{\alpha}, \boldsymbol{\beta}) | Y \Big) \right] \tag{4}$$
$$\text{s. t.} \quad H\Big(f(\boldsymbol{X}; \boldsymbol{\alpha}, \boldsymbol{\beta}) \Big) = const., \quad \sum_{s=1}^{S} \beta_s = 1, \quad \beta_s \geq 0$$

ここで，$f(\boldsymbol{x}; \boldsymbol{\alpha}, \boldsymbol{\beta})$ は判別関数であり，パラメータ $\boldsymbol{\alpha}$ と結合係数 $\boldsymbol{\beta}$ に依存して決まることを陽に表している．また，クラス条件付きエントロピー $H(f(\boldsymbol{X}; \boldsymbol{\alpha}, \boldsymbol{\beta})|Y)$ を最小化するにあたり，自明な解や過学習を避けるため，全データに対するエントロピー $H(f(\boldsymbol{X}; \boldsymbol{\alpha}, \boldsymbol{\beta}))$ を正則化項として用いる．このとき，MCEM は，$\eta > 0$ を用いて，以下のように書き直すことができる．

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left[H\Big(f(\boldsymbol{X}; \boldsymbol{\alpha}, \boldsymbol{\beta}) | Y \Big) - \eta \cdot H\Big(f(\boldsymbol{X}; \boldsymbol{\alpha}, \boldsymbol{\beta}) \Big) \right] \quad \text{s. t.} \quad \sum_{s=1}^{S} \beta_s = 1, \quad \beta_s \geq 0 \tag{5}$$

式 (5) における目的関数 $H(f(\boldsymbol{X}; \boldsymbol{\alpha}, \boldsymbol{\beta})|Y) - \eta \cdot H(f(\boldsymbol{X}; \boldsymbol{\alpha}, \boldsymbol{\beta}))$ の各項に着目し，最適化を情報論的に解釈すると，以下のように言える．

第 1 項 クラス条件付きエントロピー $H(f(\boldsymbol{X}; \boldsymbol{\alpha}, \boldsymbol{\beta})|Y)$ の最小化は，データを同一クラスごとに密集させることを意味する．

第 2 項 全データに対する負のエントロピー $-H(f(\boldsymbol{X}; \boldsymbol{\alpha}, \boldsymbol{\beta}))$ の最小化，もしくは $H(f(\boldsymbol{X}; \boldsymbol{\alpha}, \boldsymbol{\beta}))$ の最大化は，データを，属するクラスに無関係に散らばらせることを意味する．

Vol.2011-SLP-87 No.3
2011/7/21

以上を整理すると，MCEM に基づく最適化は，同一クラス内でのデータの散らばりを最小化し，異なるクラス間でのデータの散らばりを最大化しているとみなすことができる．それに対し，従来のカーネルマシンにおいて用いられるマージン最大化基準は，異なるクラス間のデータの分離度にのみ着目した基準であり，クラス内のデータの分離度については陽に考慮していない．したがって，MCEM に基づく話者照合システムは，従来のマージン最大化基準に基づくシステムと比較して，同一話者内の特徴変動に対して頑健である可能性がある．

いま，$\boldsymbol{V}_w$ をクラス内分散行列とすると，クラス条件付きエントロピー $H(f(\boldsymbol{X})|Y)$ の上界について，以下の不等式 (6) が成立する[6]．

$$H(f(\boldsymbol{X})|Y) \leq \log(2\pi)^{\frac{1}{2}} e + \frac{1}{2}\log(\boldsymbol{\alpha}^T \boldsymbol{V}_w \boldsymbol{\alpha}) \tag{6}$$

このとき，カーネルフィッシャー判別分析 (kernel Fisher discriminant analysis; KFDA)[11] の目的関数は，式 (6) の右辺と本質的に等価である．したがって，KFDA は，写像された軸上におけるデータの値 $f(\boldsymbol{x})$ を用いて計算したクラス条件付きエントロピーの上界を最小化する教師あり次元圧縮とみなすことができる．以上より，MCEM では，$H(f(\boldsymbol{X})|Y)$ を最小化する写像 f を求めるために，KFDA を近似的に用いる．

3.2 最適化アルゴリズム

MCEM の最適化アルゴリズムを**図 1** に示す．式 (4) あるいは式 (5) を $\boldsymbol{\alpha}$ と $\boldsymbol{\beta}$ の双方に対して同時に最適化するのは困難であるため，繰り返し最適化アルゴリズムを適用した．つまり，$\boldsymbol{\beta}$ の値を固定して $\boldsymbol{\alpha}$ の最適化を行い，推定された $\boldsymbol{\alpha}$ を用いて $\boldsymbol{\beta}$ を最適化する，という処理を繰り返す．この繰り返し最適化アルゴリズムの概念図を**図 2** に示す． ここで，t 回目の反復処理における $\boldsymbol{\alpha}$ と $\boldsymbol{\beta}$ の推定値を各々，$\boldsymbol{\alpha}^{(t)}$，$\boldsymbol{\beta}^{(t)}$ と書く．

$\boldsymbol{\alpha}$ の最適化は，KFDA を用いて行う．式 (6) より，$\boldsymbol{\beta}$ を固定したときの $\boldsymbol{\alpha}$ の最適解は，クラス条件付きエントロピーの上界の最小化として，KFDA により求めることができる．そのような $\boldsymbol{\alpha}$ の最適化は，式 (7) のように表される．$\boldsymbol{\beta}$ は，式 (8) により最適化される．ここでは，正則化項付きのクラス条件付きエントロピーを最小にする $\boldsymbol{\beta}$ を，ランダムサーチを用いて推定した[6]．このアルゴリズムでは，まず，平均ベクトルが $\boldsymbol{\beta}^{(t-1)}$，分散行列が単位行列であるガウス分布から P 個の $\boldsymbol{\beta}$ 候補をサンプリングする．そして，得られた $\{\boldsymbol{\beta}_p\}_{p=1}^{P}$ を用いて式 (8) の目的関数 $H(f(\boldsymbol{X};\boldsymbol{\alpha}^{(t)},\boldsymbol{\beta}_p)|Y) - \eta \cdot H(f(\boldsymbol{X};\boldsymbol{\alpha}^{(t)},\boldsymbol{\beta}_p))$ を計算し，この目的関数の値を最小にするような $\boldsymbol{\beta}$ を選択した．本研究では，エントロピーおよびクラス条件付きエントロピーは平均近傍 (mean nearest neighbor; MNN) 法[12] を用いて

MCEM: Multiple kernel learning algorithm based on conditional entropy minimization

入力: 学習データ $D=\{\boldsymbol{x}_i\}_{i=1}^{N}$，$\boldsymbol{x}_i \in \mathbb{R}^n$ とそのクラスラベル $\{y_i\}_{i=1}^{N}$，$y_i \in \{\pm 1\}$，S 個のカーネル関数 $\{k_s(\boldsymbol{x}_i,\boldsymbol{x})\}_{s=1}^{S}$，KFDA のための正則化パラメータ ζ.

初期化: D を使ってカーネル行列 $\{K_s\}_{s=1}^{S}$ を算出する．$\sum_{s=1}^{S}\beta_s^{(0)}=1$，$\boldsymbol{\beta}_s^{(0)} \geq 0$ を満たすような乱数によってカーネル結合係数 $\boldsymbol{\beta}^{(0)}=\{\beta_s^{(0)}\}_{s=1}^{S}$ を初期化する．

繰り返し: $t=1$ より収束するまで以下を繰り返す:

$\boldsymbol{\alpha}$ の最適化ステップ: KFDA の最小化問題を $\boldsymbol{\beta}^{(t-1)}$ を固定して解き，$\boldsymbol{\alpha}^{(t)}$ を得る:

$$\min_{\boldsymbol{\alpha}} \left[\boldsymbol{\alpha}^T \Big(\boldsymbol{V}_w(\boldsymbol{\beta}^{(t-1)}) + \zeta K \Big) \boldsymbol{\alpha} \right] \quad \text{s. t.} \quad \boldsymbol{\alpha}^T \boldsymbol{V}_b \boldsymbol{\alpha} = const. \tag{7}$$

$\boldsymbol{\beta}$ の最適化ステップ: 判別関数 $f(\boldsymbol{x};\boldsymbol{\alpha}^{(t)},\boldsymbol{\beta})$ の条件付きエントロピーを $\boldsymbol{\alpha}^{(t)}$ を固定した上で最小化し，$\boldsymbol{\beta}^{(t)}$ を得る:

$$\min_{\boldsymbol{\beta}} \left[H\Big(f(\boldsymbol{X};\boldsymbol{\alpha}^{(t)},\boldsymbol{\beta})|Y\Big) - \eta \cdot H\Big(f(\boldsymbol{X};\boldsymbol{\alpha}^{(t)},\boldsymbol{\beta})\Big) \right] \tag{8}$$

$$\text{s. t.} \quad \sum_{s=1}^{S} \boldsymbol{\beta}_s = 1, \quad \boldsymbol{\beta}_s \geq 0$$

出力: 収束したパラメタ $\boldsymbol{\alpha}$ と $\boldsymbol{\beta}$．これらのパラメタを用いて計算した以下の判別関数:

$$f(\boldsymbol{x};\boldsymbol{\alpha},\boldsymbol{\beta}) = \sum_{i=1}^{N}\alpha_i \sum_{s=1}^{S}\beta_s k_s(\boldsymbol{x}_i,\boldsymbol{x}). \tag{9}$$

図 1 MCEM アルゴリズム: クラス条件付きエントロピー最小化に基づくマルチカーネル学習アルゴリズム．$\boldsymbol{V}_w$ および $\boldsymbol{V}_b$ は各々，特徴空間におけるクラス内分散行列とクラス間分散行列を表す．

Fig. 1 MCEM: Multiple kernel learning algorithm based on conditional entropy minimization. $\boldsymbol{V}_w$ and $\boldsymbol{V}_b$ denote the within-class cavariance matrix and between-class covariance matrix, respectively, in the feature space.

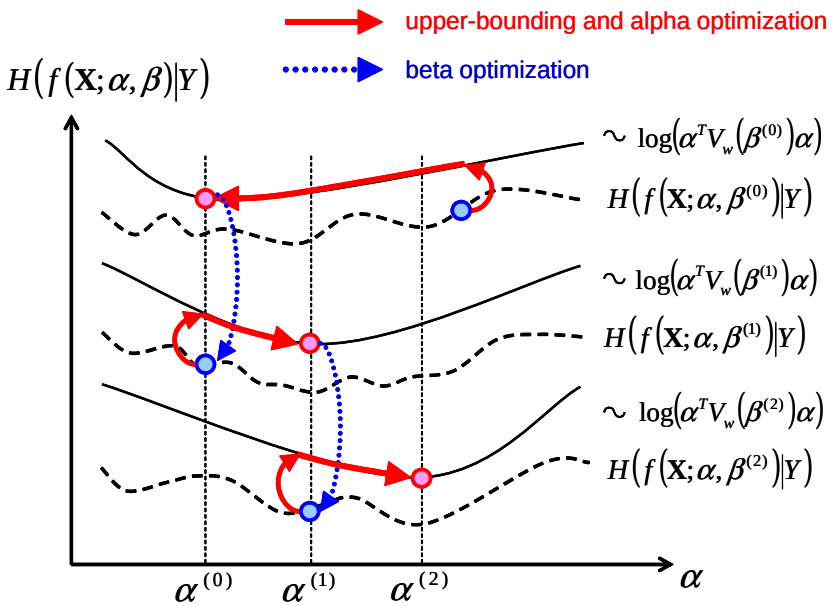

図 2 MCEM アルゴリズムにおける繰り返し最適化の概念図．点線は条件付きエントロピー，実線は条件付きエントロピーの上界 (KFDA の目的関数) を表す．

Fig. 2 Conceptual image of iterative optimization in MCEM algorithm. Dotted lines express class-conditional entropy, and solid lines express upper bounds of class-conditional entropy (i.e., objective function in KFDA).

計算した．

4. 話者照合実験

提案手法の有効性を調査するために，テキスト独立型話者照合実験を行った．本実験では，同一話者内で音響特徴に変動が生じる状況として，同一話者が異なる発話スタイルで発声した音声を扱う場合を考える．ここでは，異なる発話スタイルの音声として，通常発声音声と，騒音下で発話した際に生じる Lombard 効果を含む音声 (Lombard 発声音声) を用いる．

4.1 実験条件

4.1.1 評価項目

RMKL[4] を用いて構築した SVM に基づく話者照合システムと，MCEM を用いて構築した SVM に基づく話者照合システムの性能を比較した．また，本実験では，以下の 2 条件に対して評価を行った．

(1) 同一話者内で発話スタイル変動が生じていない場合: 通常発声音声か Lombard 発声音声のどちらかのみを話者モデルの学習データとして用いる．

(2) 同一話者内で発話スタイル変動が生じている場合: 通常発声音声と Lombard 発声音声の両方を話者モデルの学習データとして用いる．このとき，話者クラス内での音響特徴変動をモデル内部で扱う必要がある．

表 1 被験者に提示された騒音の種類と音圧

Table 1 Types and SPLs of noise exposed to subjects.

Notation	Type of noise	SPL (dB(A))
car	in-car noise	60
dep	in-department-store noise	60, 70, 75
pin	pink noise	60, 70, 75

4.1.2 音声コーパス

各話者の発話スタイル変動が話者照合システムの性能に与える影響を調査するために，日本語 Lombard 発声音声コーパス[13] を用いた．本コーパスは，通常発声のクリーン音声 (neutral-clean speech; NC) と，Lombard 発声のクリーン音声 (Lombard-clean speech; LC) を含む．LC は，女性 20 名と男性 20 名の被験者 (発話者) が，ヘッドホンを通じて騒音を提示された状態で発話した音声を，発話者の口元に置かれた指向性マイクロホンにより収録したものである．被験者が提示された騒音の種類と音圧を**表 1** に示す． 被験者に対し提示した騒音は，ATR 環境音データベースに含まれる自動車内走行音，デパート地下食品売り場騒音，ピンクノイズの 3 種類である．

本コーパスには 20 種類の評価セットが用意されている．各セットは，連続数字 10 発話と音素連鎖バランス文 10 発話から成る．連続数字は，CENSREC-1[14] において採用された 11 種類の数字 (「イチ」,「二」,「サン」,「ヨン」,「ゴ」,「ロク」,「ナナ」,「ハチ」,「キュー」,「ゼロ」,「マル」) から成る．40 名の発話者は各々，評価セットのうち 1 セット，つまり 20 発話を，表 1 に示された全ての騒音条件において発話した．

4.1.3 音響特徴抽出

音響特徴パラメータとして，MFCC 12 次元，ΔMFCC 12 次元, ΔΔ MFCC 12 次元から成る計 36 次元のパラメータを用いた．音響特徴抽出条件を**表 2** に示す．

Vol.2011-SLP-87 No.3
2011/7/21

表 2 音響特徴抽出条件
Table 2 Experimental conditions for acoustic feature extraction.

sampling frequency	16 kHz
frame length	25 ms
frame shift	10 ms
analysis window	Hamming window
pre-emphasis	$1-0.97z^{-1}$

表 3 本実験で用いたカーネル関数とパラメータ
Table 3 Kernel functions and parameters used in this study.

	$k(\boldsymbol{x},\boldsymbol{x}')$	parameter
RBF	$\exp(-\sigma\|\|\boldsymbol{x}-\boldsymbol{x}'\|\|^2)$	$\sigma=0.01, 0.05, 0.1$
polynomial	$(\boldsymbol{x}^T\boldsymbol{x}'+1)^d$	$d=2,3$
linear	$\boldsymbol{x}^T\boldsymbol{x}'$	—

4.1.4 SVM に基づく識別器

異なるフレーム長を持つ音声データに対してカーネル関数を計算するために，系列カーネル[15] を用いた．発話音声 $\boldsymbol{\mathcal{X}}=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_T)\in\mathbb{R}^{n\times T}$ および $\boldsymbol{\mathcal{X}}'=(\boldsymbol{x}'_1,\cdots,\boldsymbol{x}'_{T'})\in\mathbb{R}^{n\times T'}$ に対して，系列カーネルは以下のように定義される．このとき，$n=36$ である．

$$\mathcal{K}(\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{X}}')=\frac{1}{T\cdot T'}\sum_{t=1}^{T}\sum_{t'=1}^{T'}k(\boldsymbol{x}_t,\boldsymbol{x}'_{t'}). \tag{10}$$

式 (10) における $k(\boldsymbol{x}_t,\boldsymbol{x}'_{t'})$ は，**表 3** に示した 3 種類の RBF カーネル，2 種類の多項式カーネル，および線型カーネルである． 式 (7) における正則化パラメータ ζ は，全ての実験において $\zeta=0.001$ とした．RMKL ではソフトマージン SVM を構築するが，その際のソフトマージンパラメターは 1 とした．MCEM では，トレードオフパラメータとして $\eta=0.5$ を用いた．また，$\boldsymbol{\beta}$ の最適化ステップでは，$\boldsymbol{\beta}$ を 500 候補 ($P=500$) サンプリングした．以上のパラメータは，予備実験において最良の性能を与えたものである．また，$\boldsymbol{\alpha}$ と $\boldsymbol{\beta}$ を最適化するための繰り返し処理は，条件付きエントロピーの減少率が 0.0001 を下回ったとき，あるいは，条件付きエントロピーが増大したときに停止した．

判別は，最適な写像 f，つまり $\boldsymbol{\alpha}$ と $\boldsymbol{\beta}$ の最適解が得られたとき，式 (9) の出力値 (1 次元の判別軸への写像) を用いて行った．

表 4 発話スタイルに関する様々な条件における RMKL に基づく SVM と MCEM に基づく SVM を用いて構築した話者照合システムの Equal error rate (EER) (%)．“Improve.” は RMKL に基づくシステムの EER に対する MCEM に基づくシステムの EER の改善率 (%) を表す．
Table 4 Equal error rate (EER) (%) for the speaker verification systems using the RMKL-based SVM and MCEM-based SVM under various conditions in the speaking styles. “Improve.” is represented by an EER improvement rate (%) produced by the MCEM-based system compared to that produced by the RMKL-based one.

Speaking style		RMKL	MCEM	
Training data	Test data	EER (%)	EER (%)	Improve. (%)
NC	NC	6.25	5.86	6.24
pinLC75	pinLC75	8.00	7.67	4.13
Average		7.13	6.77	5.18
NC + pinLC75	NC	9.25	8.14	12.0
	carLC60	9.97	9.50	4.71
	pinLC60	9.75	8.75	10.26
	pinLC70	8.44	7.25	14.10
	pinLC75	8.72	8.25	5.39
	depLC60	9.25	8.06	12.86
	depLC70	9.25	8.50	8.11
	depLC75	9.83	9.00	8.44
Average		9.31	8.43	9.48

4.2 実験結果

表 4 に，学習データと評価データの発話スタイルと，対応する条件において RMKL に基づく SVM を用いて構築した話者照合システムと MCEM に基づく SVM を用いて構築した話者照合システムが与える equal error rate (EER) を示す．この表には，RMKL に基づくシステムの代わりに MCEM に基づくシステムを使用したときの EER の改善率 (“Improve.”) も示した．ここで，“NC” は，通常発声のクリーン音声を表し，“pinLC75” は，耳元で 75 dB(A) のピンクノイズ (pin) をヘッドホンを通じて提示された状態で被験者が発話した Lombard 発声のクリーン音声 (LC) を表す．“NC + pinLC75” は，通常発声と Lombard 発声の両方を含むことを表す．つまり，このときの学習データは，同一話者クラス内の特徴変動を陽に含んでいると言える．

本実験では，40 名の話者 (女性 20 名，男性 20 名) に対する 4 分割交差検定により評価を行った．つまり，各分割における評価対象話者は 10 名 (女性 5 名，男性 5 名) であり，詐称者モデルの学習は残りの 30 名 (女性 15 名，男性 15 名) のデータを用いてを行う．このときの各分割における学習データと評価データに関する条件は以下の通りである．

- 詐称者モデルの学習データは，30 名の話者 (女性 15 名，男性 15 名) が発話した音素バランス文であり，発話スタイルごとに計 300 発話である．このとき，学習データとして “NC + pinLC75” を用いた場合の詐称者モデルの学習データは，発話スタイル 2 種類分の計 600 発話となる．
- 目的話者は，全話者 40 名から詐称者モデルの学習に用いた 30 名を差し引いた，残り 10 名の話者 (女性 5 名，男性 5 名) のうち 1 名である．つまり，各分割ごとに 10 名分の目的話者モデルが構築される．目的話者モデルの学習データは，目的話者が発話した音素バランス文であり，発話スタイルごとに 10 発話である．このとき，学習データとして “NC＋pinLC75” を用いた場合の目的話者モデルの学習データは計 20 発話となる．
- 評価データは，詐称者モデルの学習に用いていない 10 名の話者 (女性 5 名，男性 5 名) が発話した連続数字発話であり，発話スタイルごとに計 100 発話である．このとき，10 名のうち 1 名の音声が目的話者として，9 名の音声が詐称者として識別されることが理想的である．本実験は，話者モデルの学習に音素連鎖バランス文が，評価に連続数字が用いられていることから，テキスト独立の条件で行われていることになる．

最終的な話者照合システムの EER は，4 交差検定で得られた結果を統合することで算出した．つまり，合計 40 話者について評価したことになる．

表 4 に示した実験結果より，MCEM に基づくシステムは，発話スタイルに関する条件に依らず，RMKL に基づくシステムを上回る性能を与えた．さらに，発話スタイル変動を陽に含む音声から学習した話者モデル (“NC＋pinLC75”) を用いたほとんどの場合 (EER 改善率: 平均 9.5%) において，発話スタイル変動を陽に含まない音声から学習した話者モデル (“NC” あるいは “pinLC75”) を用いた場合 (EER 改善率: 平均 5.2%) よりも高い EER 改善率を与えた．以上より，MCEM は，従来のマージン最大化に基づく MKL と比較して，話者照合において良好な性能を与えるとともに，MCEM に基づく話者照合システムは，マージン最大化の枠組みを用いて構築された従来のシステムと比較して，話者クラス内で特徴変動が生じる場合において，より有効であることが明らかになった．

5. ま と め

条件付きエントロピー最小化に基づくマルチカーネル学習 (MCEM) を話者照合に適用した．本手法は，従来のマージン最大化に基づく枠組みと比較して，同一話者内の特徴変動が話者照合システムの性能に与える影響を低減するのに効果的であった．話者照合実験の結果，RMKL に基づくシステムの代わりに MCEM に基づくシステムを用いたときの EER の改善率は，同一話者内での特徴変動を含まないデータを用いて学習した場合で 5.3%，同一話者内での特徴変動を陽に含むデータを用いて学習した場合で 9.5% であった．

謝辞 本研究の一部は文部科学省の科研費 (22800067) の助成を受けたものである．

参 考 文 献

1) W.M. Campbell *et al.*, “Support vector machines using GMM supervectors for speaker verification,” IEEE Sign. Process. Lett., vol.13, no.5, pp.308–311, May 2006.
2) A. Stolcke *et al.*, “MLLR transforms as features in speaker recognition,” Proc. INTERSPEECH, pp. 2425–2428, Sept. 2005.
3) G. R. G. Lanckriet *et al.*, “Learning the kernel matrix with semidefinite programming,” JMLR, vol.5, pp.27–72, 2004.
4) H. Do *et al.*, “Margin and radius based multiple kernel learning,” Proc. ECML, pp.330–343, Sept. 2009.
5) C. Longworth *et al.*, “Multiple kernel learning for speaker verification,” Proc. ICASSP, pp.1581–1584, March 2008.
6) H. Hino *et al.*, “Multiple kernel learning by conditional entropy minimization,” Proc. ICMLA, Dec. 2010.
7) T. Ogawa *et al.*, “Speaker recognition using multiple kernel learning based on conditional entropy minimization,” Proc. ICASSP, pp.2204–2207, May 2011.
8) H. Aronowitz *et al.*, “Modeling intra-speaker variability for speaker recognition,” Proc. Interspeech, pp.2177–2180, Sept. 2005.
9) R. Vogt *et al.*, “Factor analysis subspace estimation for speaker verification with short utterances,” Proc. Interspeech, pp.853–856, Sept. 2008.
10) A. Rakotomamonjy *et al.*, “SimpleMKL,” JMLR, vol.9, pp.2491–2521, 2008.
11) S. Mika *et al.*, “Fisher discriminant analysis with kernels,” Proc. NNSP, pp.41–48, Aug. 1999.
12) L. Faivishevsky *et al.*, “ICA based on a smooth estimation of the differential entropy,” Proc. NIPS, pp.433–440, Dec. 2009.
13) T. Ogawa *et al.*, “Development and evaluation of Japanese Lombard speech corpus,” Proc. INTERNOISE, Sept. 2011 (to appear).
14) S. Nakamura et al., “AURORA-2J: An evaluation framework for Japanese noisy speech recognition,” IEICE Trans. Inf. & Syst., vol. E88-D, no. 3, pp.535–544, March 2005.
15) J. Mariethoz *et al.*, “A kernel trick for sequences applied to text-independent speaker verification systems,” Pattern Recognition, vol.40, no.8, pp.2315–2324, 2007.