統計数理(2005)
第 53 巻 第 2 号 201–210

特集「計算推論—モデリング・数理・アルゴリズム—」
[研究詳解]

dPLRM を用いた話者識別

松井 知子[1] ・田邉 國士[2]

(受付 2005 年 3 月 14 日; 改訂 2005 年 9 月 21 日)

要　　旨

本稿では帰納的学習機械，dual Penalized Logistic Regression Machine (dPLRM)を用いたテキスト独立型話者識別法について報告する．dPLRM を利用して，学習データだけから識別的な話者特徴を捉えることを試みる．本方法では，従来の事前知識に基づく特徴抽出処理を必要としない．話者 10 名が異なる 3 時期に発声した音声による識別実験において，256 次元の対数パワースペクトルを直接用いた dPLRM 法は，26 次元のメルケプストラム(Mel-frequency cepstral coefficient; MFCC)を用いた混合ガウス分布モデル(Gaussian mixture model; GMM)に基づく従来法と比べて，同等以上の性能であることを示す．また，特に学習データ量が少ない場合には，dPLRM 法は GMM 法よりも識別率が高いことを示す．

キーワード：話者認識，話者識別，カーネル回帰，dual Penalized Logistic Regression Machine, 帰納的学習機械，特徴抽出．

1. はじめに

近年，音声情報サービスのアクセスコントロールや音声アーカイブ中の話者検索などのアプリケーションにおいて，話者認識技術の需要が増加している．代表的な話者認識研究プロジェクトとしては，世界各国の複数の研究機関が参加する NIST 評価が挙げられる(NIST, 1996)．NIST 評価では，テキスト独立型の話者認識法について検討を進めているが，数十次元の MFCC を用いた GMM に基づく方法が一般に採用されている．GMM は話者ごとに用意し，そのパラメータは各話者の学習データから推定する．

MFCC は，周波数軸を人間の聴覚の特性を考慮したメルスケールに変換してからケプストラム分析(対数パワースペクトルの逆フーリエ変換)を行うことによって抽出される(Stevens, 1975)．音声・話者認識では，発声変動(同じ言葉を発声しても物理的性質が変動する)に対する頑健性の向上が課題の一つとなっている(古井, 1985)．10 ∼ 20 次元の低次の MFCC は発声変動に(ある程度は)頑健な特徴量であることが経験的に知られており，音声・話者認識において広く利用されている(Murthy et al., 1999)．低次の係数だけを選択する処理は，GMM のパラメータ推定の安定性にも関連しているが，高次に含まれる話者識別に有効な情報を捨てている可能性がある．

最近，田邉は特別のペナルティ項を伴った罰金付きロジスティックモデルに基づく帰納的学習機械，dPLRM を提案した (Tanabe, 2001a, 2001b, 2003a)．このモデルはカーネル回帰子による双対性を有し，学習データだけに基づいて，識別に有効な特徴を捉えることができる．

[1] 統計数理研究所：〒106–8569 東京都港区南麻布 4–6–7; tmatsui@ism.ac.jp
[2] 早稲田大学大学院 理工学研究科：〒169–0072 東京都新宿区大久保 3–14–9; tanabe.kunio@waseda.jp

本稿では，dPLRM を利用して，メルスケールなどの事前知識に基づく処理を行うことなく，256 次元の対数パワースペクトルを直接用いる話者識別法について紹介する（Matsui and Tanabe, 2004a）．なお，MFCC を用いた場合，dPLRM による方法は従来の GMM やサポートベクターマシンによる方法と同等の性能を示すことはわかっている（Matsui and Tanabe, 2004b, 2004c）．GMM による方法は話者ごとに密度関数を推定するが，各話者の特徴を学習するのに，比較的多量のデータを必要とする．一方，dPLRM ではカーネル関数により非線形性の扱いに優れ，また識別的な学習を行うために，比較的少量の学習データから各話者の特徴を捉えることができる．dPLRM はロジスティック回帰機械の双対機械として，学習データ中の（隠れた）構造を幅広く表現することができ，非常に高い帰納力を有している（Tanabe, 2001a, 2001b, 2003a）．図 1 に dPLRM による本方法と GMM による方法の話者識別の手続きを示す．

次節では，dPLRM についてその概要を説明する．3 節では話者識別実験の結果を示す．異なる時期に発声された学習データを用いた場合，対数パワースペクトルを用いた dPLRM による方法は，MFCC を用いた GMM による方法よりも，特徴抽出における処理のいくつかを省いているにも拘らず，性能が高いことを示す．また，学習データを間引いた場合についても実験結果を示す．4 節では，話者判定のための確率値の推定について考察する．

2. dPLRM による話者識別

2.1 dPLRM

今，x_j をデータを表す n 次元の列ベクトル，c_j をクラスを表すスカラー $(\in \{1,2,\ldots,K\})$ とする．dPLRM では，有限個の学習データセット $\{(\boldsymbol{x}_j, c_j)\}_{j=1,\ldots,N}$ を入力して，$\boldsymbol{x} \in R^n$ に対するクラス c の条件付き多項分布 $M(\boldsymbol{p}^*(\boldsymbol{x}))$ を生成する．ここで $\boldsymbol{p}^*(\boldsymbol{x})$ は予測確率ベクトルで，

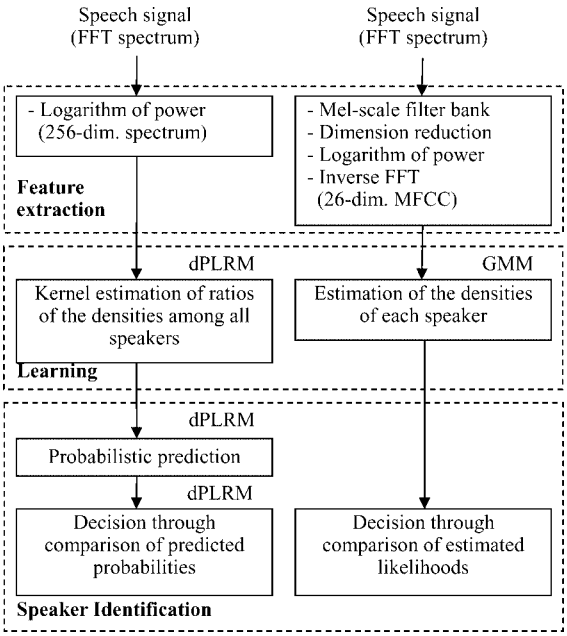

図 1. The dPLRM method (the left-hand side) vs. the GMM-based method (the right-hand side).

その第 k 要素 $p_k^*(\boldsymbol{x})$ は $\boldsymbol{x}$ のクラス c が k の値をとる確率を示す.(なお，“*” は推定量を表す.)

便宜上，クラスデータ c_j を第 j 要素が 1 の K 次元の単位ベクトル $\boldsymbol{e}_k \equiv (0,\ldots,1,\ldots,0)^t$ を用いてコード化することにより，$K \times N$ の定数行列 $\boldsymbol{Y}$ を次のように定義する.

$$\boldsymbol{Y} \equiv [\boldsymbol{y}_1;\ldots;\boldsymbol{y}_N] \equiv [\boldsymbol{e}_{c_1};\ldots;\boldsymbol{e}_{c_N}] \tag{2.1}$$

ここで第 j 列ベクトル $\boldsymbol{y}_j \equiv \boldsymbol{e}_{c_j}$ は $\boldsymbol{x}_j$ のクラスを表す.

今，次の R^N から R^K への写像を導入する.

$$\boldsymbol{F}(\boldsymbol{x}) \equiv \boldsymbol{V}\boldsymbol{k}(\boldsymbol{x}) \tag{2.2}$$

ここで $\boldsymbol{V}$ は $K \times N$ のパラメータ行列で，学習データセットから推定する．$\boldsymbol{k}(\boldsymbol{x})$ は次の R^n から R^N への写像である.

$$\boldsymbol{k}(\boldsymbol{x}) \equiv (\boldsymbol{K}(\boldsymbol{x}_1,\boldsymbol{x}),\ldots,\boldsymbol{K}(\boldsymbol{x}_N,\boldsymbol{x}))^t \tag{2.3}$$

ここで $\boldsymbol{K}(\boldsymbol{x},\boldsymbol{x}')$ は任意の正定値カーネル関数を表す．予測確率 $p(x)$ を次の多項モデルで定義する.

$$\boldsymbol{p}(\boldsymbol{x}) \equiv \hat{\boldsymbol{p}}(\boldsymbol{F}(\boldsymbol{x})) \equiv (\hat{p}_1(\boldsymbol{F}(\boldsymbol{x})),\ldots,\hat{p}_K(\boldsymbol{F}(\boldsymbol{x})))^t \tag{2.4}$$

ここで $\hat{p}_k(\boldsymbol{F}(x)) \equiv \dfrac{\exp(\boldsymbol{F}_k(\boldsymbol{x}))}{\sum_{i=1}^K \exp(\boldsymbol{F}_i(\boldsymbol{x}))}$ はロジスティック変換である.

上記モデルを仮定した場合，$\boldsymbol{p}(\boldsymbol{x})$ の負の対数尤度関数 $L(\boldsymbol{V})$ は次式の凸関数で与えられる.

$$L(\boldsymbol{V}) \equiv -\sum_{j=1}^N \log(p_{c_j}(\boldsymbol{x}_j)) = -\sum_{j=1}^N \log(\hat{p}_{c_j}(\boldsymbol{V}\boldsymbol{k}(\boldsymbol{x}_j))) \tag{2.5}$$

この目的関数 $L(\boldsymbol{V})$ は全クラスのデータを一度に用いて定義されるため，識別的な性質を持ち，カーネル関数を適切に選べば，$\boldsymbol{F}(\boldsymbol{x})$ は多様な関数を表せるので，予測確率 $\boldsymbol{p}(\boldsymbol{x})$ は真のものに近くなる．予測確率 $\boldsymbol{p}^*(\boldsymbol{x})$ は，$L(\boldsymbol{V})$ を最小化する V の ML 推定量 $\boldsymbol{V}^{**}$ $(= \arg\max_{\boldsymbol{V}} -L(\boldsymbol{V}))$ を用い，$\boldsymbol{p}^*(\boldsymbol{x}) = \hat{\boldsymbol{p}}(\boldsymbol{V}^{**}\boldsymbol{k}(\boldsymbol{x}))$ より得る.

しかしながら，学習データが有限であるので，$\boldsymbol{V}^{**}$ に関しては過学習の問題が生じる場合がある．そのために，ペナルティ項を導入し，次のペナルティ付きの負の対数尤度を最小化する $\boldsymbol{V}^*$ $(= \arg\max_{\boldsymbol{V}} -PL(\boldsymbol{V}))$ を推定値として用いる.

$$PL(\boldsymbol{V}) \equiv L(\boldsymbol{V}) + \frac{\delta}{2}\|\boldsymbol{\Gamma}^{\frac{1}{2}}\boldsymbol{V}\overline{\boldsymbol{K}}^{\frac{1}{2}}\|_F^2 \tag{2.6}$$

ここで $\|\cdot\|_F$ はフロベニウスノルムである．このペナルティ項により，V の有効な自由度を調整できる．$\boldsymbol{\Gamma}$ は任意の $K \times K$ の正定値行列である．なお，$\boldsymbol{\Gamma}$ は学習データのクラスによる偏りを補正するように，しばしば次式で与えられる.

$$\boldsymbol{\Gamma} = \frac{1}{N}\boldsymbol{Y}\boldsymbol{Y}^t \tag{2.7}$$

$\overline{\boldsymbol{K}}$ は $N \times N$ 定数行列で次式で与えられる.

$$\overline{\boldsymbol{K}} = [\boldsymbol{K}(\boldsymbol{x}_i,\boldsymbol{x}_j)]_{i,j=1,\ldots,N} \tag{2.8}$$

δ は正則化パラメータで，経験ベイズ法によって決定することができる(Tanabe, 2001a).

(2.6)式の二次のペナルティ項を導入することにより，V^* は次の行列の方等式の解として与えられる.

$$\nabla PL \equiv (\boldsymbol{P}(\boldsymbol{V}) - \boldsymbol{Y} + \delta\boldsymbol{\Gamma}\boldsymbol{V})\overline{\boldsymbol{K}} = \boldsymbol{O}_{K,N} \tag{2.9}$$

ここで $P(V)$ は第 j 列ベクトルが確率ベクトル $\boldsymbol{p}(\boldsymbol{x}_j) \equiv \hat{p}(V\boldsymbol{k}(\boldsymbol{x}_j))$ の $K \times N$ 行列である．行列 Y は(2.1)式で与えられる．予測確率 $p^*(\boldsymbol{x}) \equiv \hat{p}(V^*\boldsymbol{k}(\boldsymbol{x}))$ を与える V^* は，次のアルゴリズムにより繰り返し計算で求める．

アルゴリズム：初期値を V^0（$K \times N$ 行列）とする．$\{V^i\}$ は次式に従って計算する．

$$\boldsymbol{V}^{i+1} = \boldsymbol{V}^i - \alpha_i \Delta \boldsymbol{V}^i, \quad i = 0, \ldots, \infty \tag{2.10}$$

ここで ΔV^i は次の行列の線形等式の解である．

$$\sum_{j=1}^{N} ([\boldsymbol{p}(\boldsymbol{x}_j)] - \boldsymbol{p}(\boldsymbol{x}_j)(\boldsymbol{p}(\boldsymbol{x}_j))^t)\Delta \boldsymbol{V}^j(\boldsymbol{k}(\boldsymbol{x}_j)(\boldsymbol{k}(\boldsymbol{x}_j))^t) + \delta \boldsymbol{\Gamma} \Delta \boldsymbol{V}^i \overline{\boldsymbol{K}} = (\boldsymbol{P}(\boldsymbol{V}^i) - \boldsymbol{Y} + \delta \boldsymbol{\Gamma} \boldsymbol{V}^i)\overline{\boldsymbol{K}} \tag{2.11}$$

上記アルゴリズムの詳細については文献(Tanabe, 2001a, 2001b, 2003a)を参照されたい．ここで，狭義凸関数 $PL(V)$ の制約なしの最適化問題を解くことが，(2.9)式の簡単な行列の非線形等式を解くことと同義になることに注目されたい．

2.2 話者識別の手続き

図 2，3 に学習とテストの手続きを示す．

2.3 多項式カーネル関数の表現力

本稿の実験では，カーネル関数として多項式カーネル関数を用いた．本節では，多項式カーネル関数の表現力について，dPLRM とある意味で等価である PLRM (Penalized Logistic Regression Machine)を通して説明する．dPLRM と PLRM は互いに双対の関係にある機械である．前節で予め与えた写像 $\boldsymbol{F}(\boldsymbol{x})$ は，PLRM の枠組みでは次式のように表される（Tanabe, 2001a, 2001b, 2003a）．

$$\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{W}\boldsymbol{\varphi}(\boldsymbol{x}) \tag{2.12}$$

ここで $\boldsymbol{\varphi}(\boldsymbol{x}) = (\varphi_1(\boldsymbol{x}), \ldots, \varphi_m(\boldsymbol{x}))^t$ の各要素は任意の $\boldsymbol{x}$ の 非線形関数，W は $K \times m$ のパラメー

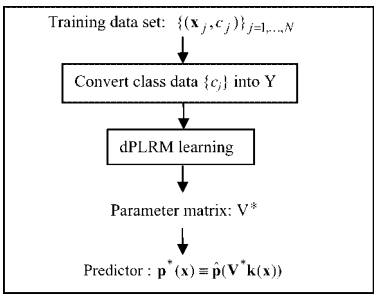

図 2. Training procedure.

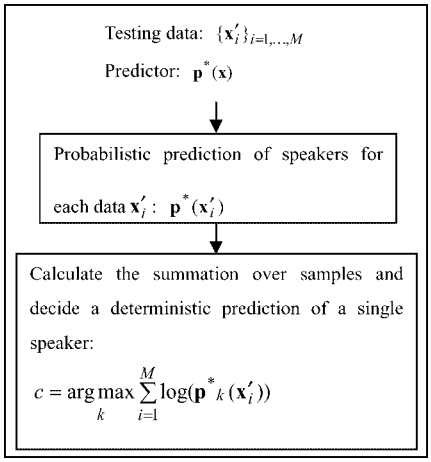

図 3. Testing procedure.

タ行列である．PLRM では次のペナルティ付尤度関数を最小化するWを推定する．

$$PL(\boldsymbol{W}) \equiv L(\boldsymbol{W}) + \frac{\delta}{2}\|\boldsymbol{\Gamma}^{\frac{1}{2}}\boldsymbol{W}\boldsymbol{\Sigma}^{\frac{1}{2}}\|_F^2 \tag{2.13}$$

ここで $\boldsymbol{\Sigma}$ は正定値行列である．$\boldsymbol{\varphi}(\boldsymbol{x}), \boldsymbol{\Sigma}, \boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x}')$，が適切に選ばれた場合，dPLRM と PLRM は全く等しい確率予測 $p^*(\boldsymbol{x})$ を与える（Tanabe, 2001a, 2001b, 2003a）．ただし，dPLRM の方が計算コストは遥かに小さい．カーネル関数として次の多項式カーネル関数

$$\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x}') = (\boldsymbol{x}^t\boldsymbol{x}' + 1)^s \tag{2.14}$$

を用いた場合，PLRM では次の m 次の列ベクトル，$\boldsymbol{\varphi}(\boldsymbol{x})$, $m \times m$ の正定値行列 $\boldsymbol{\Sigma}$ を選択することに相当する．

$$\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{\varphi}(\boldsymbol{x})^t\boldsymbol{\Sigma}^{-1}\boldsymbol{\varphi}(\boldsymbol{x}') = \sum_{j=0}^{s} {}_sC_j(\boldsymbol{x}^t\boldsymbol{x}')^j = \sum_{j=0}^{s}\left[{}_sC_j\left(\sum_{i=1}^{n}[\boldsymbol{x}]_i[\boldsymbol{x}']_i\right)^j\right] \tag{2.15}$$

ここで ${}_sC_j$ は j から s を選ぶ組合せを示し，$[\boldsymbol{x}]_i$ は $\boldsymbol{x} \in \boldsymbol{R}^n$ の各要素による i 次単項式である．今，3 節の実験のように $s = 5$ を選択した場合，$\boldsymbol{\varphi}(\boldsymbol{x})$ の要素数 m は $O(10^{10})$ の非常に大きいオーダーになる． 従って，写像 $\boldsymbol{F}(\boldsymbol{x})$ の表現力は高く，必要であれば，図 1 の GMM による方法の特徴抽出のオペレーションを模擬できると考える．dPLRM ではメルスケールのような経験的な知識を用いることなく，学習データから自動的にある種の非線形変換を構成できると考える．その場合，その変換は，必ずしも GMM による方法における特徴抽出処理に対応する変換と同じになるとは限らない．

3. 評価実験

テキスト独立型話者識別実験において，本方法と GMM による方法の性能を比較する．

3.1 実験条件

実験に用いたデータは，男性 10 名が複数の文章，単語を，約 13ヶ月に渡る 6 時期（T0 から T5 ）に，同一のコンデンサーマイクを用いて，同一の静かな部屋（録音室）で発声した音声である．サンプリング周波数は 16 kHz で，256 次元の対数パワースペクトラム特徴量，26 次元のメルケプストラム特徴量（12 次元のメルケプストラム，正規化対数パワー，それらの一次回帰係数）を，25.6 ms のハミング窓をかけて 10 ms ごとに抽出した．

学習では，DS1 と DS2 の二つのデータセットを用意した．DS1 では各話者は，時期 T0，T1，T2 に 1 文章ずつ，計 3 文章（約 12 秒）を発声している．DS2 では各話者は時期 T2 に 3 文章を発声している．テストでは時期 T3 から T5 に発声した 5 文章，もしくは 5 単語を個別に用いた．各文章，単語テキストは各話者共通であるが，学習とテストでは異なる（表 1）．テスト総数は文章，単語の場合ともに 150 である．

dPLRM ではカーネル関数としては，対数パワースペクトルのデータについては 5 次の多項式関数を，メルケプストラムのデータについては 9 次の多項式関数を用いた．(2.6)式の δ は，実験的に 1e-4∼1e-3 の値に設定した．Γ は(2.7)式に従って設定した．なお，実験に用いた計算機の 64-bit の計算精度を有効利用するために，各特徴ベクトルの全要素は $[-0.5, 0.5]$ の範囲に入るように，一定の割合でスケーリングした．

GMM による方法では，予備実験から混合ガウス分布数 16（対角分散）の GMM を用いた．GMM パラメータは HMM toolkit（HTK, 1989）を利用し，話者ごとに EM アルゴリズムで推定した．それらの初期値は全話者のデータから推定した． テストでは各発声について，その対数尤度の和が最大となる話者を選択した．

表 1. 学習とテストで用いた文章と単語内容.

	内容
学習：文章	1. 背の高さは 170 センチほどで太っている. 2. 大声を出しすぎてかすれ声になってしまう. 3. 足し算，引き算はできなくても絵はかける.
テスト：文章	1. 飛ぶ自由を得ることは人類の夢だった. 2. 初めてルーブル美術館へ入ったのは 14 年前のことだ. 3. 自分の実力は自分が一番よく知っている. 4. これまで少年野球，ママさんバレーなど地域スポーツを支え，市民に密着してきたのは無数のボランティアだった. 5. 銀鮭の卵を輸入して孵化させ，海中で育てる養殖も始まっている.
テスト：単語	1. もう一度 2. 取り返し 3. 訂正 4. 保留 5. 紹介

表 2. 時期 T0/T1/T2 に発声した計 3 文章を学習に用いた時の話者識別率%（信頼区間%）.

テスト音声	方法	識別率 (%)	
		MFCC	Log power spectrum
単語	dPLRM	90.7 (87.3, 94.7)	**92.7 (89.3, 96.0)**
	GMM	89.3 (85.3, 93.3)	84.0 (79.3, 88.7)
文章	dPLRM	99.3 (98.7, 100.0)	**100 .0 (99.3, 100.0)**
	GMM	99.3 (98.7, 100.0)	99.3 (98.7, 100.0)

表 3. 時期 T2 に発声した 3 文章を学習に用いた時の話者識別率%（信頼区間%）.

テスト音声	方法	識別率 (%)	
		MFCC	Log power spectrum
単語	dPLRM	**88.7 (84.7, 92.7)**	83.3 (78.7, 88.0)
	GMM	84.7 (80.0, 89.3)	68.0 (62.0, 74.0)
文章	dPLRM	**98.7 (97.3, 100.0)**	97.3 (95.3, 99.3)
	GMM	98.0 (96.0, 99.3)	86.7 (82.7, 91.3)

3.2 学習セット DS1 の結果

表 2 に学習に DS1 を用いた時の平均識別率とその信頼レベル 90%での信頼区間を示す．単語，文章の場合ともに，対数パワースペクトルを用いた dPLRM による方法が最も高い性能を示した．学習データに 3 時期に渡る時期差による発声変動の情報が含まれているので，dPLRM により，時期差に頑健な話者特徴量が捉えられ，本方法の識別率が高かったと考える．

3.3 学習セット DS2 の結果

表 3 に学習に DS2 を用いた時の平均識別率とその信頼レベル 90%での信頼区間を示す．dPLRM，GMM による方法ともに，対数パワースペクトルを用いるよりも，メルケプストラムを用いた方が高い性能が得られた．特に GMM による方法では，両者の識別率の差は大きかった．GMM による方法では，対数パワースペクトルは高次元（256 次元）のために，パラメータ推定に困難があったと考える．

図 4 に本方法，GMM による方法ともに MFCC を用いた時の各話者ごとの話者識別率を示す．特に単語音声に関しては，本方法の方が GMM による方法よりも識別率のばらつきが小さかった．dPLRM による方法では，どの話者も安定して識別できる傾向にあると考える．

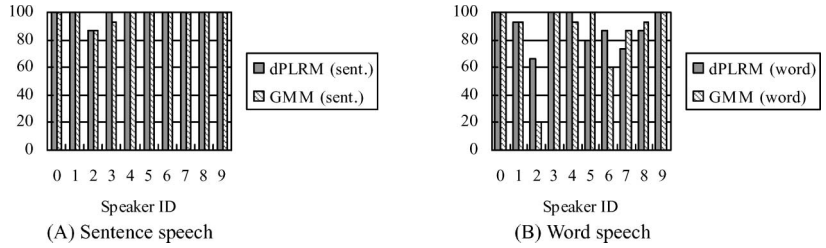

図 4. Speaker identification rates (%) using MFCCs extracted from (A) sentence and (B) word speech for each speaker.

表 4. **時期** T0/T1/T2 **に発声した計** 3 **文章を学習に用い，フレーム周期を** 10，20，30 ms **に変化させた時の話者識別率%(信頼区間%)．**

テスト音声	方法	識別率 (%)		
		10 ms shift	20 ms shift	30 ms shift
単語	dPLRM	92.7 (89.3 ,96.0)	91.3 (87.3, 94.7)	90.0 (86.0, 94.0)
	GMM	89.3 (85.3, 93.3)	86.0 (81.3, 90.7)	85.3 (80.7, 90.0)
文章	dPLRM	100.0 (99.3, 100.0)	100.0 (99.3, 100.0)	100.0 (99.3, 100.0)
	GMM	99.3 (98.7, 100.0)	98.0 (96.0, 99.3)	96.7 (94.0, 98.7)

3.4 間引いたデータの結果

表 4 **に学習に** DS1 **を用い，フレーム周期を** 10 ms **から** 20，30 ms **に長くして，データを間引いた時の平均識別率とその信頼レベル** 90%**での信頼区間を示す．**dPLRM **による方法では，フレーム周期を** 30 ms **にして，データを** 3 **分の** 1 **に間引いた場合でも，もともとのフレーム周期** 10 ms **のデータを用いた** GMM **による方法よりも高い識別率が得られた．**dPLRM **では，カーネル関数を利用して非線形性を効果的に扱うことができるので，データ量が少なくても頑健に識別できたと考える．**

4. 考察

dPLRM **はフレームごとの確率推定量を，**GMM **では尤度推定量を与える．図** 5 **にある話者が“ もう一度 ”と発声した時の音声波形，および** 10 **名の話者のフレームごとの** dPLRM **による確率推定量と** GMM **による尤度推定量を示す．なお，本発声は正しく識別されている．最下位のグラフは** dPLRM **による確率推定量から計算したフレームごとの次式のエントロピーを示す(**Tanabe, 2003b**)．**

$$H_i = -\sum_{k=1}^{K} p_k(\boldsymbol{x}_i) \log p_k(\boldsymbol{x}_i) \tag{4.1}$$

このエントロピーは各フレームの識別力を表している．

識別話者については，24 **から** 50，91 **から** 109 **フレームの音声区間に関して** dPLRM **による確率推定量は高く，エントロピーは小さい．この音声区間は，識別的な情報を比較的多く含む重要な区間と考えられる．**GMM **に関しては，上記音声区間について同じような傾向は見られない．また，始終端の区間は音声情報を含んでいないために，**dPLRM **では全話者共通して確率推定量は小さく，エントロピーは大きい．このように，**dPLRM **を用いれば，識別的な情報**

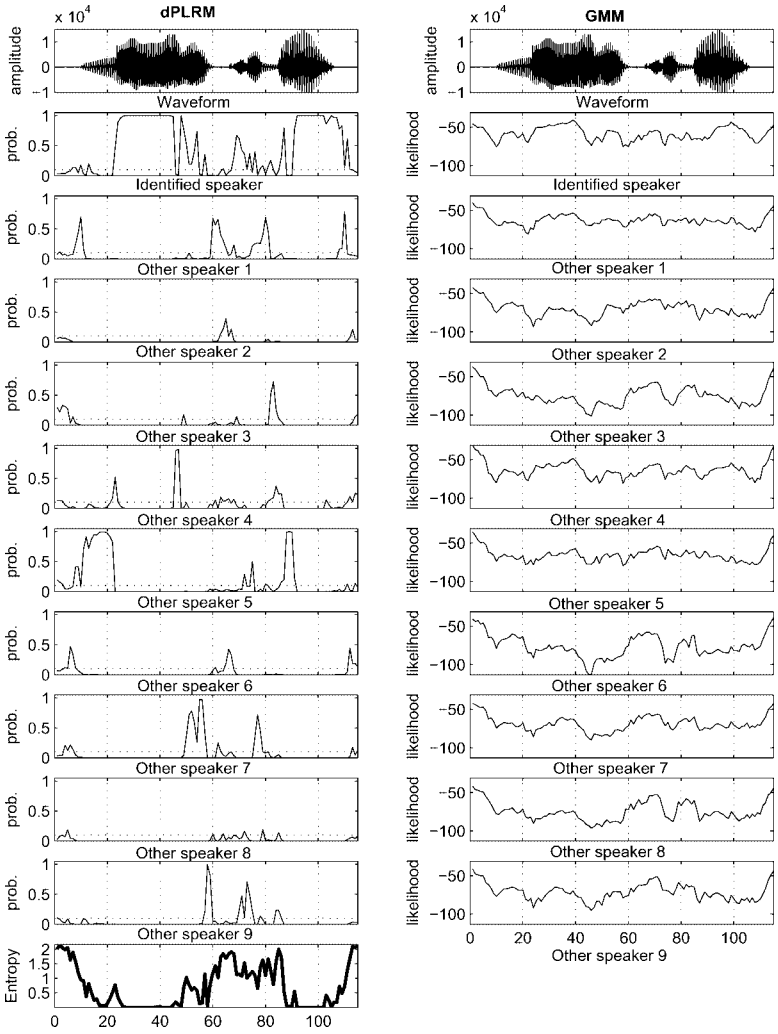

図 5. Comparison of dPLRM probability estimates (left side) and GMM likelihood estimates (right side): the waveform of the word "MOICHIDO" spoken by the identified speaker (top row), the frame-wise dPLRM probability estimates/the frame-wise GMM likelihood estimates for the speaker and other 9 speakers (2nd to 11th rows), and the entropy (12th row). The x-axis indicates frame numbers.

を含む音声区間を確率的に同定することができると考える．

5. まとめ

本稿では，対数パワースペクトルから dPLRM **により，時期差に頑健な話者特徴を捉えることができる話者識別法について述べた．話者** 10 **名が発声した，**3 **時期に渡る時期差の情報を含むデータを用いた識別実験において，対数パワースペクトルを用いた** dPLRM **による方法は，メルケプストラムを用いた場合や従来の** GMM **による方法と比べて，同等以上の性能が得られ**

ることを示した．特に，学習データ量が少ない時には，本方法は GMM による方法よりも高い性能が得られることを示した．

今後は大規模なデータベースを用いて，本方法の評価を行う予定である．

謝　辞

本研究は日本学術振興会の科学研究費補助金（B）16300036 および（C）16500092，ならびに 2004 年度 ISM プロジェクト研究費の援助により行った．

参 考 文 献

古井貞熙（1985）．『ディジタル音声処理』，東海大学出版，秦野．

HTK（1989）．The hidden Markov model toolkit, http://htk.eng.cam.ac.uk

Matsui, T. and Tanabe, K.（2004a）. Speaker recognition without feature extraction process, *Proceedings of Workshop on Statistical Modeling Approach for Speech Recognition*: *Beyond HMM*, *Kyoto*, 79–84.

Matsui, T. and Tanabe, K.（2004b）. Speaker identification with dual penalized logistic regression machine, *Proceedings of ODYSSEY 2004 - The Speaker and Language Recognition Workshop*, Toledo, 363–366.

Matsui, T. and Tanabe, K.（2004c）. Probabilistic speaker identification with dual penalized logistic regression machine, *Proceedings of 8th International Conference on Spoken Language Processing*, III-1797-1800.

Murthy, H. A., Beaufays, F., Heck, L. P. and Weintraub, M.（1999）. Robust text-independent speaker identification over telephone channels, *IEEE Transaction on Speech and Audio Processing*, **7**(5), 554–568.

NIST（1996）. NIST speaker recognition evaluations, http://www.nist.gov/speech/tests/spk/index.htm

Stevens, S. S.（1975）. *Psychophysics*, John Wiley & Sons, New York.

Tanabe, K.（2001a）. Penalized logistic regression machines: New methods for statistical prediction 1, ISM Cooperative Research Report, No. 143, 163–194.

Tanabe, K.（2001b）. Penalized logistic regression machines: New methods for statistical prediction 2, *Proceedings of Information-Based Induction Sciences 2001, Tokyo*, 71–76.

Tanabe, K.（2003a）. Penalized logistic regression machines and related linear numerical algebra, **京都大学数理解析研究所講究録**, No. 1320, 239–249.

Tanabe, K.（2003b）. Entropy-based probabilistic discrimination with dPLRM for collectively labeled noisy data and related performance indexes（manuscript）.

210 Proceedings of the Institute of Statistical Mathematics Vol. 53, No. 2, 201–210 (2005)

dPLRM-based Speaker Identification

Tomoko Matsui[1] and Kunio Tanabe[2]

[1]The Institute of Statistical Mathematics
[2]Department of Science and Engineering, Waseda University

This paper applies the dual Penalized Logistic Regression Machine (dPLRM) to text-independent speaker identification. This machine implicitly discovers speaker characteristics relevant to discrimination only from training data by the mechanism of the kernel regression. Therefore, the speaker identification method based on dPLRM does not require the conventional feature extraction process depending on prior knowledge. We show that the dPLRM method is competitive with the conventional Gaussian mixture model (GMM)-based method with data of 26-dimensional Mel-frequency cepstral coefficients (MFCCs) extracted from training speech uttered by 10 male speakers in three different sessions, even though the dPLRM method directly handles coarse data of 256-dimensional log-power spectrum. It is also shown that the dPLRM method outperforms the GMM-based method especially as the amount of training data becomes smaller.

Key words: Speaker recognition, speaker identification, kernel regression, dual Penalized Logistic Regression Machine, implicit feature extraction.