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

市野　将嗣†　　坂野　鋭*††　　小松　尚久†

† 早稲田大学理工学術院　〒 169–8555 東京都新宿区大久保 3–4–1
†† （株）NTT データ　〒 135–8671 東京都江東区豊洲 3–3–9 豊洲センタービルアネックス
*現在，日本電信電話株式会社　コミュニケーション科学基礎研究所
E-mail: †{ichino,komatsu}@kom.comm.waseda.ac.jp, ††keen@cslab.kecl.ntt.co.jp

あらまし　本稿では核非線形相互部分空間法を用いた音声に基づく話者認識のアルゴリズムを提案し, 実験的に有効性を示す. 従来より音声による話者認識において，混合ガウス分布モデルが用いられている．音声による個人認証は，連続的な音声入力を仮定しているにもかかわらず，混合ガウス分布モデルはこれを単一の音声認識問題の連続したものとして扱っている．そこで我々は，音素の連続する軌跡の形状を比較するアプローチにより連続的な入力音声を積極的に利用し，さらに高性能な認識系を構成することを目指す．さらに音声データには非線形性の存在が予想されることを踏まえ，非線形アルゴリズムのひとつである，核非線形相互部分空間法を認識アルゴリズムとして適用することで，従来より用いられている混合ガウス分布モデルに比較して高い識別率が得られることを示す.
キーワード　音声，テキスト指定型話者認識，核非線形相互部分空間法

A study on application and effectiveness of the Kernel Mutual Subspace Method in speaker recognition

Masatsugu ICHINO†, Hitoshi SAKANO*††, and Naohisa KOMATSU†

† Faculty of Science and Engineering,Waseda University,Okubo 3–4–1, Shinjuku-ku, Tokyo,169–8555 Japan
†† NTT Data Corporation,Toyosu Center Bldg Annex, 3–3–9 Toyosu, Koto-ku, Tokyo 135–8671, Japan
*The author's affiliation is NTT Comunication Science Laoratories now
E-mail: †{ichino,komatsu}@kom.comm.waseda.ac.jp, ††keen@cslab.kecl.ntt.co.jp

Abstract　We propose a method of speaker recognition based on voice by using the kernel mutual subspace method. The Gaussian Mixture Model has already been developed as an algorithm for person authentication using voice. However, it is not sufficiently discussed the verification of the algorithm using the distribution of voice data. Voice data is obtained from an audio stream. However, GMM produces continuous static-voice matching problems when used for speaker recognition. Developing a way to use continuous observation might improve the accuracy of speaker recognition by reducing noise and enabling the extraction of invariant features from voice streams. We propose a method of speaker recognition based on voice by using the kernel mutual subspace method. We experimentally demonstrate the proposed method's effectiveness with simulation results and show that the method achieved higher accuracy than that of using the Gaussian Mixture Model.
Key words　voice, text indicated speaker recognition, kernel mutual subspace method

1. ま え が き

本稿では核非線形相互部分空間法を用いた音声に基づく話者認識のアルゴリズムを提案し, 実験的に有効性を示す.

音声による個人認証はマイクロフォンなどで実装できるため，携帯電話や PC に備え付けのマイクなど，身近なところで利用できる．また，発話は人間が普段から行っている自然な動作であるため，指紋などによる個人認証に比べ，ユーザの心理的な負担が少ない．そのため，音声による個人認証は高レベルのセキュリティを必要としない場面での簡便な個人認証方式として

注目されている．

音声による個人認証のためには特定のパスワードを用いることを前提としたテキスト固定型，何を話してもよいテキスト自由型，認証システムが話す言葉を指定するテキスト指定型の 3 つの方法が提案されている．

本研究では，個人認証方式としてテキスト指定型を採用する．認証を行う際，本人の特徴の現れやすいテキストを提示することによって認証精度の向上が期待できる．また，事前に録音をしておくことが困難であるため，なりすましの問題が軽減される．それに対してテキスト固定型，テキスト自由型は録音装置から出力された音声を用いるなりすましに弱いことが指摘されている．

従来より音声による個人認証の研究は記憶した音声情報との照合の問題として捉えられ，音素の類似性を比較するアプローチが行われてきた．

その中でも，隠れマルコフモデル（Hidden Markov Model, 以下 HMM）や動的計画法 (dynamic programing，以下 DP) による認識が試みられてきた [1] [2]．これらは特定の情報を学習・認識するものであり，テキスト固定型の音声認証アルゴリズムとの親和性が高い．しかし，HMM や DP などに基づく音声認識手法では，時系列情報を利用するため，明らかにテキスト固定型以外への応用が困難である．

テキスト指定型が可能な方法として，部分空間法 [3] [4] が認識アルゴリズムとして用いられた [5]．部分空間法は時系列情報を利用しないためテキスト指定型，テキスト自由型への適用が可能である．

著者らは既に音声において非線形性が存在することを確認している [6]．音声データの分布に非線形構造が認められる場合，部分空間法のような線形性を仮定したアルゴリズムでは分布を十分正確に近似することが出来ず，認識精度の低下を引き起こすことになる．

音声データの分布の非線形性を考慮した話者認識の研究として，混合ガウス分布モデル (Gaussian Mixture Model, 以下 GMM) が広く用いられている [7]．GMM は時系列情報を利用しないためテキスト自由型，テキスト指定型の音声認証アルゴリズムとも共用することが可能である．

GMM は話者認識のアルゴリズムとして広く用いられているにもかかわらず，音声データの分布によるアルゴリズムの妥当性の検証は十分には行われていない．音声データの分布の様子や特徴を考慮して話者認識のアルゴリズムを選択することにより，さらに高性能な認識系が構成できる可能性がある．

また，部分空間法を用いた話者認識を非線形に拡張した研究として，核非線形部分空間法 [8] [9] を適用した例がある [10]．そこでは，核非線形部分空間法を適用した際の識別性能は，GMM を用いた際の識別性能に匹敵することが示されている．

音声による個人認証は，連続的な音声入力を仮定している．しかし，GMM あるいは核非線形部分空間法では，これを単一の音声認識問題の連続したものとして扱っている．

本研究では，音素の連続する軌跡の形状を比較するアプローチにより連続的な入力音声を積極的に利用し，さらに高性能な認識系を構成することを目指す．

こうした話者認識装置において重要なのは音素の連続する軌跡がどのような形状を取るかである．

音声データの特徴抽出系として LPC ケプストラムやメルケプストラムが話者認識でよく用いられている．これらの特徴量は本来，音声認識のために開発された特徴抽出系であり，個人性よりも音声信号の特徴を残している可能性がある．各個人の同一音声が個人の別の音素より離れて分布している可能性がある．つまり，各個人の音素の連続する軌跡は非線形に絡み合って存在する可能性がある．軌跡の形状を正確にモデル化できない場合には，他のカテゴリに誤認識してしまう．

以上を踏まえて，本稿においては，非線形分布が存在し，連続的な入力が仮定できる場合に強力な識別アルゴリズムとして知られている核非線形相互部分空間法を用いることを提案する．以下，**2.** では，非線形, 連続分布条件下での話者認識アルゴリズムを提案する．また，**3.** では話者認識の特徴抽出系でよく利用される LPC ケプストラムとメルケプストラムに関して実験を行い提案手法の有効性を示す．さらに，**4.** ではまとめと今後の課題である．

2. 核非線形相互部分空間法

2.1 相互部分空間法

相互部分空間法（Mutual Subspace Method, 以下 MSM) [11] [12] は，認識対象の入力として複数データが利用できる場合に適用され，入力ベクトル集合も主成分分析を用いて部分空間で表現し，テンプレートの部分空間との間の角度を類似度として識別を行う．この角度に基づいた識別は正準角の概念を用いる．

2 つの部分空間 V, W のなす正準角 θ の余弦は，次のように計算される [11].

テンプレートの部分空間を V，入力された時系列データに対する部分空間を W とする．V の部分空間の次元を M，W の部分空間の次元を N とし，$\phi_m(m = 1, \cdots, M), \psi_i(i = 1, \cdots, N)$ を各部分空間 V, W における正規直交基底ベクトルとする．次いで，式 (1) で表される行列 X の固有値問題を解き，その最大固有値を第 1 正準角に対応する類似度 S_{mutual} とする．また，$N \leqq M (1 \leqq i, j \leqq N)$ とする．

ここで，

$$X = (x_{ij}) \tag{1}$$

$$x_{ij} = \sum_{m=1}^{M} (\psi_i \cdot \phi_m)(\phi_m \cdot \psi_j) \tag{2}$$

とおくと，

$$XU = \Lambda U \tag{3}$$

となる．ここで U は X の固有ベクトル，λ_{max} は固有値 Λ の最大値である．故に，第 1 正準角に対応する類似度 S_{mutual} は，

$$S_{mutual}(V, W) = \lambda_{max} = \cos^2 \theta \tag{4}$$

となる．

さらに，式 (1) の固有値問題の第 j 固有値を第 j 正準角に対応する類似度として扱うことができる [13].

MSM では，学習データと入力データの変動の少ない統計量同士を比較することになり，扱う対象に非線形性がない場合には強力な物体認識手法になる．

2.2 核非線形相互部分空間法

前節で導入した MSM は，扱う対象に非線形性がある場合には，十分な精度を達成できないという問題がある．このような非線形性の問題を解決するために，坂野らによって相互部分空間法と核非線形主成分分析 (Kernel Principal Component Analysis，以下 KPCA) [14] を融合した核非線形相互部分空間法 [15](Kernel Mutual Subspace Method, 以下 KMS) が提案されている．

Schölkopf により提案された強力な非線形 PCA である KPCA では，非線形な関数表現を考えるために関数空間(注1)への非線形写像

$$\Psi : \mathcal{R}^n \to \mathcal{F}, \vec{x} \to \vec{X} \tag{5}$$

を考える．ただし，$\mathcal{F}$ は極めて高次元もしくは無限次元の関数空間である．

次の $m \times m$ 行列

$$K_{ij} = (\Psi(\vec{x}_i) \cdot \Psi(\vec{x}_j)) \tag{6}$$

を定義し，

$$m\lambda\alpha = \alpha K \tag{7}$$

なる固有値問題を解き，特異値分解の公式に従い，

$$V = \frac{1}{\lambda}\alpha\Psi(\vec{x}) \tag{8}$$

の形で基底ベクトルを計算する．$\Psi(\cdot)$ の選択のためには，

$$k(\vec{x}, \vec{y}) = (\Psi(\vec{x}) \cdot \Psi(\vec{y})) \tag{9}$$

を満たすような写像を選ぶ．このような写像が選択できた場合には，$\Psi(\vec{x}) \cdot \Psi(\vec{y})$ は単に関数 $k(\vec{x}, \vec{y})$ を計算することに帰着される．これより，

$$V \cdot \Psi(\vec{x}) = \frac{1}{\lambda}\sum_{i=1}^{m} \alpha_i k(\vec{x}_i, \vec{x}) \tag{10}$$

のようになり，高次元の固有ベクトル V をあらわに求めないでも，その写像を計算できるようになる．

KMS では，辞書側と入力側の部分空間の角度を類似度として扱う．ここで，V を辞書側部分空間の基底ベクトル，W を入力側部分空間の基底ベクトルとする．V，W はそれぞれ辞書側，入力側データから KPCA によって計算された部分空間の基底ベクトルであり，ノルムが正規化されていると仮定すると，辞書登録された m 個の音声群と m' 個の入力された音声系列の類似度は，$(V \cdot W)$ の大きさで評価される．$(V \cdot W)$ は $\mathcal{F}$ 上の内積であるから，この値をあらわに有限の時間で計算することができない．しかし，V, W を

$$V = \sum_{i=1}^{m} \alpha_i \Psi(\vec{x}_i) \tag{11}$$

$$W = \sum_{j=1}^{m'} \alpha'_j \Psi(\vec{x}'_j) \tag{12}$$

と表現することにより，$(V \cdot W)$ の表式は

$$V \cdot W = \sum_{i=1}^{m} \alpha_i \Psi(\vec{x}_i) \cdot \sum_{j=1}^{m'} \alpha'_j \Psi(\vec{x}'_j) \tag{13}$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{m'} \alpha_i \alpha'_j (\Psi(\vec{x}_i) \cdot \Psi(\vec{x}'_j)) \tag{14}$$

となる．

ここで，式 (14) に式 (9) を代入すると，

$$V \cdot W = \sum_{i=1}^{m}\sum_{j=1}^{m'} \alpha_i \alpha'_j k(\vec{x}_i, \vec{x}'_j) \tag{15}$$

となり，有限の時間で類似性が評価できることがわかる．実際に類似度を評価するには式 (1)，(2)，(3) に式 (15) を代入すればよい．

3. 認 識 実 験

本節では, 音声による話者認識に KMS を用いた際の実験結果を示す. 最初に実験系の概要を示し, 認識実験結果を報告し, 結果に関する考察を述べる．

3.1 実験系の概要

3.1.1 特 徴 抽 出

本研究で用いる話者認識装置では音声のうち有音のみを対象とした．発話動作による連続的な音声情報を解析するためである．

特徴抽出系として話者認識でよく用いられている LPC ケプストラムとメルケプストラムの 2 つの特徴量に関して実験を行った. LPC ケプストラムは線形予測分析により求まるケプストラムであり，ピークを重視したスペクトル包絡になる．また，メルケプストラム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を行うことにより抽出される．

KMS は，データを部分空間で表現する際，サンプル数の 3 乗に比例する処理量を有する．また，認識時の類似度を計算する際には，すべての学習データ，認識対象データの間での核関数を計算することになり，処理量が大きい．本研究では，計算時間の発散(注2)を防ぐために, 学習データのサンプル数を削減することにより処理量削減を行った．具体的には，学習データに対して k-平均法を行った際に求められるクラスタ中心を学習

(注1)：関数空間のベクトルについても有限次元のベクトル空間の記号・用語 (「行列」，転置の記号) を用いることとする．

(注2)：音声個人認証の場合，数 kHz〜数十 kHz でサンプリングしたときの数秒〜数十秒の音声データに対して前処理を行い，認証する．そのため，音声データとして数千サンプルを扱うことになる．

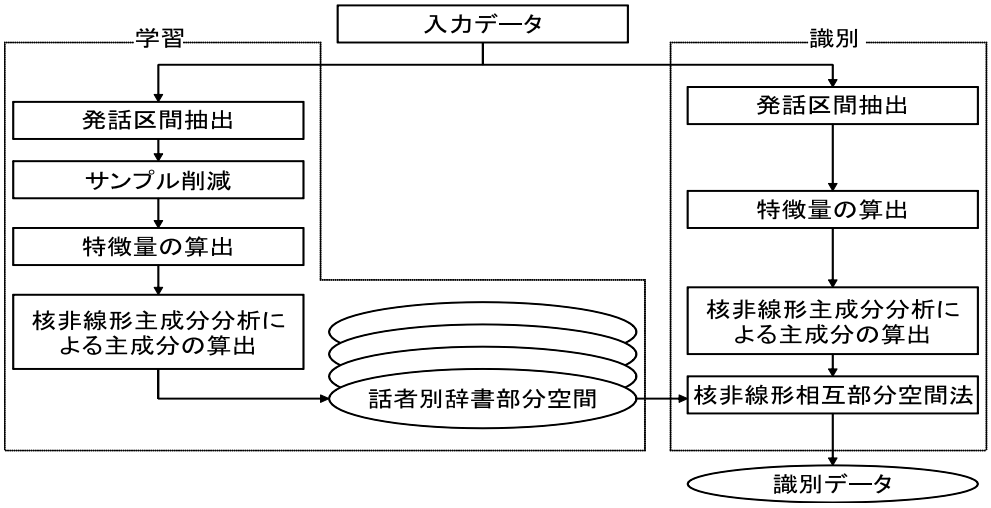

図 1　実験系の概要

データとして用いた [16].

この特徴抽出系と KMS を用いた個人認証方式の概要を図 1 に示す.

3.1.2　実験データ

実験データとして，大規模データベース XM2VTS [17] を使用した．XM2VTS は被験者を一ヶ月間隔で 4 回撮影した (1 回ごとに 1 時期とし，合計 4 時期あるとする) データベースである．発話した際の動画像 (顔画像と音声) が収録されている．

今回の実験データとして，40 人分 (男性 26 人，女性 14 人，4 時期分) を用いた．本研究では，あらかじめ「0,1,2,3,4,5,6,7,8,9」(英語) の 10 個の数字を発話した際の音声のデータをシステムに登録し，認証時に毎回数字の並びを変化させた数字列を提示し，その数字を連続で読み上げることにより認証するというテキスト指定型話者認識を想定している．そのため今回の実験では，学習データは「0,1,2,3,4,5,6,7,8,9」(英語) の 10 個の数字を連続で読み上げたデータを 1 つのデータとし，テストデータは「5,0,6,9,2,8,1,3,7,4」(英語) の 10 個の数字を連続で読み上げたデータを 1 つのデータとして評価した．そして，4 時期中の 2 時期分を学習データ，テストデータとして用いた．ただし，学習データとテストデータに関して，同一時期のデータを選んでいない．また，データの選び方に関してクロスバリデーションを行い，2 回分をあわせて実験結果としている．1 時期につき同じテキストを 2 回発話しているので 1 回の実験につき学習データは 1 人あたり 4 データ，テストデータは 1 人あたり 4 データを用いた．結局 2 回の実験でテストデータとして 40×4×2=320 データを用いた.

1 人あたり 4 データ (あわせて 16 秒前後) を用いて辞書部分空間を表現し，1 人あたり 1 データずつ (4 秒前後) を用いて入力部分空間を表現した.

発話区間抽出では，実験で使用するフレームを選択する．実験で使用するフレームとして，音声波形の有音に対応するフレームのみを手動で切り出した結果を用いた．その後，LPC ケプストラムやメルケプストラムを算出し，特徴ベクトルとして用いた.

実験諸元を表 1 に示す.

表 1　実 験 諸 元

サンプリング周波数	32[kHz]
フレーム長	32[ms]
フレーム周期	8[ms]
特徴抽出	LPC ケプストラム，メルケプストラム
特徴ベクトル次元数	24

3.2　実 験 結 果

音声に関して，KMS を適用することの有効性を確認するために，話者認識において広く用いられている GMM との比較を行った.

KMS については大幅な処理時間の削減と部分空間次元数によらず安定した識別率を得ることを考慮して，学習サンプルを 80%削減した．GMM についてはサンプル削減を行わない.

GMM は複数の入力を仮定した方法ではないため，複数回の認識処理の類似度の平均を類似度として用いた.

LPC ケプストラムを特徴量として用いる際には，KMS は第 1 正準角から第 9 正準角まで考慮した類似度の平均を類似度とした．メルケプストラムを特徴量として用いる際には，KMS は第 1 正準角から第 8 正準角まで考慮した類似度の平均を類似度とした.

核関数として，ガウス型動径基底関数

$$k(\vec{x},\vec{y}) = \exp\left(\frac{-||\vec{x}-\vec{y}||^2}{2\sigma^2}\right) \tag{16}$$

を用い，予備実験により適切と思われる σ を設定した.

LPC ケプストラムを特徴量として用いた際の識別結果を表 2，メルケプストラムを特徴量として用いた際の識別結果を表 3 に示す．LPC ケプストラム，メルケプストラムともに KMS が GMM より高い識別率を示すことがわかる.

図 2，3 に累積識別精度特性 (Cumulative Match Characteristic Curve) を示す．図において，縦軸が累積識別率 (Cumulative

Match Rate, 以下 CMR)[注3]，横軸が順位を表す．これらの図より LPC ケプストラム，メルケプストラムともに KMS は GMM と比較して高い CMR を示すことがわかる．また，KMS を用いて誤識別した場合においても識別候補の上位に本人が判定される．つまり，唇動作などの他のモダリティと統合した場合にも KMS は GMM より高精度化できる可能性があると考えられる．

また，図 4，5 に照合精度特性 (Receiver Operating Characteristic Curve, 以下 ROC 曲線) を示す．図において横軸が FRR(False Reject Rate)，縦軸が FAR(False Accept Rate) を表す．図 4，5 より LPC ケプストラム，メルケプストラムともに KMS が GMM と比較して高い認識性能を示すことが確認できた．

表 2　識別結果 (LPC ケプストラム)

識別手法	KMS	GMM
識別率（%）	90.3	80.0
σ	2.5	-
辞書次元数	13	-
入力次元数	10	-

表 3　識別結果 (メルケプストラム)

識別手法	KMS	GMM
識別率（%）	98.1	96.9
σ	2.0	-
辞書次元数	16	-
入力次元数	8	-

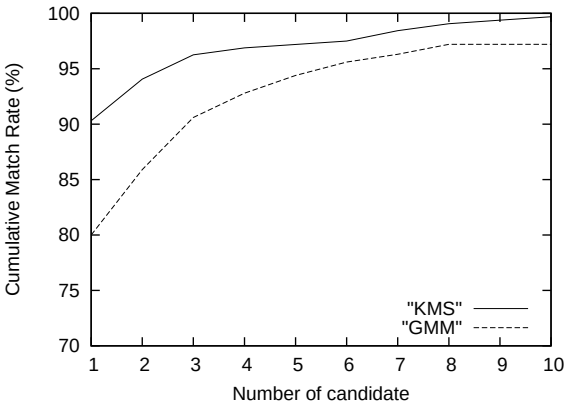

図 2　累積識別精度特性 (LPC ケプストラム)

3.3 考　　察

今回の実験において，KMS が GMM より高い識別性能を示した理由について考察する．表 2，3 に示したように LPC ケプストラムのほうがメルケプストラムより KMS の識別率と GMM の識別率の差が大きかったので，以下，LPC ケプストラムの場合に関して考察する．

KMS が GMM より高い識別性能を示した理由として次の 3 つを考えた．

- 雑音除去の効果
- 細かい特徴抽出・認識
- モデル同士を比較することにより認識

以下，これらについて説明する．

(注3)：識別アルゴリズム，識別装置あるいは個人識別システムが，同一の音声同士の識別判定で，与えられた順位以内の候補として選択する確率

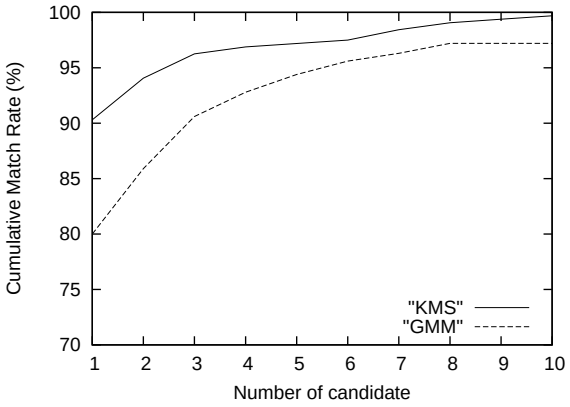

図 3　累積識別精度特性 (メルケプストラム)

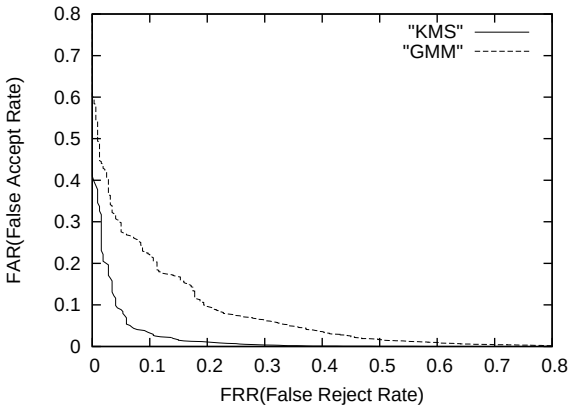

図 4　ROC 曲線 (LPC ケプストラム)

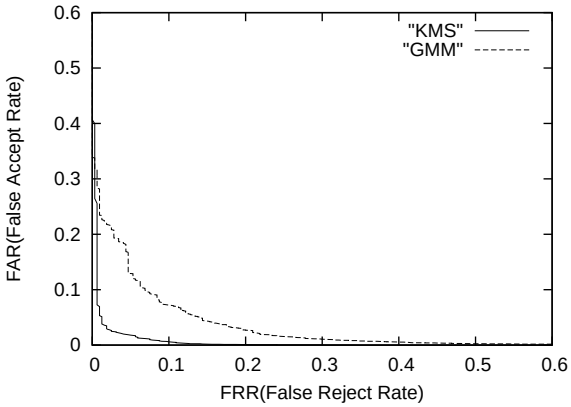

図 5　ROC 曲線 (メルケプストラム)

3.3.1 雑音除去の効果

今回の実験では，音声データの記述に LPC ケプストラムを使用した．LPC ケプストラムは本来, 音声認識のために開発された特徴抽出系でありピークを重視したスペクトル包絡を表すため，個人性よりも音声信号の特徴を残していると考えられる．つまり, 各個人の同一音声が個人の別の音素より離れて分布し

ていると考えられる．

さらに英語の音素は 44 個あるため，音声データとして取り得る状態が多く，複雑な分布になると考えられる．

つまり本人と他人の音声データは複雑に絡み合って存在していると考えられる．

それを踏まえ，次の仮説を立てた．

本人のサンプル (フレームに相当) であるにもかかわらず他人分布に近いサンプルが存在する可能性がある．このようなサンプルが多く存在する場合には 1 サンプルの入力を対象とした識別アルゴリズムでは識別精度が悪くなると考えられる．

GMM(K-平均法で考える(注4)) を用いる場合，クラスタ中心と 1 サンプルごとの距離の平均を類似度とする．他人分布に近いサンプルが多く存在する場合には本人のクラスタ中心とサンプルの距離が大きくなる．つまり距離の平均値が大きくなるため誤識別すると考えられる．ただし，他人分布に近いサンプルが多く存在していてもクラスタ中心と 1 サンプルごとの距離の平均 (以下，平均距離) が小さい値を示す場合には正しく識別できると考えられる．

KMS は学習，入力データの両方を KPCA を用いて主成分を計算し，その主成分を用いて類似度を求める方法である．つまり学習と入力の双方の変動の少ない部分同士を比較する手法である．そのため，他人分布に近いサンプルが多少存在しても影響をほとんど受けないと考えられる．本人分布に近いサンプルが多く存在していても本人分布との平均距離が他人分布との平均距離より大きい値を示す場合には本人テンプレートの分布と入力のデータの分布の形状が異なることが予想され誤識別すると考えられる．ただし，音声データの性質を考慮するとこのようなデータは多くないと考えられる．

以上より，GMM より KMS のほうが高い識別性能を示したと考える．このことを検証するために以下の実験を行った．

入力データのみに雑音を付加し，その影響を KMS と GMM の識別率を比較することにより調べた．音声データは複雑に絡み合って存在しているため，雑音を付加することによりさらに本人分布からサンプルを離し，他人分布に近づけることができる．そのとき，識別率の低下が少ないほうが雑音の変動の影響を受けずに主成分を計算できたと考えられる．

そのことを確認するために，電子協騒音データベース (人混み) [18] を使用して音声のテストデータに SN 比が 15dB,10dB,5dB になるように雑音付加の割合を変えて実験を行った．学習データは **3. 1. 2** と同じ条件のもの (雑音を付加しないデータ) である．

識別結果を表 4 に示す．この結果より雑音を強くすることに伴う識別率の低下が GMM に比べ KMS のほうが少ないことがわかる．つまり KMS のほうが GMM より変動を吸収していると考えられる．

3. 3. 2　細かい特徴抽出・認識

KPCA により求まる主成分に着目すると，低次元の主成分は分布の概形を表現し，次元数が上がるにつれて主成分は細かい情報を表す．

(注4)：要素となるガウス分布すべての分散共分散行列が等しく単位行列で，かつすべての混合重み等しい場合の EM アルゴリスムは，K-平均法とほぼ同じ振る舞いをする

表 4　雑音付加時の識別結果

	clean	15dB	10dB	5dB
KMS (%)	90.3	90.6	91.3	85.6
GMM (%)	80.0	79.7	77.2	73.1

一方，KMS の類似度に着目すると，学習部分空間と入力部分空間の間には，部分空間の複数の基底を用いると複数の正準角を定義することができる．

LPC ケプストラムは音韻性を残していると考えられるため各個人の音素の連続する軌跡の概形は似た形状になると考えられる．つまり個人を認識するためには軌跡の細かい情報も必要であると考えられる．この情報を KMS では主成分で表現しているため高い識別精度を得ることができたと考える．

このことを確認するために以下の調査を行った．

今回の実験における識別率と部分空間次元数の関係を図 6 に示す．図 6 より部分空間次元数が大きくなるにつれて識別率が高くなることが分かる．そして部分空間次元数が 6 のとき GMM より識別率が高くなり，部分空間次元数が 13 のとき識別率が最も高い．つまり分布の概形を表す主成分に加えてより細かな情報を記述した主成分が個人の識別に有効に作用していることが分かる．

使用する正準角の数と識別率の関係を表 5 に示す．使用する正準角を増やすことにより識別率が向上する．そして正準角を 7 個使うとき GMM の識別率より高くなり正準角を 9 個使うとき識別率が最も高い．つまりより細かな情報を表す正準角が識別に有効に作用していることが分かる．

KMS は **3. 3. 1** の雑音除去効果を備えながら分布の細かい部分も用いて比較することができるため高い識別性能を得ることができたと考えられる．

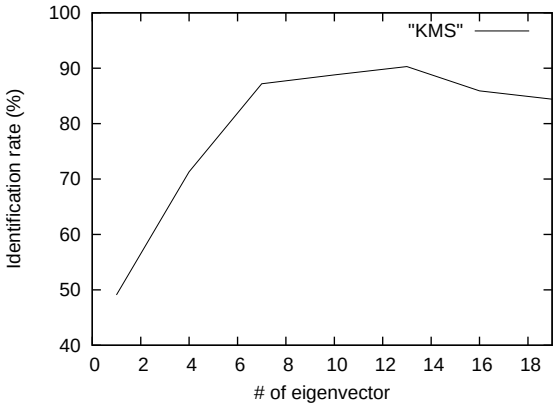

図 6　識別率と部分空間次元数の関係

3. 3. 3　モデル同士の比較

本人，他人のデータ分布が非線形に絡み合って存在している場合，音声データの分布の重なりが大きくなるため GMM のような 1 フレームのみを用いた識別は難しい．

それに対して KMS は学習，入力データを KPCA を用いて

表 5　使用する正準角の数と識別率の関係

使用する正準角の数	1	2	3	4	5	6	7	8	9	10
識別率 (Mean)(%)	35.6	51.3	64.1	70.3	75.3	79.1	84.7	88.4	90.3	82.8

主成分を計算し，モデル同士を比較する手法である．つまり，分布の重なりが大きい場合においても全体をみて一致，不一致を認証するため安定して認証が行えると考えられる．

学習，入力データを KPCA を用いて主成分を計算し，モデル同士を比較することの有効性を確認するために 1 サンプル対 1 サンプルの認識アルゴリズムである MPFC(Multiple Potential Function Classifier) [19] と 1 サンプル対 N サンプルの認識アルゴリズムである GMM と核非線形部分空間法 (Kernel based Nonlinear Subspace method，以下 KNS) との比較実験を行った．KMS と KNS の違いは入力データを KPCA を用いて主成分を計算して類似度の算出に利用しているかどうかである．MPFC，GMM，KNS と KMS の識別性能を比較することにより学習，入力データを KPCA を用いて主成分を計算し，分布同士を比較することの有効性を示すことができると考えられる．

実験データは **3. 2** と同じ条件のものを用いる．

核関数は式 (16) の動径基底関数を使用した．また，KNS に関して複数回の認識処理の類似度の平均を類似度として用いた．KMS に関して表 2 と同様に第 1 正準角から第 9 正準角までを考慮した類似度の平均を類似度とした．

識別結果を表 6 に示す．

表 6　識別結果 (KMS と MPFC，GMM，KNS の比較)

識別手法	KMS	GMM	KNS	MPFC
識別率（%）	90.3	80.0	77.5	32.8
σ	2.5	-	2.5	0.3
辞書次元数	13	-	16	-
入力次元数	10	-	-	-

この結果より 1 サンプル対 1 サンプルの認識 (MPFC) よりも 1 サンプル対 N サンプルの認識 (GMM,KNS) のほうが識別率が高いことがわかる．さらに 1 サンプル対 N サンプルの認識よりも N サンプル対 N サンプルの認識 (KMS) のほうが識別率が高く，モデル同士を比較することの有効性を示すことができたと考えられる．

さらに KMS と GMM に関して細かく調べるために本人データ，他人データそれぞれを K-平均法を用いてクラスタリングして求まるクラスタ中心からの距離分布と識別結果の関係を調べた．図 7 に示すように 1 サンプル (フレームに相当) ごとに本人データから求まるクラスタ中心との最小距離，他人データから求まるクラスタ中心との最小距離を比較した．

他人分布に近いサンプルが多少多く存在する場合でも，KPCA を用いて主成分を計算することにより変動の少ない成分 (主成分) を抽出し主成分同士を比較するため，KMS では正しく識別できると考えられる．

以下の 4 つの場合について調べた．

a)　KMS でも GMM でも正しく識別できている人物について (320 データ中 247 データ，いわゆる識別しやすいデータ)

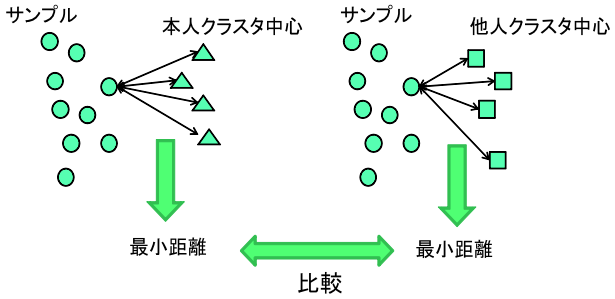

図 7　クラスタ中心からの距離分布

サンプル数 471(1 人分発話) に対して本人データのクラスタ中心よりも他人データのクラスタ中心データに近いサンプルは 367(471 サンプル数に対して 78%) であった．471 サンプル中，本人のクラスタ中心に近いサンプルが 104 サンプルあり (本人のクラスタ中心に近いサンプルが一番多い)，他人のクラスタ中心に近いサンプルに比べ多い結果を得た．平均距離が小さい傾向にあった．

b)　KMS でも GMM でも正しく識別できない人物について (320 データ中 22 データ，いわゆる識別しづらいデータ)

サンプル数 405(1 人分発話) に対して本人データのクラスタ中心よりも他人データのクラスタ中心データに近いサンプルは 389(405 サンプル数に対して 96%) であった．405 サンプル中，本人のクラスタ中心に近いサンプルが 16 サンプルであり (他人一人一人のクラスタ中心に近いサンプルのほうが多い)，他人のクラスタ中心に近いサンプルに比べ少ない結果を得た．平均距離が大きい傾向にあった．

c)　GMM では正しく識別できないが KMS では正しく識別できている人物について (320 データ中 42 データ，いわゆる a) の場合と b) の場合の中間に位置するデータ)

サンプル数 433(1 人分発話) に対して本人データのクラスタ中心よりも他人データのクラスタ中心データに近いサンプルは 383(433 サンプル数に対して 88%) であった．433 サンプル中，本人のクラスタ中心に近いサンプルが 50 サンプルあり，他人のクラスタ中心に近いサンプルと本人のクラスタ中心に近いサンプルの差は a) の場合に比べ小さい．この場合，本人のクラスタ中心に近いサンプルが一番多い場合もあれば 2 番目や 3 番目に多い場合もあった．本人分布との平均距離が他人分布との平均距離より小さな値を示す傾向にあった．

d)　GMM では正しく識別できるが KMS では正しく識別できない人物について (320 データ中 9 データ)

サンプル数 316(1 人分の発話) に対して本人データのクラスタ中心よりも他人データのクラスタ中心データに近いデータは 264(316 サンプル数に対して 83%) であった．基本的には a) の場合に分類されるはずである．

そこで 1 フレームごとの距離の変動を調べた．1 人分に関し

ての本人のクラスタ中心と入力サンプルとの距離の変動を図 8 に示す．図において，横軸がフレーム番号，縦軸が距離をあらわす．この結果より全体としては距離が小さいが，極端に距離の大きいフレームが存在していることがわかる．KMS では共分散行列を計算し固有ベクトルを求めるのでそのようなフレームがある場合，影響を及ぼす．そのため KMS では誤識別したと考えられる．それに対して GMM では平均距離を類似度とするため，そのようなフレームが存在しても平均距離が小さければ識別できると考えられる．

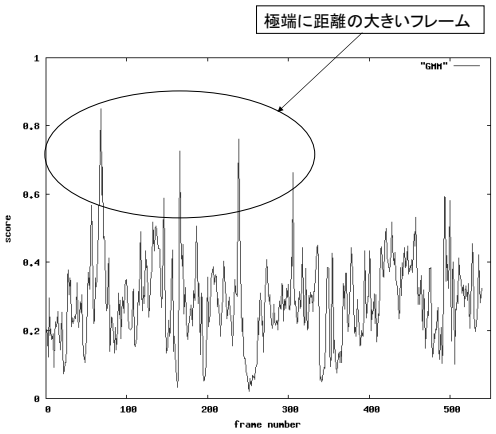

図 8　1 フレームごとの距離の変動

d) のデータよりも c) のデータが多いため GMM より KMS のほうが識別率が高かったと考えられる．

このように他人のクラスタ中心に近いサンプルが多少多い場合でも KMS は識別できていることがわかる．つまり GMM では学習データの分布と入力データの分布がぴったり張り付いていないと正しく認識できないが，KMS ではぴったり張り付いていない場合でも正しく認識できることがわかる．

KMS は **3. 3. 1** で示したように変動の少ない成分を **3. 3. 2** で示したように細かな部分も含めてモデルとして利用してモデル同士を比較することにより認識を行っているため KMS が高い認識性能を示したと考えられる．

以上より，今回の実験において，KMS が GMM より高い識別性能を示したと考えられる．

4. む す び

テキスト指定型話者認識に適用可能な音声による話者識別のアルゴリズムを提案した．

各個人の音素の連続する軌跡は非線形に絡み合って存在すると考えられるため，本研究では，音素の連続する軌跡の形状を比較するアプローチにより連続的な入力音声を積極的に利用することを考えた．具体的には認識アルゴリズムに KMS を用いることを提案し，識別実験を通して，提案手法が有効であることを示した．

今後は，より大規模なデータで再検証を行い，KMS を用いることの有効性を確認していく予定である．

さらに，より認証精度を高めるために本人と他人の部分空間の正準角を広げるもしくは本人同士の部分空間の正準角を狭めるようなアプローチを考えていきたい．

謝辞　本研究の一部は，日本学術振興会科学研究費補助金 (特別研究員奨励費) の補助による．

文　　献

[1] J.P. Campbell,“Speaker recognition: A tutorial,” Proc.IEEE, vol.85, no.9, pp.1437-1462, 1997.

[2] K.Yu, J.Mason and J.Oglesby,“Speaker recognition using hidden Markov models, dynamic time warping and vector quantisation,” Vision, Image and Signal Processing, IEE Proceedings- vol.142, issue 5, pp.313 - 318, October 1995.

[3] S.Watanabe and N.Pakvasa, “Subspace method of pattern recognition,”Proc. 1st Int.J.Conf. on Pattern Recognition, 1973.

[4] E. Oja,“Subspace Methods of Pattern Recognition,”Research Studies Press, 1983.

[5] J.B. Attili, M. Savic, and J.P. Campbell,“A TMs32020-based real time, text-independent, automatic speaker verification system,“ IEEE International Conference ICASSP’88, pp. 599-602, April 1988.

[6] 市野将嗣，高倉大樹，坂野　鋭，小松尚久，“母音音素分布の非線形性について，” 信学総大，D-14-14，March 2004.

[7] R.A. Reynolds and R.C. Rose,“Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models,” Speech and Audio Processing, IEEE Transactions on , Volume:3, Issue:1, pp.72-83, January 1995.

[8] 津田宏治，“ヒルベルト空間における部分空間，” 信学論 (D-II)，vol.J82-D-II，no.4，pp.592-599，April 1999.

[9] 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．” 信学論 (D-II)，vol.J82-D-II，no.4，pp.600-612，April 1999.

[10] 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，” 信学技報，PRMU2000-127,2000 年.

[11] 前田賢一，渡辺貞一，“局所構造を導入したパターン・マッチング法，” 信学論（D），vol.J68-D，no.3，pp.345-352，March 1985.

[12] 山口　修，福井和広，前田賢一，“動画像を用いた顔認識システム，” 信学技報，PRMU97-50，June 1998.

[13] 福井和広，山口　修，鈴木　薫，前田賢一，“制約相互部分空間法を用いた環境変動にロバストな顔画像認識–照明変動の影響を抑える制約部分空間の学習–，” 信学論（ D-II），vol.J82-D-II，no.4，pp.613-620，April 1999.

[14] B.Schölkopf *et al.*,“Nonlinear component analysis as a Kernel eigenvalue problem,” Neural Computation, vol.10, pp.1299-1319, 1998

[15] 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，” 信学論（D-II），vol.J84-D-II，no.8，pp.1549-1556，August 2001.

[16] 市野将嗣，坂野　鋭，小松尚久，“クラスタリングを用いた核非線形相互部分空間法の処理量削減手法，” 信学論 (D)，vol.J90-D，pp.2168-2181，August 2007.

[17] K.Messer, J.Matas, J. Kittler, J. Luettin and G. Maitre, “XM2VTSDB: The extended M2VTS database,” in Proc. of Int. Conf. on Audio and Video based Biometric Person Authentication, Washington, USA,1999.

[18] 板橋秀一，“騒音データベースと日本語共通音声データ DAT 版，” 音響誌, vol.47, no.12, pp.951-953, 1991.

[19] H. Sakano, T. Suenaga, “Classifiers under continuous observation,” In Ed. T. Caeli, et al., Structual, Syntactic, and Statistical Pattern Recognition 2002, Lecture Note on Computer Science 2396, pp.798-805, 2002, Springer-Verlag Berlin Heidelberg 2002 Proc. in IAPR Intl. Workshop. SPR 2002