

音素部分空間の統合による音声特徴量抽出の検討

朴 玄信[†] 滝口 哲也[†] 有木 康雄[†]

[†] 神戸大学工学研究科 〒 657-8501 兵庫県神戸市灘区六甲台町 1-1

E-mail: †silentbattle@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

あらまし 本稿では、事前学習による音声特徴量抽出の枠組みの中で、音素部分空間の統合により、得られた空間へ観測ベクトルを線形射影する特徴量抽出法を提案する。近年、音声認識システムで最も使われている特徴量空間はMFCC(Mel-Frequency Cepstrum Coefficient) 空間である。MFCC 空間は対数メルフィルタバンク出力に離散コサイン変換を適用することにより得られる空間であるが、この射影軸は対象とする音声データの音響特徴に依存することなく一意に決まる。音声認識システムの実用化のためには、実環境で収録された音声データに潜在する音素情報以外の情報を除去する必要がある。MFCC は様々なノイズ抑制手法と組み合わせて使われているが、想定できないノイズが混入すると性能の劣化は避けられない。そこで本研究では、対象データに対し事前学習を行い、観測データから音素情報だけを抽出する部分空間（射影行列）を推定する。具体的には各音素データに対して主成分分析や判別分析を用い特定音素情報を抽出する部分空間を求め、さらに各音素部分空間へ射影された全てのデータに対して主成分分析を行い各音素部分空間を統合する。この統合された部分空間は対象とする音声の音素情報が含まれる空間になり、観測データから音素情報だけ抽出することができる。評価実験としては提案する特徴量を用い、音素 HMM(Hidden Markov Model) を学習し、クリーン音声と残響音声の孤立単語認識を行った。評価実験の結果、提案手法の有効性が確かめられた。

キーワード 主成分分析, 判別分析, 部分空間統合, 特徴量抽出, 単語認識,

Integration of Phoneme-subspaces for Speech Feature Extraction

Hyunsin PARK[†], Tetsuya TAKIGUCHI[†], and Yasuo ARIKI[†]

[†] Graduate School of Engineering, Kobe University

Rokkodaicho 1-1, Nada-ku, Kobe, Hyogo, 657-8501 Japan

E-mail: †silentbattle@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

Abstract In this paper, we propose a speech feature extraction method that is estimating each phoneme-subspace and integrate each subspace within a framework of feature extraction by pre-learning. The most commonly used speech feature for speech recognition is MFCC that is computed applying DCT to the mel-scale filter bank output. This feature space dose not depend on target speech data set and is decided by uniquely. To make speech recognition system fit for practical use, noise that is latent in observed data and useless for recognition must be removed. MFCC is uesd combined with other removing noise methods but performance degradation is inescapable if unexpected noises are mixed in observed data. Consequently, subspaces (projection matrix) that only extract phonemic information are estimated by pre-learning with observed data. Specifically, PCA or LDA are applied to each phoneme data set and each phoneme-subspace were estimated. Additionally, all phoneme-subspaces are integrated by PCA. This integrated subspace will have phonemic information of target speech data set and extract only that information. In evaluation experiment, we modeled phoneme HMM by proposed feature and carried out isolated word recognition experiments. The experiment results showed that the proposed method is effective compared to conventional methods.

Key words PCA, LDA, Subspaces integration, Feature extraction, Word recognition

1. はじめに

現在の音声認識システムは接話マイクなど限られた制約条件の基では高い認識率を示している。しかしハンズフリーなど実環境では収録した音声データに存在するノイズの影響で音声認識システムの性能が急激に落ちる。音声認識システムを実用化するためには実環境にロバストである必要がある。

ロバスト性を保つための従来研究として特徴量ベース手法とモデルベース手法に分けられる。さらに特徴量ベースの手法には特徴量抽出のための事前学習の有無に分けられる。事前学習無しの特徴量抽出法としては RASTA [1], チャネル正規化 [2], 雑音推定 [3], 残響抑制 [4] などがあり, 事前学習有りの特徴量抽出法としては主成分分析 [5] [6] [7], カーネル主成分分析 [8] [9], 線形判別分析 [10], 独立成分分析 [11] などを用いた手法がある。

近年の音声認識システムで一般的に使われる MFCC は事前学習無しの特徴量抽出法であり, 実環境では様々なノイズ抑制手法を組み合わせて使われている。しかし, 観測される音声信号には音素時系列情報, 発話者の身体や感情の情報, 環境情報など様々な情報が混在する。このような情報の中から, 音声認識システムが主に必要とする情報は音素時系列情報である。MFCC などの学習無し型特徴量抽出法は, 予想できないノイズが混入すると有効な情報だけを抽出するのは困難になる。このような理由で本研究ではロバスト性を保つため, 音声信号に対して主成分分析や線形判別分析を用い, 音素時系列情報を強調させる事前学習有り型の特徴量抽出法を提案する。

提案する手法は音声の音素情報を用いて音素情報だけを抽出する部分空間 (射影行列) を求める手法である。各音素性を表す音素部分空間は主成分分析や線形判別分析で推定する。ここまでは [10] で述べられた手法と似ている。本研究ではさらに各音素情報を表す部分空間の統合を主成分分析を用いて行う。統合を行うことで重複する情報の削減と音素間の相関情報を表すことが可能になる。

以降 2 章では音素部分空間作成のための主成分分析と判別分析を概説し, 3 章では提案手法を説明する。4 章では評価実験の条件と結果の考察を行う。最後に 5 章で結論としてまとめる。

2. 部分空間作成

本研究で音素部分空間を作成するために, 主成分分析 (Principal Component Analysis; PCA) と線形判別分析 (Linear Discriminant Analysis; LDA) を用いた。本節では, この 2 つの手法について概説する。

2.1 主成分分析

主成分分析とは, 対象とするデータ集合の表現に適した正規直交基底行列 (空間) を求め, 元のデータベクトルを求めた空間へ線形射影する手法である。この手法は空間の次元圧縮や特徴量係数間の無相関化に用いられている。

P 次元空間の t 番目データベクトルを \mathbf{x}_t とすると, データ集合の共分散行列 S は次式のように定義される。

$$S = \frac{1}{N} \sum_{t=1}^N (\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})' \quad (1)$$

$\bar{\mathbf{x}}$ は平均ベクトル, N はデータ数, \mathbf{x}' は \mathbf{x} の転置ベクトルである。

正規直交基底行列は共分散行列 S を次式のように固有値分解することで求まる。

$$S\phi_k = \lambda_k\phi_k, (k = 1, 2, \dots, P) \quad (2)$$

ϕ_k は固有値 λ_k に対応する固有ベクトルである。固有値が大きい順に並べ, $Q (< P)$ 個の固有ベクトルを選択することで, データの部分空間が求まる。

新しい空間上のデータベクトル \mathbf{y}_t は次のような線形射影により求まる。

$$\mathbf{y}_t = \Phi^T (\mathbf{x}_t - \bar{\mathbf{x}}) \quad (3)$$

$$\Phi = (\phi_1, \phi_2, \dots, \phi_Q) \quad (4)$$

主成分分析により求まる固有値は対応する固有ベクトルへ射影されたデータ集合の分散を表わす。一般的に音声認識システムで用いられている特徴量空間上のクリーンの音声データに対して主成分分析を行うと, 認識に有効な音声情報は多数の音素情報が含まれているため, 話者情報などに比べ変動が大きいと考えられ, 有効な特徴量が抽出できる。

主成分分析による部分空間法は特に話者の固有部分空間を推定して識別を行う手法 [13] [14] [15] でよく用いられている。

2.2 線形判別分析

線形判別分析の目的は, 多クラスデータ集合のクラス判別能力を向上させる射影行列を求めることである。

M 個のクラスの中, i 番目クラスは N_i 個のサンプルを持っているとすると, クラス内分散 S_W とクラス間分散 S_B は次のように定義される。

$$S_W = \sum_{i=1}^M S_i \quad (5)$$

$$S_B = \sum_{i=1}^M N_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T \quad (6)$$

S_i は i 番目クラスの共分散行列であり, 式 1 のように求まる。 $\bar{\mathbf{x}}_i$ はクラス i の平均ベクトル, $\bar{\mathbf{x}}$ はグローバル平均である。

クラス間分散 S_W 対クラス内分散 S_B 比を最大にするには, $S_W^{-1} S_B$ の固有値分解をし, 固有値が大きい固有ベクトルで射影行列を構成することと等価であることが [17] で示されている。

3. 音素部分空間の統合による特徴量抽出

人の声をマイクロフォンで収録すると, 収録されたデータの中には音素の時系列情報だけではなく, 話者の身体や感情の情報, さらに環境音など様々な情報が混在する。音声認識システムで一般的に使われている特徴量空間は MFCC 空間であるが, この空間は人の聴音構造はスペクトルの包絡特徴で現れると仮定して, メル周波数フィルタバンク出力に対し離散コサイ

ン変換から得られる空間である．この特徴量はクリーンな環境では優れた認識率を保証するが，ノイズが入ると包絡の形状が変わってしまい認識率が低下する問題がある．

この問題に対し我らの先行研究 [12] では主成分分析を用い音素部分空間を推定し，統合する手法を提案した．特に音素部分空間は分散の低い固有ベクトルに音素固有の情報が含まれていると仮定しその基底ベクトルを用いて推定して，実験を行った結果残響に頑健な結果が得られた．

本研究の提案手法は先行研究の [12] の拡張である．音素部分空間は主成分分析だけではなく判別分析を用いても推定を行う．主成分分析の場合は一般的な方法で分散の大きい固有ベクトルで音素部分空間を推定する．これは先行研究と違う点であるが，音素の部分空間の次元数を先行研究で選ばれた分散の小さい基底ベクトルまである程度含まれるように設定し，後段に主成分分析を行い統合する．後段の主成分分析も分散が大きい固有ベクトルを選ぶので，最終的に求まる軸は音素間識別の特徴（音素間の変動成分，音素の固有特徴）が得られると考えられる．

提案する特徴量空間は大きく 2 つの役割をする．1 つ目は PCA を用いて空間を統合したので，音素情報と無関連な情報は除去される．2 つ目は各音素空間を統合したことで，観測データが統合された空間に射影されて得られる特徴量には観測データの各音素との相関情報が含まれるようになる．これは主に話者識別で多数の話者モデルを作成して，観測データの多数モデルとの尤度を用いて識別する手法と似ている．

図 1 は音素部分空間学習法を表わしている．

3.1 節以降では音素部分区間の学習と統合の定式化を行う．

3.1 音素部分空間

主成分分析や線形判別分析を用いて音素部分空間作成を定式化する．

まず， i 番目の音素データの共分散行列を S^i として次式のように定義する．

$$S^i = \frac{1}{N} \sum_{t=1}^N (\mathbf{x}_t^i - \bar{\mathbf{x}}^i)(\mathbf{x}_t^i - \bar{\mathbf{x}}^i)^T, (i = 1, \dots, M) \quad (7)$$

\mathbf{x}_t^i は i 番目音素の観測ベクトルを表わし， $\bar{\mathbf{x}}^i$ は平均ベクトルである． M は対象とする音声の全音素数を表わす．

PCA を用いて音素部分空間を求める場合は，各 S^i に対して固有値分解を行う．

$$S^i \phi_k^i = \lambda_k^i \phi_k^i, (k = 1, 2, \dots, P) \quad (8)$$

P 次元のデータから P 個の固有値と固有ベクトルが推定され，対応する固有値が大きい順に Q 個の固有ベクトルを選び，音素部分空間とする．

LDA を用いて音素部分空間を求める場合は，各 $S_W^{-1} S_B^i$ に対して固有値分解を行う． S_W^i と S_B^i は全音素クラスを i 番目の音素クラスと残りの音素クラスの 2 クラスに再分割して求めたクラス内共分散とクラス間共分散である．

観測ベクトル \mathbf{x}_t を i 番目音素の部分空間へ射影した特徴ベクトル \mathbf{y}_t^i は次式のように表わす．

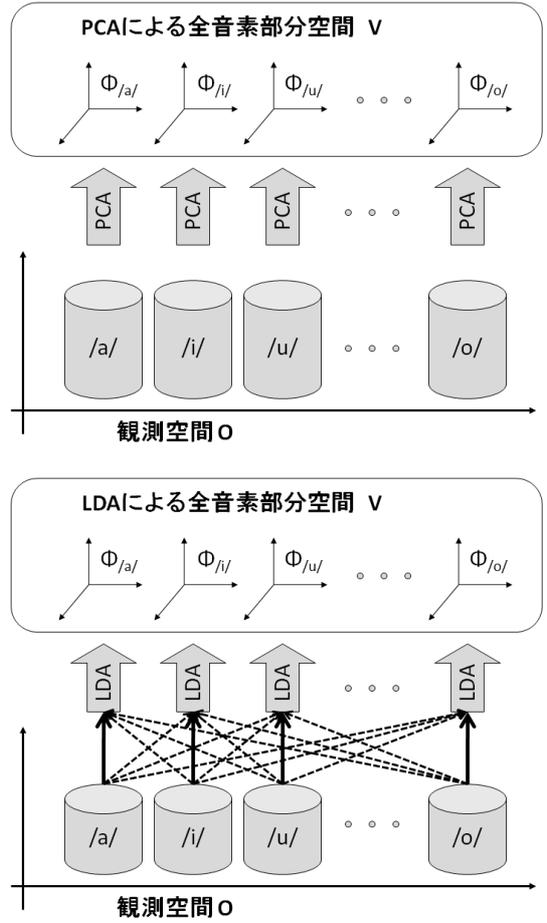


図 1 上：PCA による音素部分空間学習
下：LDA による音素部分空間学習

$$\mathbf{y}_t^i = \Phi^{iT} (\mathbf{x}_t - \bar{\mathbf{x}}^i) \quad (9)$$

$$\Phi^i = (\phi_1^i, \phi_2^i, \dots, \phi_Q^i) \quad (10)$$

ここで全音素部分空間 V と全音素平均ベクトル C を次のように定義する．

$$V = [\Phi^1, \Phi^2, \dots, \Phi^M] \quad (11)$$

$$C = [(\Phi^{1T} \bar{\mathbf{x}}^1)^T, (\Phi^{2T} \bar{\mathbf{x}}^2)^T, \dots, (\Phi^{MT} \bar{\mathbf{x}}^M)^T]$$

V と C を用いて全音素部分空間へ射影された特徴ベクトル \mathbf{y}_t は，次のように求まる．

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{y}_t^1 \\ \mathbf{y}_t^2 \\ \vdots \\ \mathbf{y}_t^M \end{bmatrix} = \begin{bmatrix} \Phi^{1T} [\mathbf{x}_t - \bar{\mathbf{x}}^1] \\ \Phi^{2T} [\mathbf{x}_t - \bar{\mathbf{x}}^2] \\ \vdots \\ \Phi^{MT} [\mathbf{x}_t - \bar{\mathbf{x}}^M] \end{bmatrix} \quad (12)$$

$$= \begin{bmatrix} \Phi^{1T} \mathbf{x}_t - \Phi^{1T} \bar{\mathbf{x}}^1 \\ \Phi^{2T} \mathbf{x}_t - \Phi^{2T} \bar{\mathbf{x}}^2 \\ \vdots \\ \Phi^{MT} \mathbf{x}_t - \Phi^{MT} \bar{\mathbf{x}}^M \end{bmatrix}$$

$$= V^T \mathbf{x}_t - C^T$$

LDA の場合は平均を引く操作はないので， $\mathbf{y}_t = V^T \mathbf{x}_t$ となる．

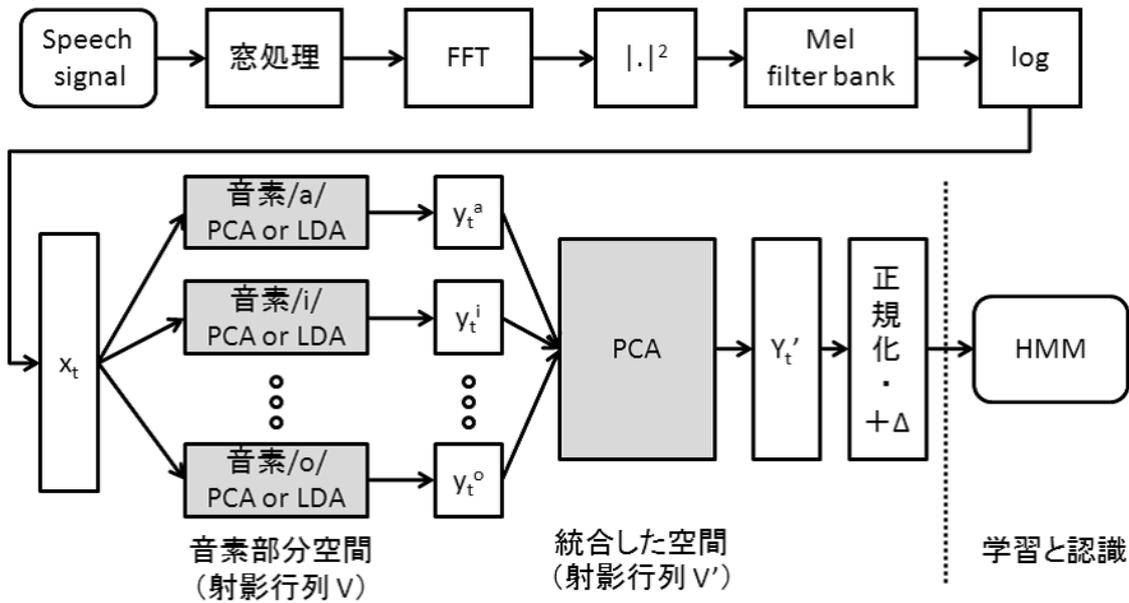


図 2 特徴量抽出過程

3.2 音素部分空間の統合

前項で音素部分空間をまとめた全音素部分空間 V は各音素部分空間の基底ベクトルを並べたにすぎない。この空間はまず次元数が $Q \times M$ (Q は各音素部分空間の次元数, M は全音素数) であり, 各音素間に似ている基底ベクトルを持っている可能性がある。似ているベクトルは無駄であるため, 本研究ではもう一度 PCA を用い, 各音素部分空間の統合 (全音素空間の圧縮) を行う。PCA で統合することで音素と無関連な情報がさらに抑圧される可能性もある。

全音素部分空間 V に射影されたデータ y_t を用いて, 統合のための射影行列 V' を推定し, 次のように最終特徴量ベクトル y'_t を求める。

$$y'_t = V'^T (y_t - \bar{y}) \quad (13)$$

統合された全音素部分空間は $V' = [\phi^1, \phi^2, \dots, \phi^R]$ で, 固有値が大きい順に R 本の固有ベクトルで構成される。

4. 孤立単語音声認識実験

提案手法の有効性を確認するため, クリーン音声と残響音声の孤立単語認識実験を行った。実験条件と実験結果を技術し, 考察を述べる。

4.1 実験条件

評価実験のデータとして, ATR の日本語データベース A-set から男女 2 名ずつ計 4 名に話者のデータを用いた。また, テストデータとして残響音声を作成するためには RWCP 実環境音声・音響データベース [18] より残響インパルス応答を用いた。先行研究 [12] で残響時間を変えながら単語認識を行った結果, 残響時間が 380ms のところで認識率が著しく下がったため, 本研究ではこの残響時間のインパルス応答を用いてクリーン音声に重畳させた。データ数は話者ごとに学習音声データ用 2,620 単語発話, クリーンと残響音声データ用 1,000 単語発話を用いた。

音声信号は 16 bits, 12 kHz でサンプリングされ時間の長さが 32ms のハミング窓を 8ms シフトさせフレーム分析を行った。各フレームに離散フーリエ変換をした後, 32 チャンネルのメル周波数フィルタバンク分析を行った。本研究ではこの 32 次元対数エネルギーを持つ空間をベース特徴量空間とする。以上の音響分析には HTK の toolkits [19] を用いた。

評価実験で比較のため行った特徴量抽出法を以下にまとめる。

- *LogMFB* 上で説明した代数メルフィルタバンク出力。以下の特徴量抽出の入力となる。
- *DCT* MFCC を求める手法。*LogMFB* に対して離散コサイン変換 (Discrete Cosine Transform; DCT) を行う。
- *PCA* 全音声データに対して PCA を用いて射影行列を求め, 線形射影する。
- *LDA* 全音素クラスを判別するため LDA を用いて射影行列を求め, 線形射影する。
- *PPCA* 音素部分空間を PCA を用いて求め, さらに PCA を用いて統合し, 線形射影する。
- *PLDA* 音素部分空間を LDA を用いて求め, さらに PCA を用いて統合し, 線形射影する。

表 1 は各特徴量抽出のパラメータであるフレーム数と次元数を表わしている。*PPCA* と *PLDA* に対しては, 統合前後のパラメータを併記している。

表 1 特徴量空間推定のための手法, フレーム数, 次元数

分析手法	分析フレーム数	次元数
<i>LogMFB</i>	—	32
<i>DCT</i>	—	16
<i>PCA</i>	4,000	16
<i>LDA</i>	100×54	16
<i>PPCA</i>	100×54 / 4,000	16×54 / 16
<i>PLDA</i>	100×54 / 4,000	16×54 / 16

以上で得られた特徴量は短時間内の静的な特徴量であるため、続く HMM ベース識別機には特徴量の基本係数と係数が正規化され入力される。本研究では、クリーンの音素データを用い、54 個のコンテキスト独立のモノフォン HMM を学習した。各 HMM は自己ループの 3 状態で各状態には 4 つのガウス分布が混合されている。HMM の学習と認識にも HTK の toolkits [19] を用いた。

特徴量抽出のための射影行列と音素 HMM は 4 人話者の特定モデルと共通モデルを学習した。

図 2 は提案する特徴量抽出法のフローを表す。

4.2 実験結果

孤立単語認識実験結果を図 3 と図 4 に示す。図 3 は 4 人のクリーンの音声 1000 単語認識結果の平均を表す。特定モデルに対しては、*PLDA* が一番いい結果になり、4 人共通モデルに対しては、従来の手法の *PCA* の結果が一番良かった。特定モデルより、4 人共通モデル認識率が一般的に低い、4 人共通モデルの *PCA* の結果が特定モデルの *LDA* よりいい結果になっている。

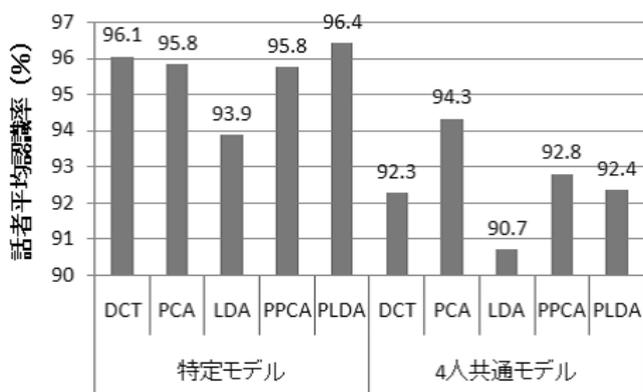


図 3 4 人のクリーン音声孤立単語認識結果

図 4 は 4 人の 380ms 残響音声 1000 単語認識結果の平均を表す。特定モデルの認識では、*PLDA* と *PPCA* の順に高い認識結果を表わしている。4 人共通モデルの認識では、*PPCA* と *PLDA* の順の高い認識結果が得られた。

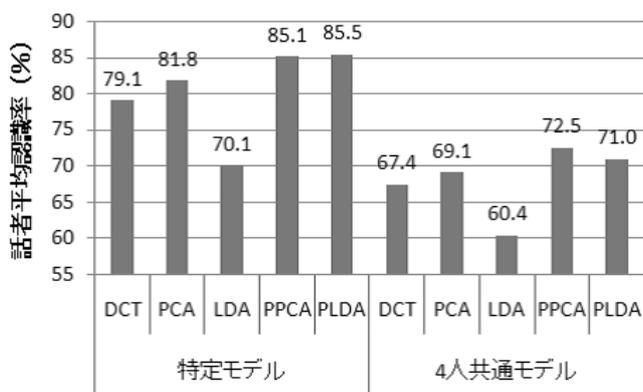


図 4 4 人の 380ms 残響音声孤立単語認識結果

実験結果をまとめると、クリーン音声の 4 人共通モデルで認

識した結果を除けば、提案手法が有効であることが分かる。

4.3 考察

前項の実験結果から、一般的にクリーン音声の特定話者モデルの場合の認識率が一番高く、残響音声の 4 人共通モデルの場合の認識率が一番低い結果が得られた。残響音声には残響成分が混じり、4 人共通モデルには話者の変動成分が混じったので認識率の低下が生じた。

従来手法の *LogMFB* に対する *DCT*, *PCA*, *LDA* を比較してみると、*PCA* はクリーンの特定モデルの場合を除くと、*DCT* より約 2 % の認識率の向上が得られたが、*LDA* は全体的に *DCT* より認識率が下がった。まず、*PCA* の手法は音声の部分空間がうまく推定でき、音声情報と無関連な情報は除去できたと考えられる。反面 *LDA* の場合は音素クラスの識別空間がうまく推定できなかった。一番大きい理由としては、元の次元数が 32 次元でこの空間上で 54 個のクラスを識別する部分空間を推定することに無理があったと考えられる。

次に提案手法である *PPCA* や *PLDA* の *DCT* との差を見てみると、特に残響音声の認識において約 5 ~ 6 % の認識率向上が得られた。また、*PCA* との比較をしてみても、約 3 % の認識率改善ができた。この結果から特徴量空間から各音素部分空間を推定しさらに統合することが、特徴量空間に音素識別的な情報を入れることができ、音声認識に寄与したと考えられる。

最後に *PPCA* と *PLDA* の比較を行う。特定モデルに対しては *PLDA* が *PPCA* より認識率がいいが、4 人共通モデルになると *PPCA* が *PLDA* よりいい結果を示している。これは *PLDA* が特定話者の音素間識別には優れているが、多数話者の音素間識別になると、音素データ内存在する話者変動成分の影響で、音素クラス内共分散が大きくなり、音素間データ音素間識別能力が劣ってしまう。これに対して *PPCA* は特定音素データに対して主成分分析を行うため、得られる基底ベクトルは音素の構造だけでなく、話者の変動性分も含まれるが、各音素データの中の変動成分が似たような構造で現れると仮定すると、後の全音素空間統合時の主成分分析により音素間の共通情報（話者の変動成分）は除去されるので、*PLDA* よりいい性能が得られたと考えられる。このため不特定話者音声認識には *PLDA* より *PPCA* の方がいいといえる。

5. 結論

HMM をベースにした音声認識システムにおける新しい特徴量抽出法を提案し、評価を行った。対数メルフィルタバンク出力は、*PCA* や *LDA* を用いて求めた各音素部分空間へ線形射影され、さらに、*PCA* により統合された空間へ線形射影され、後の HMM のベースの認識システムの入力となる。評価実験により、特徴量空間を対象とする音声の音素情報を用いて変換することが HMM ベース音声認識に有効であることが示された。提案手法はモデルベースの耐雑音・残響手法と組み合わせることで性能の改善が期待できる。今後は、各音素部分空間を確率空間として定義する方法やカーネルトリックを用いた非線形射影、独立成分分析などを用いた新しい部分空間統合法を検討していく予定である。

文 献

- [1] H. Hermansky and N. Morgan, "RASTA Processing of Speech," IEEE Trans. on Speech and Audio Processing, Vol. 2, No. 2, pp. 578-589, 1994.
- [2] C. Avendano, S. Tivrewala, and H. Hermansky, "Multiresolution channel normalization for ASR in reverberant environments," Proc. Eurospeech1997, pp. 1107-1110, 1997.
- [3] W. Li, K. itou, K. Takeda and F. Itakura, "Two-Stage Noise Spectra Estimation and Regression Based In-Car Speech Recognition Using Single Distant Microphone," Proc. ICASSP2005, pp. 533-536, 2005.
- [4] K. Kinoshita, T. Nakatani and M. Miyoshi, "Efficient Blind Dereverberation Framework for Automatic Speech Recognition," Proc. Interspeech2005, pp. 3145-3148, 2005.
- [5] R. Vetter, N. Virag, P. Renevey and J.-M. Vesin, "Single Channel Speech Enhancement Using Principal Component Analysis and MDL Subspace Selection," Proc. Eurospeech99, pp. 2411-2414, 1999.
- [6] S.-M. Lee, S.-H. Fang, J.-W. Hung and L.-S. Lee, "Improved MFCC Feature Extraction by PCA-Optimized Filter Bank for Speech Recognition," Proc. ASRU2001, pp. 49-52, 2001.
- [7] F. Asano, Y. Motomura, H. Asoh and T. Matsui, "Effect of PCA Filter in Blind Source Separation," Proc. ICA2000, pp. 57-62, 2000.
- [8] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," Neural Computation, Vol. 10, pp. 1299-1319, 1998.
- [9] 滝口哲也, 有木康雄, "Kernel PCA を用いた残響下における口バースト特徴量抽出の検討," 情報処理学会論文誌, Vol. 47, No. 6, pp. 1767-1773, 2006.
- [10] 柘植寛, 黒岩真吾, 任福継, 北研二, "音素依存線形判別分析の検討," 日本音響学会 2001 年秋季講演論文集, Vol. 1, pp. 177-178, 2001.
- [11] Jong-Hwan Lee¹, Ho-Young Jung², Te-Won Lee³, Soo-Young Lee¹ "Speech feature extraction using independent component analysis," Proc. ICASSP2000, Vol. 3, pp. 1631-1634. 2000
- [12] 朴玄信, 滝口哲也, 有木康雄, "音素 PCA を用いた残響下における音声特徴量抽出," 電子情報通信学会技術報告誌, SP2007-1, pp. 1-6, 2007.
- [13] Y. Ariki and K. Doi, "Speaker Recognition based on Subspace Method," Proc. ICSLP94, pp. 1859-1862, 1994.
- [14] Y. Ariki, S. Tagashira and M. Nishijima, "Speaker Recognition and Speaker Normalization by Projection to Speaker Subspace," Proc. ICASSP96, Vol. 1, pp. 319-322, 1996.
- [15] P. Ding and L. Zhang, "Speaker Recognition Using Principal Component Analysis," Proc. ICONIP2001, Paper ID 72, 2001.
- [16] 赤津 達也, 加藤 正治, 小坂 哲夫, 好田 正紀, "話者ベクトルを用いた話者識別法における音響モデルの検討," 日本音響学会 2006 年秋季講演論文集, pp. 114-115, 2006 .
- [17] D. X. Sun, "Feature dimension reduction using reduced-rank maximum likelihood estimation for hidden Markov model," Proc. ICSLP1996, pp. 224-247, 1996.
- [18] S. Nakamura, K. Hiyane, F. Asano, T. Nishimura and T. Yamada, "Acoustical Sound Database in Real Environments for Sound Scene Understanding and Hands-Free Speech Recognition," Proc. LREC2000, Vol. 2, pp. 965-968, 2000.
- [19] S. Young et. al., "The HTK Book," Entropic Labs and Cambridge University, 1995-2002 .