音声特徴量抽出のための音素部分空間統合法の検討*

◎朴玄信, 滝口哲也, 有木康雄 (神戸大)

1 はじめに

近年の音声認識システムで一般的に使われている特徴量 MFCC は対数メルフィルタバンク出力に離散コサイン変換をして得られるが，この特徴量空間は音声データに依存することなく一意に決まる．そのため，音声信号に雑音が混入すると，認識性能の低下が起こる．文献 [1] では，部分空間を用いて雑音が混入した観測信号から目的信号を強調する方法が述べられている．我々はこの部分空間手法を用いて，音素間の相関情報を持つ，データ依存型特徴量空間を提案してきた [2]．本稿では [2] をベースに，独立成分分析を用いた音素部分空間の統合法を提案する．また，クリーン音声と残響音声の単語認識実験において提案手法の有効性を示す．

2 線形変換による従来特徴抽出

2.1 主成分分析による特徴抽出

観測音声信号の t 番目短時間フレームのパワースペクトル $\mathbf{x}_t$（q 次元の列ベクトル）を次のように定義する．

$$\mathbf{x}_t = \mathbf{s}_t + \mathbf{n}_t \tag{1}$$

ここで，$\mathbf{s}_t$ はクリーン音声，$\mathbf{n}_t$ は加法性雑音を意味する．クリーン音声と雑音が無相関であると仮定すると，観測信号の主成分分析を行うことで，相互に直交するクリーン音声部分空間と雑音部分空間が得られる．観測信号をクリーン音声部分空間へ射影することで，雑音を抑圧することができる．

以下でアルゴリズムを簡単に説明する．まず，観測信号から N フレームをランダムに選び，共分散行列 S を次式のように求める．

$$S = \frac{1}{N}\sum_{t=1}^{N}(\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})^T \tag{2}$$

$\bar{\mathbf{x}}$ は観測信号の平均ベクトルである．次に，S を

$$S\phi_k = \lambda_k\phi_k, (k = 1, 2, \cdots, q) \tag{3}$$

のように固有値分解する．λ と ϕ は固有値と固有ベクトルである．固有値が大きい順に $p(< q)$ 個の固有ベクトルをクリーン音声部分空間 $\Phi(pq)$ の列ベクトルすると，次のように観測信号からクリーン音声を抽出することができる．

$$\mathbf{y}_t = \Phi^T(\mathbf{x}_t - \bar{\mathbf{x}}) \tag{4}$$

部分空間法による信号フィルタリングの詳細は文献 [1] を参照されたい．

2.2 独立成分分析による特徴抽出

独立成分分析は元の信号源が互いに独立であることを仮定し，観測信号から元の信号を復元する手法である．観測信号 $\mathbf{x}$ の線形混合モデルを $\mathbf{x} = A\mathbf{s}$ であるとする．A は混合行列，$\mathbf{s}$ は元の信号である．復元行列 W を考えると，$W = A^{-1}$ であれば，完全に $\mathbf{s}$ が復元されるが，観測時に A に関する情報はなく，$\mathbf{x}$ だけで復元を行わなければいけない．復元行列 W により復元されたベクトル $\mathbf{y}$ を次のように表わす．

$$\mathbf{y} = W\mathbf{x} \tag{5}$$

復元行列 W は $\mathbf{y}$ の要素間の独立性が最大になるように推定する．独立性の基準は４次のキュムラントを用い，最急降下法で推定を行う事ができる．文献 [3] では，様々な独立性の基準と推定手法が述べられているので参照されたい．

3 提案手法

近年音声認識システムで一般的に使われている特徴量 MFCC は音声データ非依存であるため，観測信号に音声と雑音が混じった場合はシステムの性能低下が起こる．本研究では，クリーン音声データを用いて特徴量空間を設計することで，雑音にロバストなデータ依存型特徴量を提案する．まず，主成分分析をクリーンな各音素データ集合に対して行い，音素部分空間を求める．次に，各音素部分空間へ射影されたベクトル集合に対して独立成分分析を行い，各音素部分空間を統合する変換行列を求める．前段の主成分分析は雑音のフィルタリングと音素情報抽出の役割をし，後段の独立成分分析は各要素が独立で，各音素間の相関を表わす情報を抽出する．

3.1 主成分分析による音素部分空間

まず，観測信号 $\mathbf{x}$ を対数メルフィルタバンク出力とし，M 個の音素の中 i 番目音素の部分空間を表わす射影行列 Φ^i と平均ベクトル $\bar{\mathbf{x}}^i$ を主成分分析により求める．対数演算により，スペクトル上の乗法性雑音を足し算で表すことができるので，主成分分析により乗法性雑音も除去される．すべての音素を表わ

*Integration of Phoneme-subspaces for Speech Feature Extraction
by Hyunsin PARK, Tetsuya TAKIGUCHI, and Yasuo ARIKI (Kobe Univ.)

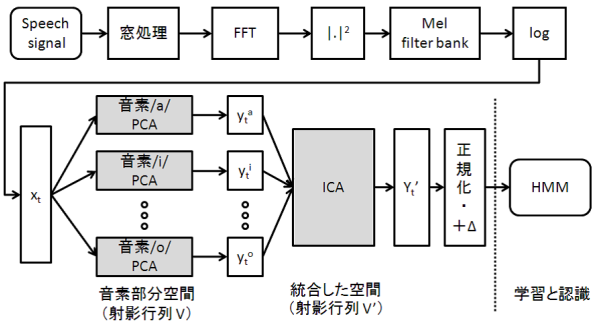

Fig. 1　特徴量抽出過程

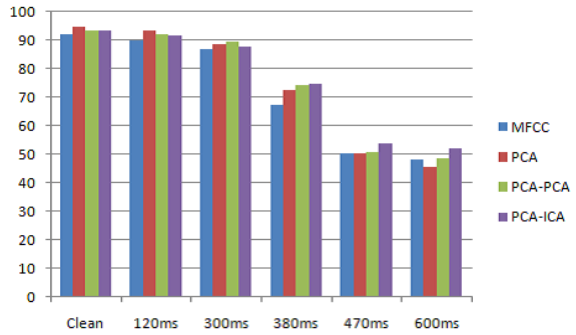

Fig. 2　孤立単語音声認識結果：４人の平均認識率
横軸：残響時間，縦軸：認識率 [%]

す空間の射影行列 V と射影された平均ベクトル C を次のようにまとめる．

$$V = [\Phi^1, \Phi^2, \cdots, \Phi^M]$$
$$C = [({\Phi^1}^T \bar{\mathbf{x}}^1)^T, ({\Phi^2}^T \bar{\mathbf{x}}^2)^T, \cdots, ({\Phi^M}^T \bar{\mathbf{x}}^M)^T] \quad (6)$$

すべての音素部分空間へ射影されたベクトル $\mathbf{y}_t$ は次のように表わすことができる．

$$\mathbf{y}_t = V^T \mathbf{x}_t - C^T \quad (7)$$

3.2　独立成分分析による音素部分空間の統合

式 (7) の $\mathbf{y}_t$ の集合に対して，独立成分分析を行い射影行列 V' を求める．本研究では，文献 [3] の FastICA を用いて，V' を求めた．V' により射影されるベクトル $\mathbf{y}'_t$ は次のようになる．

$$\mathbf{y}'_t = V' \mathbf{y}_t \quad (8)$$

提案手法による特徴量抽出の流れを図 1 で示す．対数メルフィルタバンク出力に対して式 (7) と式 (8) の線形変換を行なう．さらに，正規化と Δ 係数追加を行い，音声特徴量として使用する．学習と認識には HMM を用いる．

4　評価実験

4.1　実験条件

評価実験として，孤立単語音声認識を行った．データは ATR の SpeechDatabase SET-A から男女 2 名ずつ 4 人のデータを用いた．学習には各話者の 2620 単語を，テストには各話者 1000 単語を用いた．音声信号は 16 bit / 12 kHz でサンプリングされ，32 ms のハミング窓を 8 ms シフトさせながらフレーム化した．対数メルフィルタバンク出力は 32 次元にした．対数メルフィルタバンク出力に対して離散コサイン変換した MFCC，従来の主成分分析を適用した PCA，主成分分析を用いた音素部分空間統合による特徴量 PCA-PCA，提案手法である独立成分分析を用いた音素部分空間統合による特徴量 PCA-ICA(すべての特徴量は 16+Δ 16 次元) を用いて，3 状態 4 混合の 54 個音素 HMM の学習と認識を行った．各音素部分空間の次元は全て１６次元にした．学習した統合音素部分空間と HMM はクリーン音声データを用いた４人共通モデルである．テストはクリーン音声とインパルス応答を畳み込みした残響音声を用いて行った．

4.2　実験結果と考察

評価実験の結果を図 2 に示す．提案手法の PCA-ICA は全ての残響時間において，MFCC より高い認識率が得られ，残響時間が 380ms 以上の時は，全ての特徴量の中で一番いい結果が得られた．提案手法の前段の主成分分析により残響が抑圧され，後段の独立成分分析により，独立に動く発話器官の情報が得られたと考えられる．今回の実験では，各音素部分空間の次元数を全て 16 次元にしたため，統合時に余分な音素部分空間の次元の影響で音素間の相関情報がうまくとれず，残響時間が短い場合，従来の PCA よりやや低い認識結果が得られたと考えられる．

5　おわりに

主成分分析を用いて各音素の部分空間を学習した後，独立成分分析による部分空間統合を行い，音声特徴量を抽出する手法を提案した．また，評価実験により有効性を確かめた．今後は，話者数を増やした音声データを用いて部分空間の学習と実験を行い，また，各音素部分空間の次元数や統合した部分空間の次元数の最適解について検討する予定である．

参考文献

[1] K. Hermus *et al.*, EURASIP Journal on Advances in Signal Processing, vol. 2007, Article ID 45821, 15 pages, 2007.
[2] 朴　他，信学技報，SP2007-137，2007
[3] A. Hyvarinen, E. Oja, Neural networks, Vol. 13, pp. 411-430, 2000