

PCA を用いた音素ベクトル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)

1 はじめに

実環境で使われる音声認識システムには音声信号に雑音が入り込み認識性能が低下する．音声特徴量として一般的に用いられる MFCC は雑音にロバストではないため，雑音抑制手法がともに用いられる．

観測される音声信号には様々な情報が入っており，特定の情報を抽出するため部分空間法 [1] を用いる手法がある．本論文では各音素の部分空間を PCA を用いて求め，各部分空間をマージした空間を定義し，このマージした空間へ射影した値を特徴量として用いる手法を提案する．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 提案手法

2.1 固有音素空間

先行研究 [3] で提案した音素間の関係情報を持つ特徴量空間である固有音素空間 (Eigen Phoneme Space : EPS) について説明する．

観測信号の中の特定情報を強調する手法として PCA を用いる部分空間法がある．我々の先行研究 [2] では全音素データに対して PCA を用いた音声強調手法を提案した．さらにその後の研究 [3] では音声データを音素データに分け，各音素データ集合に対して PCA を行い，各音素の構造を表す部分空間を求める方法を提案した．各部分空間をマージすることで，マージされた空間には各音素間の情報が含まれるようになる．この各音素部分空間がマージされた空間を固有音素空間 (EPS) とする．また元の観測ベクトルを固有音素空間 (EPS) へ射影して得られる新しい特徴量を音素ベクトル (Phoneme Vector : PV) とする．

Fig. 1 は観測データから固有音素空間 (EPS) を求める概念を図示している． $\Phi_{/a/}$, $\Phi_{/i/}$, \dots は各音素データ集合に対して PCA を行い得られた音素部分空間である．固有音素空間 (EPS) は各音素部分空間をマージした空間である．

2.2 PCA を用いた音素ベクトル

PCA を用いた音素ベクトルの定式化を行う． i 番目の音素データの共分散行列 S^i は

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

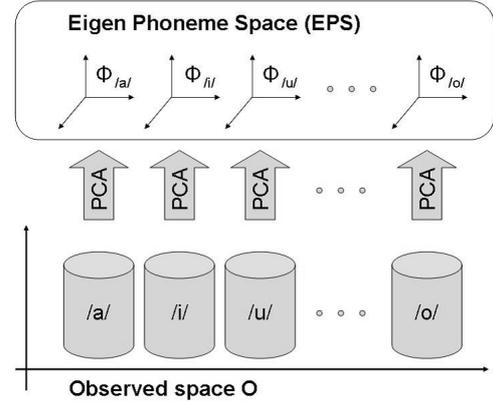


Fig. 1 Observed space and Eigen Phoneme Space のように定義される． \mathbf{x}_t^i と $\bar{\mathbf{x}}^i$ は i 番目音素データの観測ベクトルと平均ベクトル， M は全音素数を表す． S^i を次のように固有値分解する．

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

λ_k^i は固有値， ϕ_k^i は固有ベクトルであり， P は \mathbf{x}_t^i の次元数である．固有値が大きい順に Q 個の固有ベクトルを用いて i 番目音素の部分空間を求める．観測ベクトル \mathbf{x}_t^i はこの i 番目音素の部分空間へ射影され \mathbf{y}_t^i が次のように求まる．

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

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

ここで， V (全音素部分空間行列) と C (全音素平均ベクトル) を次のように定義する．

$$V = [\Phi^1, \Phi^2, \dots, \Phi^M] \quad (5)$$

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

音素ベクトル \mathbf{y}_t は V と C を用いて次のように求まる．

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{y}_t^1 \\ \mathbf{y}_t^2 \\ \vdots \\ \mathbf{y}_t^M \end{bmatrix} = \begin{bmatrix} \Phi^{1T} [\mathbf{x}_t - \bar{\mathbf{x}}^1] \\ \Phi^{2T} [\mathbf{x}_t - \bar{\mathbf{x}}^2] \\ \vdots \\ \Phi^{MT} [\mathbf{x}_t - \bar{\mathbf{x}}^M] \end{bmatrix} \quad (6)$$

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

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

* Feature extraction with Phoneme Vector based on PCA. by Hyunsin PARK, Tetsuya TAKIGUCHI, and Yasuo ARIKI (Kobe Univ.)

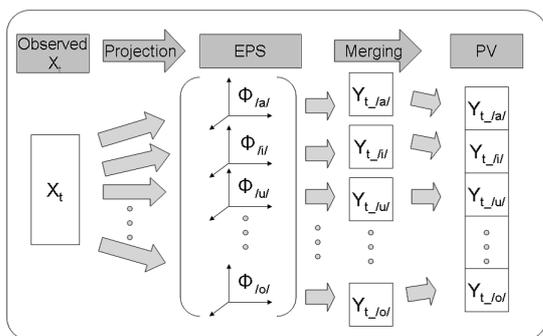


Fig. 2 Phoneme Vector (PV) extraction

式 (6) で求めた y_t を音声特徴量と使うには次元数 ($Q \times M$) が大きいので, y_t に対して PCA を行い, 次元圧縮を行った y'_t を求める.

$$y'_t = V'^T (y_t - \bar{y}) \quad (7)$$

ここで $V' = [\phi^1, \phi^2, \dots, \phi^R]$ は固有値が大きい順で R 個の固有ベクトルで得られる圧縮空間を意味する. Fig. 2 に音素ベクトルの抽出方法を図示する.

3 単語認識実験

3.1 実験条件

メル周波数フィルタバンク出力に対して DCT, 全音素 PCA, 提案手法 (圧縮された音素ベクトル) を適用して得られる特徴量を MFCC, PCA, PV とし比較実験を行う. 元のメル周波数フィルタバンク出力は 32 次元であり, 各特徴量は基本係数 16 次元とそのデルタ係数を用い 32 次元で共通である. 提案手法の特徴量は, 圧縮前の各音素空間の次元数 (Q) を 5 から 21 まで 2 ずつ増やしながらか求めた. 音声データは ATR 日本語音素バランス文 A-SET から男女各 2 名の単語発話データを用いた. 学習データとしては話者ごと 2,620 単語を用いて, 54 音素のモノフォン HMM を求めた. テストデータは話者ごと 1,000 単語のクリーンと 380 ms 残響音声を用いて行った. その他の音響分析, PCA, 音響モデルなどの条件は文献 [3] と同じである.

3.2 単語認識実験結果

Fig. 3 は圧縮前の音素ベクトルにおいて, 各音素部分空間の次元数 (Q) を変えながらクリーンと残響音声認識実験を行った結果である. 4 人の単語認識率の平均を表している. クリーン音声認識率は各音素部分空間の次元数にそれほど影響を受けていないが, 残響音声認識率は各音素部分空間の次元数が大きくなるほど向上している. Fig. 4 は提案手法の特徴量で一番良い認識率と従来特徴との比較である. クリー

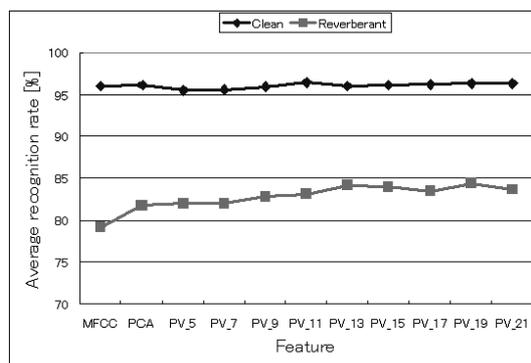


Fig. 3 Recognition result for clean and reverberant speech by adjusting the number of phoneme subspace dimension (Q)

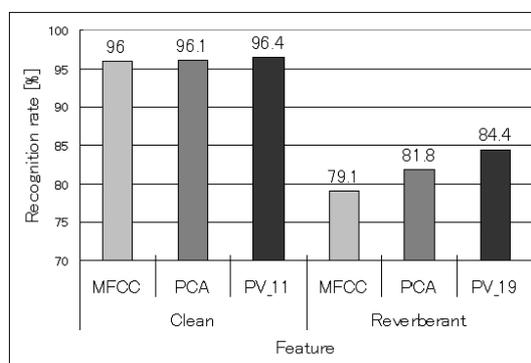


Fig. 4 Recognition result for clean and reverberant speech

ン音声と残響音声に対して認識率の向上が得られている.

4 おわりに

PCA を用いた音素間相関情報を持つ特徴量空間と特徴ベクトルを提案し, 実験を行い有効性を確かめた. 提案手法は従来 MFCC 特徴量ベースの耐雑音処理と組み合わせることができ, 認識率の向上が期待できる. 今後は, PCA 以外の統計手法や確率手法を用いた固有音素空間と音素ベクトルを研究する予定である.

参考文献

- [1] Y. Ariki and K. Doi, Proc. ICSLP94, 1859-1862, 1994.
- [2] 滝口, 有木, 情処論, Vol. 47, 1767-1773, 2006.
- [3] 朴, 滝口, 有木, 信学技報, Vol. 107, No. 77, 1-6, 2007