

Random Projection を用いた音声特徴量抽出 における Random Matrix の統合*

吉井麻里子，滝口哲也，有木康雄 (神戸大)，Jeff Bilmes (University of Washington)

1 はじめに

Random Projection とは，元の特徴量空間における任意の2点間のユークリッド距離が，写像先の特徴量空間においても高い確率で保存される，という性質を持つ空間写像の手法で，その変換写像行列の各要素がある確率分布に従うランダムな値として定義される点に特徴を持つ．我々はこのランダムプロジェクトを用いて音声特徴量抽出が行えないかと研究を行ってきた [1]．本稿では複数のランダム写像行列を用いて音声特徴量を変換し，各々の特徴量を用いて音声認識を行い，その認識結果を投票により統合することで最適な認識結果を得る手法を報告する．

2 Random Projection

ランダムプロジェクトは n 次元ユークリッド空間から k 次元ユークリッド空間へランダムに写像する空間写像の手法である．ある n 次元の元特徴量ベクトル y が与えられたとき， k 次元 ($k \leq n$) の変換後の特徴量ベクトル x は次のように表わされる．

$$x = Ry$$

ここで R は $n \times k$ の写像行列である．ランダムプロジェクトでは，任意の2点間の距離が高い確率で $(1 \pm \epsilon)$ に収まることが証明されている ($0 \leq \epsilon \leq 1$)．また，写像行列 R は確率的にある値をとる行列として定義されるが， R の各値が $N(0,1)$ に従うランダムな値からなるとき，その距離保存の性質が成り立つことが示されている [2, 3]．本稿では次のようにランダム写像行列 R を設定する．

- 標準正規分布 $N(0,1)$ に従う要素を持つ $n \times k$ の行列 R を作成する．
- グラムシュミットの直交化手法を用いて R を直交化し，列ベクトルを大きさ1で正規化する．

ランダム写像行列 R は，標準正規分布 $N(0,1)$ から無限に生成することができる．

3 音声特徴量抽出

2章で示した手法により，ランダム写像行列 R は無限に生成することができる．従って，ランダムプロジェクトを用いて音声特徴量抽出を行う際，無限

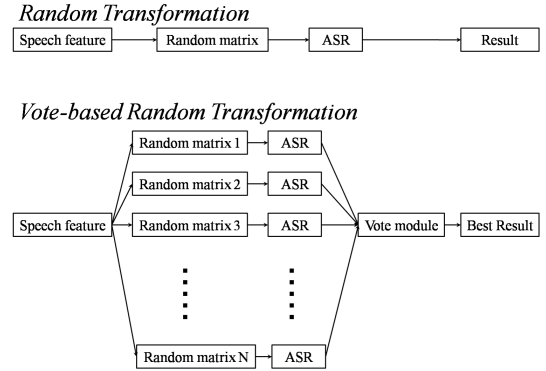


Fig. 1 Overview of Random Transformation

のランダム写像行列から最も音声認識に適したランダム写像行列を見つけるか，もしくは複数のランダム写像行列から最適な音声認識結果を求める必要がある．本稿では最適な認識結果を得るために，ROVER[5]を用いた特徴量統合の手法を検討する．

3.1 音声特徴量変換

ROVER とは，複数の音声認識システムから得た認識結果に対して投票を行い，最適な認識結果を出力する手法である．Fig. 1 に本稿で提案する統合システムの流れ図を示す．上図は基本的なランダムプロジェクトを用いた音声特徴量抽出手法を示している．下図では，複数のランダム写像行列を用いて特徴量抽出を行い，それぞれ音声認識を行った後に ROVER を用いて統合を行っている．このように統合を行うことにより，ランダム写像行列に優劣を付けることなく最適な認識結果を得ることが可能となる．

3.2 音声特徴量

本稿で用いる音声特徴量を示す．

- (1) MFCC(12 次元) $\rightarrow$ RP(12 次元)
- (2) MFCC + Δ + $\Delta\Delta$ (36 次元) $\rightarrow$ RP(36 次元)
- (3) MFCC(12 次元) $\rightarrow$ RP(12 次元) + Δ + $\Delta\Delta$ (36 次元)
- (4) Gabor(60 次元) $\rightarrow$ RP(30 次元)
- (5) Gabor + Δ + $\Delta\Delta$ (180 次元) $\rightarrow$ RP(30 次元)

(1)，(2) では MFCC 特徴量やその Δ $\Delta\Delta$ に対してランダムプロジェクトを行っている．(3) では MFCC

*Random-Projection-Based Feature Transformation Using Random-Matrices Combination, by Mariko Yoshii, Tetsuya Takiguchi, Yasuo Ariki (Kobe University), Jeff Bilmes (University of Washington)

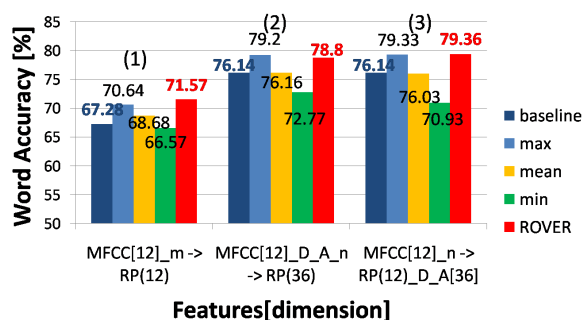


Fig. 2 Word recognition rate of MFCC-RP

特徴量に対してランダムプロジェクションを行い，その $\Delta + \Delta\Delta$ を計算している．(4)，(5)では2-D Gabor特徴量 [4] を音声特徴量として用い，ランダムプロジェクションによって次元を削減して音声特徴量としている．

4 評価実験

4.1 実験条件

提案手法の有効性を示すために，自動車内音声認識の評価用データベース CENSREC-3[6] を用いて単語音声認識実験を行う．音声認識評価環境には Condition4 を用い，その学習データはアイドリング走行時の遠隔マイクロホン音声 3,608 発話，評価データは低速，高速走行時の遠隔マイクロホン音声 8,836 発話である．評価対象語彙数は 50 単語からなり，学習音声は音素バランス文となっている．音声の標準化周波数は 16kHz，語長 16bit であり，音響モデルは音素 HMM である．また，各 HMM の状態数は 5，状態あたりの混合分布数は 32 である．

4.2 単語認識実験

単語認識実験の結果を Fig. 2, 3 に示す．それぞれ，左端にベースラインの変換前の認識率，中央にランダム写像行列を 100 種類用いた最大認識率，平均認識率，最小認識率を示し，右端の値が ROVER を用いて特徴量統合を行った際の認識率である．ランダム写像行列によっては元の特徴量空間よりも音声認識に適した空間が得られることがわかったが，認識率が低くなる空間も存在することがわかった．しかしながら ROVER を用いて複数のランダム写像行列を統合することにより，安定して高い認識率が得られることが示された．

5 おわりに

本稿では，ランダムプロジェクションを用いた新しい特徴量抽出手法について提案した．我々は複数のランダム写像行列を用いて音声特徴量を変換し，各々

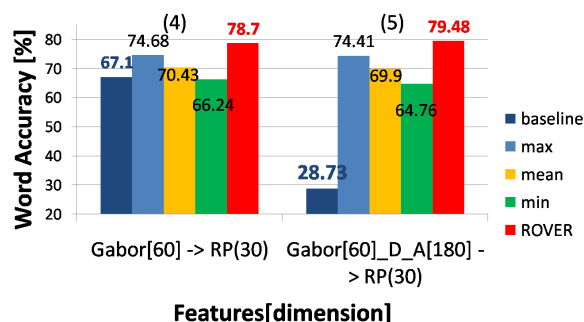


Fig. 3 Word recognition rate of Gabor-RP

のランダム写像に対する音声認識結果に投票を行うことで，最適な認識結果を求めた．ランダムプロジェクションは単純で容易に得られる写像行列から空間写像を行うにも関わらず，認識に有益な特徴量空間を生成することが示された．今後は，多くのランダムプロジェクション特徴量を統合する方法をさらに考えると共に，有用なランダム写像行列とそうでないランダム写像行列を見分ける方法や，一つのランダム写像行列の中から，認識に適した次元を選び出す手法を考えていく必要がある．より有効なランダムプロジェクションの活用方法を提案していきたい．

参考文献

- [1] 吉井，他，ランダムプロジェクションを用いた音声特徴量抽出，音講論（春），pp. 159-160，2009.
- [2] S. Kaski. “Dimensionality reduction by random mapping,” In Proc. Int. Joint Conf. on Neural Networks, volume 1, pp. 413-418, 1998.
- [3] E. Bingham, H. Mannila, “Random projection in dimensionality reduction: applications to image and text data,” In Proc. of the seventh ACM SIGKDD Int. Conf. on Knowledge discovery and data mining, pp. 245-250, 2001.
- [4] M. Kleinschmidt and D. Gelbart, “Improving word accuracy with Gabor feature extraction,” ICSLP, 2002.
- [5] J. G. Fiscus, “A post-processing system to yield reduced word error rates: Recogniser output voting error reduction (ROVER),” Proc. IEEE ASRU Workshop, pp. 347-352, 1997.
- [6] 藤本，他，実走行車内単語音声データベース CENSREC-3 と共通評価環境の構築，第 55 回音声言語情報処理研究会 (SLP)，pp. 41-46, 2005.