Vol.2011-SLP-87 No.15
2011/7/23

# Eigen-SPLICEを用いた雑音環境下における音声認識

千々岩　圭吾[†1]　鈴木雅之[†2]　齋藤大輔[†1]
峯松信明[†1]　広瀬啓吉[†1]

特徴量変換による雑音抑制手法であるSPLICEは比較的少ない計算量で高い性能を発揮する．しかし，このSPLICEにも雑音の変動に頑健でないという問題がある．今回は変換方法そのものを未知の雑音環境に適応することを試みた．本報告では，主成分分析を用いて，未知の雑音環境下において推定すべきパラメータを削減し，少数の適応データで適応できる手法を提案する．また，AURORA–2 データベース testb セットにおいて提案手法を評価し，理想的な適応データが得られた場合には12.0%，粗い雑音推定を用いた場合でも1.9%の誤り削減率を得られた．

# Speech recognition in noisy environments using Eigen-SPLICE

Keigo Chijiiwa,[†1] Masayuki Suzuki,[†2]
Daisuke Saito,[†1] Nobuaki Minematsu[†1]
and Keikichi Hirose[†1]

SPLICE is one of the speech enhancement methods using feature conversion, which shows a high performance with a relationally small amount of calculation. It supposes that input noisy environments are static and similar to training data environments, so it is not guaranteed to work well in unseen environments. Therefore, we propose a new method to adapt conversion functions to work well in unseen environments. The proposed method reduces the number of parameters through Principal Component Analysis, so that the adapted conversion function is obtained with a smaller amount of adaptive data. Experiments show 12.0% reduction of error rate in the ideal adaptive condition and 1.9% reduction even in unfavorable conditions.

†1 東京大学 情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo
†2 東京大学　工学系研究科
Graduate School of Engineering, The University of Tokyo

## 1. はじめに

近年，スマートフォンやカーナビゲーションシステムの普及により，雑音環境下における音声認識の必要性が高まってきている．しかし，雑音によって歪められた音声は，クリーンな音声を用いて学習した音響モデルとのミスマッチを生じ，認識精度が著しく低下するという問題がある[1)]．

この問題に対応するため，独立成分分析[2)]やスペクトルサブトラクション[3)]などの手法が多く提案されている．その中でも音声認識に用いる特徴量に対するもので，特徴量の統計的性質を表すモデルを用いて特徴量を変換する手法が着目されている．これは，音声特徴量の統計的な性質を混合正規分布 (GMM:Gaussian Mixture Model) などで事前に学習し，そのモデルを用いて歪んだ特徴量をクリーンな特徴量に変換しており，高精度な雑音抑制を実現する．代表的なものとしては，雑音付加音声の特徴量の分布をGMMで学習し，クリーン音声と雑音付加音声のパラレルデータから変換を学習するSPLICE(Stereo-based Piecewise LInear Compensation for Environments)[4)]，クリーン音声の特徴量のGMMと，適応用の雑音付加音声特徴量からベクトルテーラー展開 (Vector Taylor Series:VTS) を用いて適応する手法[5)]などがある．これらは雑音環境下の音声認識用データベースAURORA-2[6)]において高い性能を示している[5)]．

しかし，これらの特徴量変換による雑音抑制手法にもいくつかの問題がある．まずVTSを用いた手法については，計算の単純な対数スペクトルなどの特徴量においては計算量が小さくて済むが，MFCCなどのより複雑な特徴量に関しては，多くの計算量が必要になるという問題がある．一方，SPLICEは定常雑音環境を暗に仮定しており，非定常雑音環境下においては十分な性能を発揮することが保証されていない．SPLICEのこの問題点を解決するために，まず入力音声の雑音環境の種類を推定し，その推定結果の雑音環境に応じたGMMと区分的線形変換を用いて，雑音抑制するEMS(Evironmental Model Selection)の手法も提案されている．しかし，この手法も事前に学習した変換方法でしか変換できないため，未知の雑音環境下において高い性能を発揮することが保証されていない．

そこで今回は，特徴量の変換方法を入力の雑音環境に対して適応することで，未知の雑音環境下においても十分な性能を発揮させることを試みる．変換方法を適応するためには，本来であれば各コンポーネント毎に高次元の補正ベクトルのパラメータを推定する必要があ

り，多くの計算量を要してしまう．それを避けるため，提案手法では変換を表すベクトルに主成分分析を施すことで推定すべきパラメータを削減する．これにより，少量の適応データでも適切な特徴量変換を推定できるようにした．

本報告では，まず研究のベースとなった SPLICE について詳しく説明する．その上で，提案手法の手順を具体的に述べ，その有効性を実験的に示す．最後に，現状の問題点と今後の展望について説明する．

## 2. SPLICE

この節では，SPLICE[4] について詳しく説明する．SPLICE は，雑音付加音声の特徴量の統計的な性質を GMM でモデル化し，その GMM の各コンポーネント毎に特徴量を線形変換することで，雑音付加音声からクリーン音声への変換を実現する．これは，雑音付加音声の特徴量からクリーン音声の特徴量への非線形変換を，区分的線形変換によって近似的に実現している．SPLICE は，変換を学習する学習段階と，学習した変換を用いた強調段階に分かれている．以下，SPLICE の仮定や手順について見ていく．

### 2.1 仮　　定

SPLICE は，まず雑音環境下の音声の特徴量ベクトルの分布が GMM で表現出来ると仮定している．その分布を EM(Expectation-Maximization) アルゴリズムによって学習する．

$$p(\boldsymbol{y}) = \sum_s p(\boldsymbol{y}, s) = \sum_s p(\boldsymbol{y}|s)p(s), \text{where}$$
$$p(\boldsymbol{y}|s) = N(\boldsymbol{y}; \mu_s, \Sigma_s) \tag{1}$$

ここで $\boldsymbol{y}, s, \mu, \Sigma$ はそれぞれ，雑音付加音声の特徴量ベクトル，GMM の各コンポーネントのインデックス，平均，分散を表す．

ここで，以下の仮定を設ける．雑音付加音声の特徴量ベクトルとそれが所属するコンポーネントが与えられたとき，その変換後のクリーン音声の確率分布も正規分布によって表され，その正規分布の平均は雑音付加音声のアフィン変換によって表されると仮定する．つまり次式で表される仮定を設ける．

$$p(\boldsymbol{x}|\boldsymbol{y}, s) = N(\boldsymbol{x}; \boldsymbol{A}_s\boldsymbol{y} + \boldsymbol{r}_s, \Gamma_s) \tag{2}$$

ここで，$\boldsymbol{x}, \boldsymbol{r}_s, \boldsymbol{A}_s, \Gamma_s$ はそれぞれ，クリーン音声の特徴量ベクトル，コンポーネント $s$ における平均ベクトルの補正ベクトル，線形変換を表す行列，分散を表す．

### 2.2 変 換 手 法

以上のような仮定を設けることによって，ある入力の雑音付加音声の特徴量ベクトルが与えられたときの，出力のクリーン音声の特徴量は，下記の期待値として推定することができる．

$$\hat{\boldsymbol{x}} = E_x[\boldsymbol{x}|\boldsymbol{y}] = \sum_s p(s|\boldsymbol{y})E_x[\boldsymbol{x}|\boldsymbol{y}, s] \tag{3}$$

ここで，(2) を用いると，

$$E_x[\boldsymbol{x}|\boldsymbol{y}, s] = \boldsymbol{A}_s\boldsymbol{y} + \boldsymbol{r}_s \tag{4}$$

となる．よって，雑音付加音声の特徴量から推定したクリーン音声の特徴量は，以下のようになる．

$$\hat{\boldsymbol{x}} = \sum_s p(s|\boldsymbol{y}) \left(\boldsymbol{A}_s\boldsymbol{y} + \boldsymbol{r}_s\right) \tag{5}$$

この変換は，入力が GMM のどのコンポーネントに所属しているかという事後確率を求め，その事後確率を重みとして各コンポーネントでの変換結果の重み付け和と解釈することが出来る．

### 2.3 学　　習

以上の特徴量の変換に必要なコンポーネント $s$ における $\boldsymbol{r}_s, \boldsymbol{A}_s$，つまり $\boldsymbol{y}, s$ が分かったときの条件付き確率分布 $p(\boldsymbol{x}|\boldsymbol{y}, s)$ は最尤推定で，クリーン音声とそれに雑音が付加された雑音付加音声の対を用いて以下のように学習する．

$$\boldsymbol{r}_s = \frac{\sum_t p(s|\boldsymbol{y}_t)(\boldsymbol{x}_t - \boldsymbol{y}_t)}{\sum_t p(s|\boldsymbol{y}_t)},$$
$$\boldsymbol{A}_s = \frac{\sum_t p(s|\boldsymbol{y}_t)\left(\boldsymbol{x}_t - \boldsymbol{r}_s\right)\boldsymbol{y}_t^{\mathrm{T}}}{\sum_t p(s|\boldsymbol{y}_t)\boldsymbol{y}_t\boldsymbol{y}_t^{\mathrm{T}}}, \text{where}$$
$$p(s|\boldsymbol{y}_t) = \frac{p(\boldsymbol{y}_t|s)p(s)}{\sum_s p(\boldsymbol{y}_t|n)p(s)} \tag{6}$$

このとき $t$ は特徴量系列のインデックスである．あるコンポーネントの補正関数は $\boldsymbol{r}_s$ は，雑音付加音声の特徴量がそのコンポーネントに所属する確率を重みとして，雑音付加音声とクリーン音声の差分の重み付け平均とすることが出来る．また，そのときの所属する事後確率は，ベイズの定理を用いて求めることが出来る．このときのクリーン音声と雑音付加音声の対，パラレルデータは実環境では口元のマイクで録音した歪が殆ど無い音声と離れたところで録音した歪んだ音声というようにして得ることが可能である．今回用いた AURORA–2 データベースではシミュレーションによって，クリーン音声に雑音を重畳している．

この SPLICE は用いる GMM の混合数を上げていくと，つまり特徴量空間をより細かく

分割していくと，行列 $\boldsymbol{A}_s$ を単位行列にしても十分な効果が得られることが知られている[4]．ただし，本報告では混合数を抑えたため，$\boldsymbol{A}_s$ も非単位行列として推定した．

### 2.4 SPLICE の問題点

以上の SPLICE は定常雑音環境を暗に仮定しており，非定常雑音環境下においては十分な性能を発揮することが保証されていない．そこで， 環境毎に GMM および区分的線形変換の方法を切り替える EMS が SPLICE の改善手法として提案されている[7]．これは， 各々の学習環境毎に GMM を学習し，その GMM を用いて学習環境毎の区分的線形変換を学習する．そして変換の際には， 入力系列 $\boldsymbol{y}_t$ がどの環境 $e$ に依存しているかを推定し，

$$\hat{e} = \underset{e}{\operatorname{argmax}}\, p(\boldsymbol{y}_t|e) \tag{7}$$

もっとも尤度の高い環境の GMM とその区分的線形変換関数を用いて変換する．このとき，環境の推定誤差を抑えるために，推定された環境の系列を時間方向でスムージングする．こうすることで，非定常な雑音環境においてもその雑音環境に適した変換を用いて雑音除去することが出来るようになる．

しかし，この EMS も事前に学習した変換方法でしか変換出来ないため，未知の雑音環境下において高い性能を発揮することが保証されていない．そこで，今回は入力環境に対して変換関数のパラメータを適応することを試みる．具体的には，少数の適応データで変換関数を適応できるように， 変換関数を表すベクトルに主成分分析を施し，推定すべきパラメータを削減する．

## 3. 提 案 手 法

提案手法は，区分的線形変換のパラメータを入力環境に対して適応することで，未知雑音環境下でも性能を発揮することを目指す．また，少数のデータでも適応出来るように，変換を表すベクトルに主成分分析を施すことで，推定すべきパラメータを削減する．今回は簡単のため各コンポーネント毎の変換関数のうち，補正ベクトル $\boldsymbol{r}_s$ の項のみについて適応した．主成分分析によって推定すべきパラメータを削減する手法は，Eigen–MLLR[8] や固有声に基づいた性質変換[9] などで広く用いられている．しかし，提案手法は確率分布のパラメータではなく変換関数のパラメータを推定するので，入力音声だけでなく雑音付加音声とクリーン音声のパラレルデータを必要とする点でこれらの手法と異なる．ところが，未知環境においてパラレルデータを得ることは難しい．そのため提案手法では，雑音付加音声から雑音のみの区間を抽出し，それを手持ちの学習データのクリーン音声に重畳することで擬似的にパラレルデータを作成する．

### 3.1 主成分の学習

まず，従来の SPLICE を用いて学習データの中の全ての環境に共通な変換関数の行列項として $\boldsymbol{A}_s^0$ を学習する．次に，学習データの中の特定の種類・SNR の雑音環境での変換関数の補正ベクトル項 $\boldsymbol{r}_s^i$ を次式のように求める．ただし，このとき添字 $i$ は特定の種類・SNR の雑音環境を表すインデックスである．

$$\hat{\boldsymbol{r}_s^i} = \underset{\boldsymbol{r}_s^i}{\operatorname{argmin}} \sum_t \sum_s p(s|\boldsymbol{y}_t) \left\{\boldsymbol{x}_t - (\boldsymbol{A}_s^0 \boldsymbol{y}_t + \boldsymbol{r}_s^i)\right\}^2 \tag{8}$$

こうすることで学習データ中の特定の環境のための変換関数 $\boldsymbol{A}_s^0, \boldsymbol{r}_s^i$ を得る．ただし，行列 $\boldsymbol{A}_s^0$ に関しては全ての環境で共通である．

次に，GMM の各コンポーネントの補正関数を全て連結することによって，特定の環境の変換関数を表すスーパーベクトル $\boldsymbol{SV}^i$ を得る．ただし，$S$ は GMM の混合数である．

$$\boldsymbol{SV}^i = \{\hat{\boldsymbol{r}_1^i}, \cdots, \hat{\boldsymbol{r}_s^i}, \cdots, \hat{\boldsymbol{r}_S^i}\} \tag{9}$$

このスーパーベクトルは，学習データの中の全ての環境に関して各々学習する．そして，得られた複数のスーパーベクトルに対して主成分分析を施す．学習データの中の全ての環境の変換関数のスーバーベクトルの平均を表すバイアスベクトル $\boldsymbol{BV}$ と，その主成分を表すベクトル $\boldsymbol{PC}^m$ を得る．ただし，$m$ は主成分のインデックスである．

$$\boldsymbol{BV} = \{\boldsymbol{b}_1, \cdots, \boldsymbol{b}_s, \cdots, \boldsymbol{b}_S\} \tag{10}$$

$$\begin{aligned}\boldsymbol{PC}^1 &= \{\boldsymbol{c}_1^1, \cdots, \boldsymbol{c}_s^1, \cdots, \boldsymbol{c}_S^1\}\\ &\vdots\\ \boldsymbol{PC}^M &= \{\boldsymbol{c}_1^M, \cdots, \boldsymbol{c}_s^M, \cdots, \boldsymbol{c}_S^M\}\end{aligned} \tag{11}$$

これらのバイアスベクトルと主成分を用いて，ある環境での変換関数は以下のように表せる．

$$\begin{aligned}\hat{\boldsymbol{x}}_t &= \sum_s p(s|\boldsymbol{y}_t) \left(\boldsymbol{A}_s^0 \boldsymbol{y}_t + \boldsymbol{B}_s \boldsymbol{w} + \boldsymbol{b}_s\right), \text{where}\\ \boldsymbol{B}_s &= \{\boldsymbol{c}_s^{1\mathrm{T}}, \cdots, \boldsymbol{c}_s^{M\mathrm{T}}\}\end{aligned} \tag{12}$$

ただし，ここで添字 $\boldsymbol{w}$ は主成分の重み付けを表す．また，添字 T は行列の転置を表す．

### 3.2 重みの推定

以上の操作によって，新たな雑音環境が現れたときに，適応すべき変換関数のパラメータ

が高次元の補正ベクトルから，低次元の重みベクトルになった．次にこの重み $\boldsymbol{w}$ の推定について述べる．この重みベクトルは，少数の未知環境下における雑音付加音声とクリーン音声のパラレルデータを用いて，最小誤差基準で下記のように推定される．

$$\hat{\boldsymbol{w}} = \underset{w}{\operatorname{argmin}} \sum_t \left\{ \boldsymbol{x}_t - \sum_s p(s|\boldsymbol{y}_t) \left( \boldsymbol{A}_s \boldsymbol{y}_t + \boldsymbol{B}_s \boldsymbol{w} + \boldsymbol{b}_s \right) \right\}^2 \tag{13}$$

これを解くと，以下のような重み付けになる．

$$\begin{aligned} \hat{\boldsymbol{w}} &= \left( \sum_t \boldsymbol{M}_t^{\mathrm{T}} \boldsymbol{M}_t \right)^{-1} \left( \sum_t \boldsymbol{M}_t^{\mathrm{T}} \boldsymbol{E}_t \right), \text{where} \\ \boldsymbol{M}_t &= \sum_s p(s|\boldsymbol{y}_t) \boldsymbol{B}_s \\ \boldsymbol{E}_t &= \boldsymbol{x}_t - \sum_s p(s|\boldsymbol{y}_t) \left( \boldsymbol{A}_s \boldsymbol{y}_t + \boldsymbol{b}_s, \right) \end{aligned} \tag{14}$$

以上の手順によって，従来の SPLICE を少数の適応データを用いて，重みを推定することで未知の雑音環境下における変換関数を得られるようになった．この手法を Eigen–SPLICE と今後記述することにする (図 1 参照)．

### 3.3 擬似パラレルデータ

通常，声質変換などの場合だと未知の入力話者と出力話者が同じ内容の発話をしたものを得ることは難しい．同様に，雑音除去においても未知の雑音環境下において雑音付加音声とクリーン音声のパラレルデータを得ることは難しい．しかし未知の雑音付加音声は，学習用のクリーン音声に未知雑音を付加することで擬似的に作成可能である．具体的には，未知の雑音付加音声が得られたときに，雑音のみの区間を取り出して，それを学習データ中のクリーン音声に付加することで，擬似パラレルデータを得る．そして，この擬似パラレルデータを用いて未知環境における主成分の重み付けを推定する．

## 4. 実　　験

以上のような提案手法の性能を評価するために，AURORA–2 データベースを用いて以下の実験を行った．

まず，AURORA–2 の train セット 4 タイプ× 4SNR 計 16 雑音環境すべてを使って，雑音環境非依存の変換関数 $\boldsymbol{A}_s^0$ を学習する．次に $\boldsymbol{A}_s^0$ を用いて，(8) 式のように 16 雑音環境

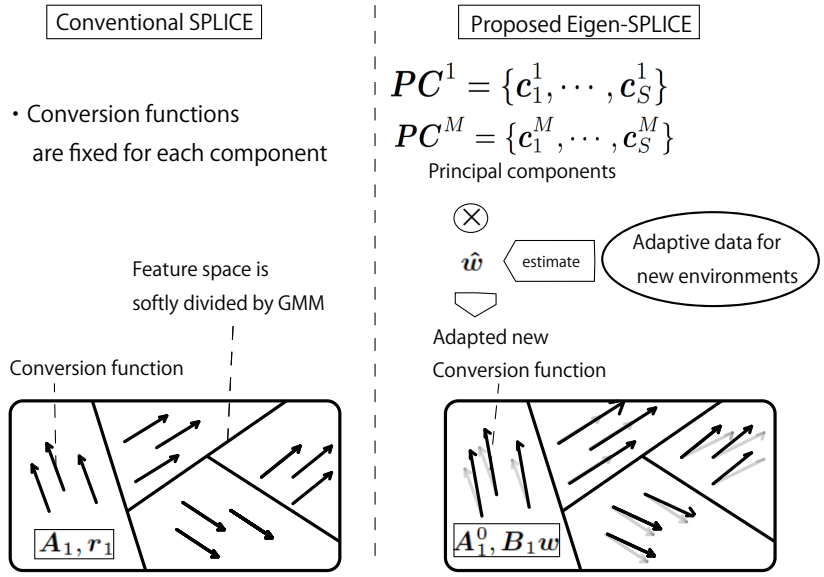


**図 1**　Overview of Eigen-SPLICE

毎の $\boldsymbol{r}_s^i$ を学習する．その後，スーパーベクトルを構成し，主成分分析を施し，主成分とバイアスベクトルを得る．

得られた主成分を用いて，AURORA–2 の testb セットに対して認識実験を行った．testb セットは train セットにはない雑音のタイプが様々な SNR で収録されている．つまり，testb セットは未知の雑音環境である．

### 4.1 実　験　1

まず，擬似パラレルデータが完璧に求められたと仮定したときの，提案手法の性能を見るために，テストセットの一部を主成分の重み学習に用いた．test セットには雑音環境毎に 1000 発声あるが，そのうち 800 を認識用，200 を重み学習用に分ける．このとき，適応に用いる発声の数を変えて認識実験を行った．また，用いる主成分の数は予備実験において高い性能を示した 6 とした．特徴量としては MFCC13 次元に Δ と Δ Δ を加えた 39 次元 (HTK での MFCC_D_A_0)，雑音付加音声を表現する GMM の混合数は 64 混合とした．認識結果は，testb セットの N1 から N4，0[dB] から 20[dB] の Accuracy の平均を用いて評価した．今回，認識用の HMM は train-set のクリーン音声のみから学習し，1 単語あたり 18

Vol.2011-SLP-87 No.15
2011/7/23

状態，1 状態あたり 20 混合の GMM を持つ単語 HMM を用いた．

雑音除去後の特徴量に対する認識実験の結果は，図２のようになった．ここでいう従来手法は，単純に SPLICE の処理を施したものである．今回は，計算量を抑えるために GMM 混合数を 64 としたため，従来手法は 81.14%と性能が低い．

重み付けを学習するパラレルデータは重み付け学習用の 200 発声のうちからランダムに選んだため，少数の場合には性能が不安定になるものの，データを増やすと安定して高い性能を発揮している．適応データを３２発声分まで用いると，平均で 12.0%の誤り削減率が見られた．環境毎の詳細な認識精度をベースラインや従来手法とともに表 1 の (a),(b),(c) に示す．

### 4.2 実　験　2

次に，理想的なパラレルデータを用いた場合ではなく，擬似パラレルデータを用いて重み付けを学習した場合の提案手法の性能を確認するために以下の実験を行った．まず，得られた未知の雑音付加音声と学習データのクリーン音声から擬似データを作成した．今回は単純に，雑音付加音声の前後 0.25[s] は雑音のみの区間だと仮定した．その雑音のみの区間を切り出し，連結し，0.50[s] の雑音のみの区間を得る．次に得られた雑音を波形領域で，クリーン音声に繰り返し重畳することで擬似的な雑音付加音声を得る．これで，擬似的なパラレルデータを得る．この擬似パラレルデータを用いて，実験 1 と同様に testb セットに対して雑音除去，認識を行い，その認識精度 (Accuracy) の平均をみた．また，そのときの条件は高い性能を発揮した主成分数 6，適応データ数を 32 発声とした．そのときの結果が表 1 の (d) である．

従来手法より 1.9%の誤り率の改善を得ることができたが，実験 1 での理想的な状態よりも性能が低下してしまっている．前後 0.25[s] が雑音区間であるという非常に大雑把な仮定をおいて，擬似雑音付加音声を作成したことが原因のひとつと考えられる．この雑音のみの区間をカルマンフィルタを用いた手法などでより厳密に推定したり，話すのを少し待ってもらうことでより長い雑音のみの区間を得るたりすることが出来れば，理想的な状態に近づき，擬似パラレルデータを用いた場合でもより高い精度を得られると考えられる．

## 5. ま　と　め

雑音付加音声の特徴量のモデルを用いた雑音抑制手法の一つである SPLICE は，定常雑音環境を暗に仮定しており，非定常雑音環境下においては十分な性能を発揮することが保証されていない．そこで本報告では未知の雑音環境に対して変換自体を適応することを試み

**表 1** Word recognition accuracy(testb-set) (a) without enhancement, baseline. (b) with the conventional enhancement. (c) with the proposed method using ideal parallel data. Principal Components: 6, Adapted data: 32 utterances. (d) with the proposed method using quasi parallel data. Principal Components: 6, Adapted data: 32 utterances.

| (a) | Rest. | Street | Airport | Station | Avg. |
|---|---|---|---|---|---|
| 20dB | 90.16 | 94.63 | 86.88 | 88.18 | 89.96 |
| 15dB | 71.31 | 84.59 | 66.39 | 69.01 | 72.83 |
| 10dB | 44.77 | 59.90 | 40.42 | 42.77 | 46.97 |
| 5dB | 10.51 | 31.08 | 11.08 | 15.82 | 17.12 |
| 0dB | -15.2 | 11.24 | -6.26 | 0.00 | -2.55 |
| Avg. | 40.31 | 56.29 | 39.70 | 43.16 | 44.86 |
| (b) | Rest. | Street | Airport | Station | Avg. |
| 20dB | 99.36 | 98.36 | 99.02 | 98.86 | 98.90 |
| 15db | 98.53 | 97.28 | 98.51 | 97.32 | 97.91 |
| 10dB | 94.96 | 90.87 | 94.78 | 93.06 | 93.42 |
| 5dB | 82.64 | 70.56 | 80.47 | 72.82 | 76.62 |
| 0dB | 49.43 | 33.68 | 43.16 | 29.25 | 38.88 |
| Avg. | 84.99 | 78.14 | 83.19 | 78.26 | 81.14 |
| (c) | Rest. | Street | Airport | Station | Avg. |
| 20dB | 99.45 | 98.16 | 99.18 | 98.76 | 98.89 |
| 15db | 98.47 | 96.96 | 98.63 | 97.40 | 97.87 |
| 10dB | 95.37 | 91.69 | 96.26 | 94.01 | 94.33 |
| 5dB | 83.81 | 74.41 | 83.51 | 76.56 | 79.57 |
| 0dB | 55.90 | 39.46 | 52.06 | 38.1 | 46.38 |
| Avg. | 86.60 | 80.13 | 85.93 | 80.97 | 83.41 |
| (d) | Rest. | Street | Airport | Station | Avg. |
| 20dB | 99.41 | 98.16 | 99.26 | 98.64 | 98.87 |
| 15dB | 98.47 | 96.99 | 98.22 | 97.36 | 97.76 |
| 10dB | 94.67 | 89.74 | 95.00 | 92.92 | 93.08 |
| 5dB | 82.05 | 67.79 | 81.29 | 72.40 | 75.88 |
| 0dB | 51.08 | 33.82 | 48.65 | 33.83 | 41.84 |
| Avg. | 85.14 | 77.30 | 84.48 | 79.03 | 81.49 |

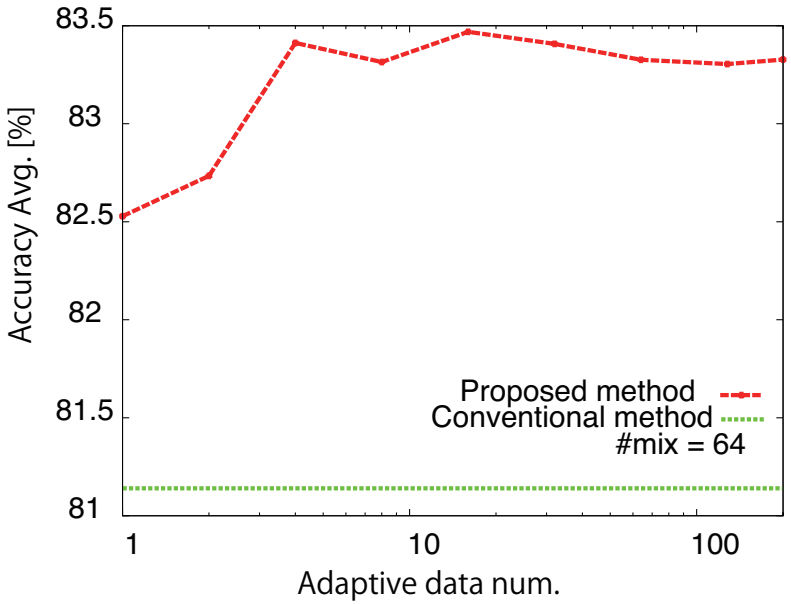


**図 2** Word recognition accuracy after proposed enhancement and after conventional method

た．変換方法を適応するためには，本来であれば各コンポーネント毎に高次元の補正ベクトルのパラメータを推定する必要があり，多くの計算量を要してしまう．それを避けるため，提案手法では変換を表すベクトルに主成分分析を施すことで推定すべきパラメータを削減した．これにより，少量の適応データでも適切な特徴量変換を推定できるようになった．

### 5.1 今後の展望

以上のように SPLICE の性能改善を実現した提案手法であるが，未だにいくつかの問題がある．まず，擬似パラレルデータを作成する際に，雑音区間を非常に雑に推定したため，理想的な場合と比べ性能が大きく低下してしまっている．より厳密な雑音区間推定をすることで，擬似データの場合でも性能を向上させることを考えている．また，今回は比較しなかった EMS との性能比較もする予定である．

## 参 考 文 献

1) 猿渡　洋：招待講演 音声信号処理における雑音抑圧技術の最新動向 (音声)，電子情報通信学会技術研究報告， Vol.110, No.401, pp.1–6 (2011).
2) Hyvarinen, A., Krhunen, J. and Oja, E.: 詳解　独立成分分析，東京電機大学出版局 (2005).
3) Boll, S.: Suppression of acoustic noise in speech using spectral subtraction, *Acoustics, Speech and Signal Processing, IEEE Transactions on*, Vol.27, No.2, pp.113–120 (1979).
4) Droppo, J., Deng, L. and Acero, A.: Evaluation of the SPLICE algorithm on the Aurora2 database, *Seventh European Conference on Speech Communication and Technology* (2001).
5) Stouten, V.: Robust automatic speech recognition in time-varying environments, *KU Leuven, Ph.D. Thesis* (2006).
6) Pearce, D., Hirsch, H. et al.: The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions, *Sixth International Conference on Spoken Language Processing*, Citeseer (2000).
7) Droppo, J., Acero, A. and Deng, L.: Efficient on-line acoustic environment estimation for FCDCN in a continuous speech recognition system, *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, Vol.1, IEEE, pp.209–212 (2001).
8) Chen, K., Liau, W., Wang, H. and Lee, L.: Fast speaker adaptation using eigenspace-based maximum likelihood linear regression, *Proc. ICSLP*, Vol.3, pp. 742–745 (2000).
9) 戸田智基，大谷大和，鹿野清宏：固有声に基づく声質変換，信学技報, SP2006-39，pp. 25–30 (2006).