雑音特徴量を用いた劣化音声特徴量変換に関する検討*

☆鈴木雅之 (東大), 吉岡拓也, 渡部晋治 (NTT CS 研), 峯松信明 (東大)

1 はじめに

近年，携帯端末での通話をはじめ，個人での動画撮影，自動音声認識の利用など，音声を実環境で収音する機会が増えている．しかし，実環境で収音された音声には，ほぼ不可避的に雑音が重畳がされるため，音声品質や音声認識性能が劣化してしまう．そのため，実環境で収音された音声を対象とするアプリケーションにとって，雑音除去技術が不可欠である．

雑音除去技術は，大きく二つのアプローチに分類できる．一つ目は，スペクトルサブトラクション [1] やウィーナーフィルタ [2] など，パワースペクトルドメインで雑音除去を行う手法である．これらは，少ない計算量の割に有効な手法であることが知られているが，雑音除去後のスペクトル歪みが大きい問題がある．

二つ目は，VTS (Vector Taylor Series) compensation [3] や SPLICE (Stereo Piecewise LInear Compensation for Environments) [4] など，MFCC など音声認識の特徴量として利用されるドメインにおいて，あらかじめ学習された音声のモデルを用いて特徴量変換を行う手法である．この手法は，ある程度計算コストがかかるものの，音声のモデルを用いるために低歪みな雑音除去が実現できる．本稿では，近年の計算機の計算能力向上を鑑みて，後者の特徴量変換のアプローチついて考察する．

従来の特徴量変換アプローチの問題点として，非定常雑音をうまく扱えないことがある．例えば，VTS compensation を用いる場合，雑音のモデルが更新される度に逆行列計算を行う必要があるため，実用的な計算コストで処理できない．また SPLICE を用いる場合には，雑音が定常であることを暗に仮定して劣化音声とクリーン音声の対応関係を学習するため，そもそも非定常雑音が考慮されていない．

そこで本稿では，SPLICE を雑音特徴量を用いるように拡張し，非定常雑音にも有効な劣化音声特徴量変換を目指す．ここで SPLICE は，劣化音声の特徴量空間を「空間分割」し，分割されたそれぞれの空間の中で「線形変換」を行う特徴量変換技術であるので，SPLICE の「空間分割」「線形変換」それぞれに対し，雑音特徴量を用いるように拡張する手法を検討する．

実験の結果，特に SPLICE の「線形変換」に雑音特徴量を用いるように拡張した手法が，雑音特徴量の推定値を一定にした場合にも SPLICE と同等の性能を実現し，かつ，雑音特徴量が正しく推定できた場合には SPLICE を超える性能を実現できることがわかった．

2 SPLICE について

本節では，劣化音声特徴量変換法として広く知られている SPLICE について説明を行う．まず SPLICE による劣化音声特徴量変換法を具体的に説明し，その後解釈と問題点について述べる．

2.1 SPLICE による劣化音声特徴量変換

SPLICE の目標は，観測された劣化音声の特徴量 $\boldsymbol{y}$ から，クリーン音声の特徴量の推定値 $\hat{\boldsymbol{x}}$ を求めることである．ここで $\boldsymbol{y}$，$\hat{\boldsymbol{x}}$ は，それぞれ列ベクトルで表された特徴量ベクトルであり，具体的には MFCC 及びその動的特徴量（Δ，$\Delta\Delta$）が用いられることが多い．

SPLICE の学習データとして，クリーン音声の特徴量系列 $\boldsymbol{X} = [\boldsymbol{x}_1 \boldsymbol{x}_2 \cdots \boldsymbol{x}_T]$，それに対応する劣化音声の特徴量系列 $\boldsymbol{Y} = [\boldsymbol{y}_1 \boldsymbol{y}_2 \cdots \boldsymbol{y}_T]$ を与える．ここで T は特徴量系列のデータ数で，ある時刻 t における $\boldsymbol{x}_t$，$\boldsymbol{y}_t$ は，時間的に同期しているものとする．このような学習用データは，クリーン音声と雑音のみを録音したものを別々に用意し，計算機上でそれらを重畳することにより，比較的容易に用意することができる．

まず準備として，$\boldsymbol{y}$ を出力する混合数 K の GMM (Gaussian Mixture Model) を，学習データとして $\boldsymbol{Y}$ を用いて学習する．これにより，GMM の k 番目のコンポーネントの重み $p(k)$，およびそのコンポーネントに対応する正規分布 $p(\boldsymbol{y}|k) = \mathcal{N}(\boldsymbol{y}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ のパラメタ $\boldsymbol{\mu}_k$，$\boldsymbol{\Sigma}_k$ が推定される．ここで $\boldsymbol{\Sigma}_k$ は対角共分散行列と仮定される．

これを用い，SPLICE では以下のように $\boldsymbol{y}$ から $\hat{\boldsymbol{x}}$ を推定する．

$$\hat{\boldsymbol{x}} = \sum_{k=1}^{K} p(k|\boldsymbol{y}) \boldsymbol{A}_k \boldsymbol{y}' \quad (1)$$

$$p(k|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|k)p(k)}{\sum_{k=1}^{K} p(\boldsymbol{y}|k)p(k)} \quad (2)$$

ただし $\boldsymbol{y}'$ は $[1\ \boldsymbol{y}^T]^T$ の拡張特徴量ベクトルである．

*A Study on Feature Enhancement Using Noise Features by M.Suzuki (Tokyo Univ.), T.Yoshioka, S.Watanabe (NTT), N.Minematsu (Tokyo Univ.)

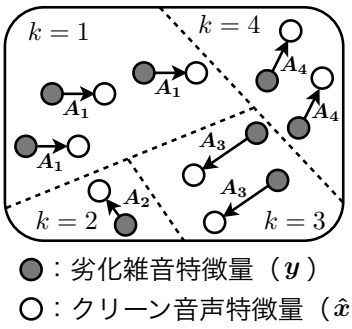

Fig. 1　Transformation from noisy speech features $\boldsymbol{y}$ to estimated clean speech features $\hat{\boldsymbol{x}}$ using SPLICE

また $\boldsymbol{A}_k$ は $N \times (N+1)$ の行列であり，例えば以下ように ML 推定する（重み付き二乗誤差最小化）.

$$\boldsymbol{A}_k^{\mathrm{ML}} = \underset{\boldsymbol{A}_k}{\operatorname{argmin}} \sum_{t=1}^{T} p(k|\boldsymbol{y}_t)|\boldsymbol{x}_t - \boldsymbol{A}_k \boldsymbol{y}_t'|^2 \quad (3)$$

なお，この解析解は，

$$\boldsymbol{A}_k^{\mathrm{ML}} = \boldsymbol{X'PY'}^T(\boldsymbol{Y'PY'}^T)^{-1} \quad (4)$$

と計算できる [5]．ただし，$\boldsymbol{X'}$ および $\boldsymbol{Y'}$ はそれぞれ $\boldsymbol{X}$，$\boldsymbol{Y}$ を拡張特徴量ベクトル系列にしたものであり，$\boldsymbol{P}$ は $p(k|\boldsymbol{y}_1), p(k|\boldsymbol{y}_2), \cdots p(k|\boldsymbol{y}_T)$ を対角成分にもつ対角行列である.

なお SPLICE には，式 (1) の線形変換 $\boldsymbol{A}_k\boldsymbol{y'}$ の代わりに単純なシフト $\boldsymbol{y}+\boldsymbol{b}_k$ を用い $\boldsymbol{b}_k$ を学習する方法や，$\boldsymbol{A}_k$（もしくは $\boldsymbol{b}_k$）を識別学習する方法などもあるが，今回は先に述べた $\boldsymbol{A}_k$ を ML 推定する SPLICE を利用する.

2.2　SPLICE の解釈

式 (1) を見返してみると，$p(k|\boldsymbol{y})$ を重みとした，線形変換 $\boldsymbol{A}_k\boldsymbol{y'}$ の重み付け和により，$\hat{\boldsymbol{x}}$ が求められていることがわかる．ここで $p(k|\boldsymbol{y})$ は，GMM のあるコンポーネント k に $\boldsymbol{y}$ が属する確率である．また，それぞれのコンポーネント k ごとに，対応する $\boldsymbol{A}_k$ が用意されている.

仮に，$p(k|\boldsymbol{y})$ が最大となるような k のときに $p(k|\boldsymbol{y}) = 1$ とし，それ以外の k で $p(k|\boldsymbol{y}) = 0$ と近似すると，SPLICE の動作は Fig. 1 のように図示できる．すなわち，まず入力の劣化音声特徴量 $\boldsymbol{y}$ がどのコンポーネント k に属するかを求め，対応するコンポーネント用に学習してある $\boldsymbol{A}_k$ により，$\boldsymbol{y}$ を $\hat{\boldsymbol{x}}$ に特徴量変換していることになる.

簡単にまとめると，SPLICE は，GMM で「空間分割」し，分割されたそれぞれの空間で「線形変換」を行う劣化音声特徴量変換法といえる．なお，このように「空間分割」と「線形変換」を利用して特徴量変換を行う手法は，雑音除去以外にも，例えば GMM に基づく声質変換などに利用されている [6].

2.3　SPLICE の問題点

SPLICE の大きな特色の一つに，雑音の特徴量 $\boldsymbol{n}$ の情報を陽に用いていないことがある．ただし $\boldsymbol{n}$ は，$\boldsymbol{y}$ および $\hat{\boldsymbol{x}}$ と同じく列ベクトルで表された特徴量ベクトルである．ここで，真のクリーン音声の特徴量ベクトルを $\boldsymbol{x}$ として，$\boldsymbol{x}$，$\boldsymbol{y}$，$\boldsymbol{n}$ には $\boldsymbol{x} = f(\boldsymbol{y}, \boldsymbol{n})$ といった関係が成立する．そのため，$\boldsymbol{n}$ の情報は，$\hat{\boldsymbol{x}}$ を推定する上で非常に有効なものである．実際，他の多くの雑音除去手法では，$\boldsymbol{n}$ に関するパラメタを推定し，雑音除去に用いている．ところが SPLICE では，$\hat{\boldsymbol{x}}$ を計算する過程において，$\boldsymbol{y}$ のみが用いられており，$\boldsymbol{n}$ に関する情報は陽に使われていない.

SPLICE が $\boldsymbol{n}$ を陽に用いずに特徴量変換を行うことは，利点であり欠点でもある．利点は，$\boldsymbol{n}$ に関するパラメタを推定する計算コストがかからないことである．しかしこれは，近年の計算機の性能向上や $\boldsymbol{n}$ の推定技術の発展を鑑みれば，あまり問題にならない．欠点は，非定常雑音に弱いことである．$\hat{\boldsymbol{x}}$ を計算するのに $\boldsymbol{y}$ しか用いないのは，$\boldsymbol{n}$ が変化しない，すなわち雑音の定常性を仮定していることに他ならない．しかし実際の環境では雑音は変化しうるため，この仮定は適切でない.

3　雑音特徴量を用いた SPLICE

ここまで，SPLICE による劣化音声特徴量変換法について述べ，その問題点を指摘した．本節では，その非定常雑音に弱いという問題点に対処するため，SPLICE において雑音特徴量 $\boldsymbol{n}$ を陽に用いる手法を提案する．これまで見てきたとおり，SPLICE は「空間分割」と「線形変換」による特徴量変換なので,「空間分割」に $\boldsymbol{n}$ を陽に用いる方法,「線形変換」に $\boldsymbol{n}$ を陽に用いる方法について，それぞれ検討する．このように SPLICE に $\boldsymbol{n}$ を陽に用いることで，$\boldsymbol{n}$ の変化の情報が特徴量変換に反映され，非定常雑音にも強い雑音抑圧ができると予想される.

以降，雑音特徴量を用いた SPLICE の学習データとして，クリーン音声の特徴量系列 $\boldsymbol{X} = [\boldsymbol{x}_1\boldsymbol{x}_2\cdots\boldsymbol{x}_T]$，劣化音声の特徴量系列 $\boldsymbol{Y} = [\boldsymbol{y}_1\boldsymbol{y}_2\cdots\boldsymbol{y}_T]$，雑音の特徴量系列 $\boldsymbol{N} = [\boldsymbol{n}_1\boldsymbol{n}_2\cdots\boldsymbol{n}_T]$ を与える．ある時刻 t における $\boldsymbol{x}_t$，$\boldsymbol{y}_t$，$\boldsymbol{n}_t$ は，時間的に同期しているものとする.

3.1　空間分割に雑音特徴量を用いる SPLICE

式 (1) の $p(k|\boldsymbol{y})$ の代わりに，$p(k|\boldsymbol{y}, \boldsymbol{n})$ を用いれば，空間分割に雑音の特徴量 $\boldsymbol{n}$ を用いることができる．しかし，$\boldsymbol{n}$ は観測されるものではなく，劣化音声の特徴量 $\boldsymbol{y}$ から推定するため，完璧には推定できない．そこで，$p(\boldsymbol{n})$ が正規分布から出力されるとモデル化し，

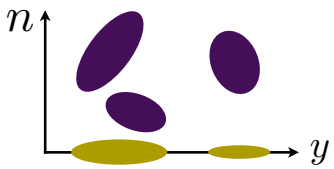

Fig. 2　Effect to use $\boldsymbol{n}$ for space division of SPLICE

$p(\boldsymbol{n}) = \mathcal{N}(\boldsymbol{n}; \boldsymbol{\mu_n}, \boldsymbol{\Sigma_n})$ の $\boldsymbol{\mu_n}$，$\boldsymbol{\Sigma_n}$ の推定値 $\hat{\boldsymbol{\mu}}_{\boldsymbol{n}}$，$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}$ をパラメタとして用いる．

まず準備として，$\boldsymbol{y}$ と $\boldsymbol{n}$ の結合ベクトル $[\boldsymbol{y}^T \boldsymbol{n}^T]^T$ を出力する混合数 K の GMM を，$\boldsymbol{Y}$ と $\boldsymbol{N}$ を結合した $[\boldsymbol{Y}^T \boldsymbol{N}^T]^T$ を学習データとして学習する．これにより，GMM の k 番目のコンポーネントの重み $p(k)$，およびそのコンポーネントに対応する正規分布 $p([\boldsymbol{y}^T \boldsymbol{n}^T]^T|k) = \mathcal{N}([\boldsymbol{y}^T \boldsymbol{n}^T]^T; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ のパラメタ $\boldsymbol{\mu}_k$，$\boldsymbol{\Sigma}_k$ が推定される．

これを用いて，以下のように $\hat{\boldsymbol{x}}$ を計算する．

$$\hat{\boldsymbol{x}} = \sum_{k=1}^{K} p(k|\boldsymbol{y}, \hat{\boldsymbol{\mu}}_{\boldsymbol{n}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}) \boldsymbol{A_k y'} \quad (5)$$

$$p(k|\boldsymbol{y}, \hat{\boldsymbol{\mu}}_{\boldsymbol{n}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}) = \frac{p(\boldsymbol{y}, \hat{\boldsymbol{\mu}}_{\boldsymbol{n}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}|k)p(k)}{\sum_{k=1}^{K} p(\boldsymbol{y}, \hat{\boldsymbol{\mu}}_{\boldsymbol{n}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}|k)p(k)} \quad (6)$$

ただし，$p(\boldsymbol{y}, \hat{\boldsymbol{\mu}}_{\boldsymbol{n}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}|k)$ は，$\boldsymbol{\mu}_k$ の $\boldsymbol{y}$ に対応する部分を $\boldsymbol{\mu}_{k(\boldsymbol{y})}$，$\boldsymbol{n}$ に対応する部分を $\boldsymbol{\mu}_{k(\boldsymbol{n})}$ とおき，さらに $\boldsymbol{\Sigma}_k$ は対角共分散行列と仮定し，$\boldsymbol{y}$ に対応する部分を $\boldsymbol{\Sigma}_{k(\boldsymbol{y})}$，$\boldsymbol{n}$ に対応する部分を $\boldsymbol{\Sigma}_{k(\boldsymbol{n})}$ として，

$$\begin{aligned} & p(\boldsymbol{y}, \hat{\boldsymbol{\mu}}_{\boldsymbol{n}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}|k) \\ &= \int_{\boldsymbol{m}} p([\boldsymbol{y}^T \boldsymbol{m}^T]^T|k) \mathcal{N}(\boldsymbol{m}; \hat{\boldsymbol{\mu}}_{\boldsymbol{n}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}) d\boldsymbol{m} \\ &= \mathcal{N}(\boldsymbol{y}; \boldsymbol{\mu}_{k(\boldsymbol{y})}, \boldsymbol{\Sigma}_{k(\boldsymbol{y})}) \mathcal{N}(\hat{\boldsymbol{\mu}}_{\boldsymbol{n}}; \boldsymbol{\mu}_{k(\boldsymbol{n})}, \boldsymbol{\Sigma}_{k(\boldsymbol{n})} + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{n}}) \end{aligned} \quad (7)$$

と計算する．

Fig. 2 に，空間分割に雑音特徴量を陽に用いる効果を概念的に示す．$\boldsymbol{n}$ を用いない場合は，黄色のように分割されていたものが，$\boldsymbol{n}$ を用いることで，紫色のように分割されることになる．この図でいうと，特に左側のクラスタについて，$\boldsymbol{n}$ を用いることで、黄色の一つのクラスタが、新たに紫色の 2 つのクラスタに分かれており，空間分割が $\boldsymbol{n}$ に依存してより精緻にできることになる．すなわち，$\boldsymbol{n}$ のパラメタが正確に推定できるという仮定の下，提案手法は SPLICE よりも $\boldsymbol{n}$ の変化の情報を特徴量変換に反映できると期待される．

3.2　線形変換に雑音特徴量を用いる SPLICE

式 (1) の $\boldsymbol{A_k y'}$ の代わりに，$\boldsymbol{A_k}[\boldsymbol{y'}^T \hat{\boldsymbol{\mu}}_{\boldsymbol{n}}^T]^T$ を用いれば，線形変換に $\boldsymbol{n}$ を用いることができる．

$$\hat{\boldsymbol{x}} = \sum_{k=1}^{K} p(k|\boldsymbol{y}) \boldsymbol{A_k} [\boldsymbol{y'}^T \hat{\boldsymbol{\mu}}_{\boldsymbol{n}}^T]^T \quad (8)$$

$$\hat{x}_k = \boxed{A_k}\ y'$$

$$\hat{x}_k = \boxed{A_k} \begin{pmatrix} y' \\ n \end{pmatrix}$$

Fig. 3　Effect to use $\boldsymbol{n}$ for linear transformation of SPLICE

ここでは，$\boldsymbol{A_k}$ は $N \times (2N+1)$ の行列となる．このような $\boldsymbol{A_k}$ は，通常の SPLICE で $\boldsymbol{A_k}$ を学習する際，$\boldsymbol{Y}$ を $[\boldsymbol{Y}^T \boldsymbol{N}^T]^T$ に置き換えるだけで，まったく同じ方法で学習することができる．

Fig. 3 に，線形変換に雑音特徴量を陽に用いる効果について示す．通常の SPLICE と異なり，$\hat{\boldsymbol{\mu}}_{\boldsymbol{n}}$ も $\hat{\boldsymbol{x}}$ の推定に用いられている．そのため，$\hat{\boldsymbol{\mu}}_{\boldsymbol{n}}$ 推定精度が高く，次元呪いの問題のない範囲では，提案手法は SPLICE よりも $\boldsymbol{n}$ の変化の情報を特徴量変換に反映できると期待される．

3.3　雑音の一部の情報を用いる SPLICE

ここまで，雑音の特徴量 $\boldsymbol{n}$ を，$\boldsymbol{y}$ や $\hat{\boldsymbol{x}}$ と同じく，N 次元の特徴量として議論してきた．ここで，$\boldsymbol{n}$ を別の特徴量としても，同様の議論を行うことができる．例えば，$\boldsymbol{n}$ として雑音のパワー n_0 のみを用いることもできるし，雑音パワーと第一次項 $[n_0 n_1]^T$ の 2 次元を用いることもできる．このように雑音特徴量の次元を減らすことで，最終的に推定すべきパラメタの数を調整することが可能になる．雑音のパワーは最も容易に推定でき，かつ最も雑音除去性能に影響をもつと考えられるので，以降本稿では，雑音のパワー n_0 のみを雑音特徴量として利用する SPLICE について検討する．

4　実験

提案した雑音特徴量を用いる SPLICE に関して，

- 雑音のパラメタの推定精度が低い場合，通常の SPLICE から精度が下がらないかどうか
- 雑音のパラメタが正確に推定できた場合，通常の SPLICE から精度がどれだけ上がるか

の二つについて評価を行う．前者の評価は，提案手法の最低性能の評価で，これが SPLICE と同程度以上であれば，提案手法でも安定した雑音除去が実現できるものと考えられる．後者の評価は，提案手法の最高性能の評価で，実際に雑音のパラメタを逐次推定しながら雑音除去する場合には，最高性能と最低性能の間の性能が得られると考えられる．

提案手法として，空間分割に雑音パワーを用いる／用いない，線形変換に雑音パワーを用いる／用い

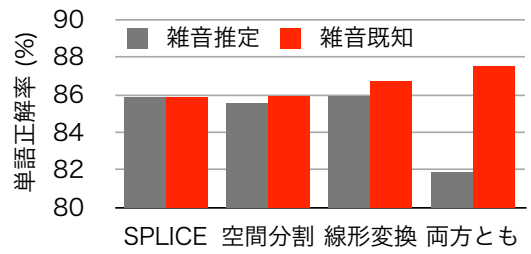

Fig. 4　The average of word accuracy

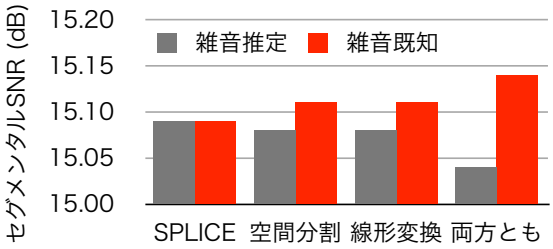

Fig. 5　The average of segmental SNR

ないの計 4 手法を試す．なお空間分割にも線形変換にも雑音パワーを用いないものが，通常の SPLICE である．これらの提案手法それぞれに対し，雑音パワーを，音声ファイルの先頭無音区間から推定しそれを常に用いる場合（最低性能の評価）と，雑音パワーの真値を与える場合（最高性能の評価）を実験する．

実験の具体的な条件を述べる．性能評価には，音声認識の単語正解率とセグメンタル SNR を用いた．データベースには AURORA-2 データベースを用い，評価にはセット A を用いた．音声認識に用いる HMM の各種パラメタは，すべて AURORA-2 の complex backend に準拠した形で実験を行った．$\boldsymbol{y}$ や $\hat{\boldsymbol{x}}$ の特徴量には，MFCC とそのパワー（ケプストラムの 0 次元），およびその Δ，ΔΔ の合計 $(12+1)\times 3 = 39$ 次元を用いた．また $\boldsymbol{n}$ の特徴量には，前述の通りパワーのみ 1 次元を用いた．特徴量に関するその他パラメタ（窓の種類やシフト長など）は，すべて AURORA-2 の標準で用いられているものを採用した．空間分割に用いる GMM の混合数 K は，1024 とした．雑音パワーを音声ファイルの先頭から推定する際には，先頭 20 フレームの平均・分散により推定を行った．また，雑音除去時に雑音を推定しながら提案手法を用いる場合，学習データとして用いる $\boldsymbol{N}$ の代わりに，$\boldsymbol{Y}$ から推定された $\hat{\boldsymbol{N}}$ を用いた方が，学習と雑音除去のミスマッチがなくなると予想される．しかし，予備実験の結果，ほとんど性能は変化しなかったため，学習データには常に雑音特徴量の真値 $\boldsymbol{N}$ を用いることにした．

単語正解率の平均を Fig. 4 に示す．まず，雑音パワーを先頭 20 フレームから推定して用いた場合の結果を見ると，空間分割のみ，もしくは線形変換のみに雑音パワーを用いた場合には，SPLICE とほぼ同等の結果が得られた．しかし，空間分割にも線形変換にも雑音パワーを用いた場合には，SPLICE より精度が低下してしまっている．これは，雑音パワーの変化の情報が特徴量変換に強く反映されすぎてしまい，雑音パワーの推定精度の低さが精度に大きく影響してしまったためだと考えられる．次に，雑音パワーの真値を与えて推定した場合の結果を見ると，通常の SPLICE，空間分割のみ，線形変換のみ，両方とも，の順で精度が高くなっている．以上の結果より，線形変換のみに雑音パワーを用いて SPLICE を行うことで，頑健性を保ちつつ，通常の SPLICE より高い精度を得ることができると分かる．

セグメンタル SNR の平均を Fig. 5 に示す．セグメンタル SNR については，単語正解率と同様の傾向が見られるものの，改善幅は非常に限定的であった．主な理由は，特徴量はスペクトル包絡に対応するため，特徴量強調ではスペクトルの微細構造は回復されないためと考えられる．このことから，SPLICE ならびに提案法は，雑音除去そのものよりも，音声認識の前処理としての用途に適しているといえる．

5　まとめ

本稿では，非定常雑音に弱いという SPLICE の欠点を改善するために，SPLICE において雑音特徴量を用いる方法をいくつか提案し，実験的に比較検討した．実験の結果，特に SPLICE の線形変換部分に雑音パワーを用いることで，頑健性を保ちつつ，より高い精度が得られることが分かった．

雑音特徴量を用いた SPLICE の拡張によって一定の性能改善が得られたが，この枠組みについては更なる検討の余地が残されている．特に，本稿では対角行分散行列を用いたが，雑音特徴量と劣化音声特徴量との間には本来強い相関があるため，全共分散行列を用いたほうがよいと予想される．これらについては今後の課題である．

参考文献

[1] S.F. Boll, IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. 27, No. 2, pp. 113–120 (1979).
[2] J.S. Lim, A.V. Oppenheim, IEEE Trans. on Acoustics, Speech, Signal Processing, Vol. 37, No. 11, pp. 1659–1671 (1989).
[3] J.C. Segura, A. De la Torre, M.C. Beneitez, A.M. Penado, Proc. Eurospeech (2001).
[4] J. Droppo, L. Deng, A. Acero, Proc. Eurospeech (2001).
[5] Y. Qiao, N. Minematsu, Proc. ICASSP (2009).
[6] H. Zen, Y. Nankaku, T.Keiichi, IEEE Trans. on Acoustics, Speech, and Language Processing, Vol. 19, No. 2, pp. 417–430 (2011).