2007－SLP－67 (12)
2007／7／21

GMMに基づく最尤変換法による携帯電話音声の帯域拡張

藤敦 渉† 関本 英彦† 戸田 智基† 猿渡 洋† 鹿野 清宏†

† 奈良先端科学技術大学院大学 情報科学研究科 〒630–0192 奈良県生駒市高山町 8916–5
E-mail: †{wataru-f,hidehi-s,tomoki,sawatari,shikano}@is.naist.jp

あらまし 帯域拡張は狭帯域音声のみから広帯域音声を再構築する有効な技術である．従来の典型的な手法として，混合正規分布 (GMM: Gaussian Mixture Model) を用いた最小二乗誤差推定 (MMSE: Minimum Mean Square Error) に基づく帯域拡張が提案されている．MMSE 基準の手法は比較的高い変換精度を実現できるものの，1) フレーム間の相関を考慮していないため，時間方向に不適切な特徴量遷移が生じる事がある，2) 汎下処理により推定された広帯域スペクトル包絡が過剰に平滑化されてしまう，といった問題点が残されている．これらの問題点を解決するために，動的特徴量と系列内変動を考慮した最尤推定を GMM に基づく帯域拡張に導入をする．実験的評価により，提案手法の有効性を示す．

Bandwidth Extension of Cellular Phone Speech based on Maximum Lilelihood Estimation with GMM

Wataru FUJITSURU†, Hidehiko SEKIMOTO†, Tomoki TODA†, Hiroshi SARUWATARI†, and Kiyohiro SHIKANO†

† Graduate School of Information Science, Nara Institute of Science and Technology
Takayama-cho 8916–5, Ikoma-shi, Nara, 630–0192 Japan
E-mail: †{wataru-f,hidehi-s,tomoki,sawatari,shikano}@is.naist.jp

Abstract Bandwidth extension is a useful technique for reconstructing wideband speech from only narrowband speech. As a typical conventional method, bandwidth extension algorithm based on minimum mean square error (MMSE) with GMM has been proposed. Although the MMSE-based method has reasonably high conversion-accuracy, there still remain some problems to be solved: 1) inappropriate spectral movements are caused by ignoring a correlation between frames, and 2) the converted spectra are excessively smoothed by the statistical modeling. In order to address those problems, we propose a bandwidth extension algorithm based on maximum likelihood estimation (MLE) considering dynamic features and the global variance (GV) with a Gaussian Mixture Model (GMM). Results experimental evaluations of a subjective test demonstrate that the proposed algorithm outperforms the conventional MMSE-based one.

1. はじめに

近年，爆発的に普及した携帯電話によって，音声コミュニケーションをより頻繁かつ容易に行う事が可能となった．一般的に用いられている携帯電話音声は 3.4 kHz 以下の狭帯域音声であり，コミュニケーションを可能とする明瞭性は保たれているものの，その音質は十分とは言い難い．特に 3.4 kHz 以上に重要なエネルギーが分布する摩擦音や破裂音のようないくつかの音素で，音質の劣化が著しく目立つ．そこで，より高品質な音声コミュニケーションを実現するために，広帯域音声に対応した符号化方式 [1] が盛んに研究されている．その音質改善効果は絶大である一方で，基本的に広帯域音声に対応した符号化方式は，狭帯域音声の符号化方式より多くの情報量が必要となる．災害などによる回線混雑時を考えると，情報量を増加させずに，高品質な広帯域音声コミュニケーションを実現する技術の構築が望まれる．

情報量を増加させずに広帯域音声コミュニケーションを実現する手法として帯域拡張法が研究されている．帯域拡張は狭帯域音声のみから広帯域音声を再構築する技術であり，これまでに数々の統計的スペクトル変換に基づく帯域拡張法が提案されている．コードブックマッピング法に基づく帯域拡張 [2] では，個々のフレームにお

いて狭帯域スペクトル特徴量をベクトル量子化し，狭帯域スペクトル特徴量セントロイドに対応する広帯域スペクトル特徴量セントロイドへと置き換えることで広帯域音声を再構築する．ハードクラスタリングに基づくアプローチであり，大きな量子化誤差が生じやすい．量子化誤差を大幅に低減する手法の1つとして，ソフトクラスタリングと連続的なマッピングを可能にする混合正規分布モデル (Gaussian Mixture Model :GMM) に基づく帯域拡張法 [3] が提案されている．基本的なマッピングアルゴリズムとして，声質変換 [4] において提案されたものが適用されている．従来の GMM に基づく手法の多くは最小二乗誤差基準で変換が行われる [3], [5]．比較的高い変換精度を実現できる一方で，1) フレーム間の相関を考慮していないため，時間方向に不適切な特徴量遷移が生じる事がある，2) 汎下処理により推定された広帯域スペクトル包絡が過剰に平滑化されてしまう，といった問題が残されている．これらの問題点を緩和する手法として，スペクトル系列の動的特性をモデリングするアプローチ [6], [7] やマッピングと符号化処理を併用するアプローチ [8], [9] が研究されている．

近年，動的特徴量と系列内変動を考慮した最尤推定によって，GMM に基づく声質変換性能が著しく向上した [10]．この技術は声質変換に限らず，帯域拡張においても性能改善をもたらすものと期待させる．本稿では動的特徴量と系列内変動を考慮した最尤推定に基づく帯域拡張法を提案する．実験評価によって，動的特徴量と系列内変動の有効性を示す．

以下，2. で従来法である最小二乗誤差推定に基づく帯域拡張法について述べ，3. で最尤推定に基づく帯域拡張法について述べる．4. で帯域拡張処理について述べ，5. で評価実験について述べる．最後に 6. において本稿の結論と今後の課題について述べる．

2. 従来法：最小二乗誤差推定に基づく帯域拡張 [3]

2.1 学 習 [11]

フレーム t において D_x 次元の狭帯域特徴量を $\boldsymbol{x}_t$，D_y 次元の広帯域特徴量を $\boldsymbol{y}_t$ とする．狭帯域と広帯域の特徴量ベクトルの結合確率密度は，次式に表す GMM によってモデル化される．

$$P\left(\boldsymbol{z}_t|\theta\right) = \sum_{m=1}^{M} \omega_m \mathcal{N}(\boldsymbol{z}_t; \boldsymbol{\mu}_m^{(z)}, \Sigma_m^{(z)}) \tag{1}$$

ここで，$\boldsymbol{z}_t$ は結合ベクトル $\left[\boldsymbol{x}_t^\top, \boldsymbol{y}_t^\top\right]^\top$ である．$^\top$ は転置を表す．混合数は M であり，m 番目の混合重みは ω_m である．$\mathcal{N}(\cdot;\boldsymbol{\mu},\boldsymbol{\Sigma})$ は $\boldsymbol{\mu}$，$\boldsymbol{\Sigma}$ の正規分布を表す．θ は重み，平均ベクトル，共分散行列からなるモデルパラメータである．m 番目の分布における平均ベクトル $\boldsymbol{\mu}_m^{(z)}$ と共分散行列 $\boldsymbol{\Sigma}_m^{(z)}$ は次式で表される．

$$\boldsymbol{\mu}_m^{(z)} = \begin{bmatrix} \boldsymbol{\mu}_m^{(x)} \\ \boldsymbol{\mu}_m^{(y)} \end{bmatrix}, \boldsymbol{\Sigma}_m^{(z)} = \begin{bmatrix} \boldsymbol{\Sigma}_m^{(xx)} & \boldsymbol{\Sigma}_m^{(xy)} \\ \boldsymbol{\Sigma}_m^{(yx)} & \boldsymbol{\Sigma}_m^{(yy)} \end{bmatrix} \tag{2}$$

ここで，$\boldsymbol{\mu}_m^{(x)}$ と $\boldsymbol{\mu}_m^{(y)}$ は m 番目の分布における狭帯域特徴量，広帯域特徴量の平均ベクトル，$\boldsymbol{\Sigma}_m^{(xx)}$ と $\boldsymbol{\Sigma}_m^{(yy)}$ は狭帯域特徴量，広帯域特徴量の共分散行列である．$\boldsymbol{\Sigma}_m^{(yx)}$ は m 番目の分布における狭帯域特徴量と広帯域特徴量の相互共分散行列である．GMM は学習データセットの結合ベクトルを用いて EM アルゴリズムで学習される．

2.2 最小二乗誤差推定に基づく変換 [12]

次式に表す最小二乗誤差推定に基づき，入力された狭帯域特徴量に対応する広帯域特徴量を推定する．

$$\begin{aligned} \hat{\boldsymbol{y}}_t &= E\left[\boldsymbol{y}_t|\boldsymbol{x}_t\right] \\ &= \sum_{m=1}^{M} P(m|\boldsymbol{x}_t, \theta)\boldsymbol{E}_{m,t} \end{aligned} \tag{3}$$

ここで

$$P(m|\boldsymbol{x}_t, \theta) = \frac{\omega_m \mathcal{N}(\boldsymbol{x}_t; \boldsymbol{\mu}_m^{(x)}, \boldsymbol{\Sigma}_m^{(xx)})}{\sum_{j=1}^{M} \omega_j \mathcal{N}(\boldsymbol{x}_t; \boldsymbol{\mu}_j^{(x)}, \boldsymbol{\Sigma}_j^{(xx)})} \tag{4}$$

$$\boldsymbol{E}_{m,t} = \boldsymbol{\mu}_m^{(y)} + \boldsymbol{\Sigma}_m^{(yx)}(\boldsymbol{\Sigma}_m^{(xx)})^{-1}(\boldsymbol{x}_t - \boldsymbol{\mu}_m^{(x)}) \tag{5}$$

である．

2.3 問 題 点

最小二乗誤差推定は比較的高い変換精度を実現できるものの，1) フレーム間の相関を無視しているため，不適切なスペクトル遷移が生じる，2) 汎下処理により推定されたスペクトルが過剰に平滑化される，といった問題がある．

3. 提案法：最尤推定に基づく帯域拡張

従来手法の二つの問題を解決するために，動的特徴量と系列内変動を考慮した最尤推定に基づく帯域拡張を提案する．動的特徴量の導入により，フレーム間の相関を考慮に入れた広帯域特徴量系列の推定が可能となる．さらに系列内変動の導入により，推定スペクトルの過剰な平滑化を緩和することが可能となる．

3.1 学 習

フレーム t において，静的・動的特徴量からなる $2D_x$ 次元の狭帯域音声特徴量を $\boldsymbol{X}_t = \left[\boldsymbol{x}_t^\top, \Delta\boldsymbol{x}_t^\top\right]^\top$，$2D_y$ 次元の広帯域音声特徴量を $\boldsymbol{Y}_t = \left[\boldsymbol{y}_t^\top, \Delta\boldsymbol{y}_t^\top\right]^\top$ とする．学習データとしてフレーム毎に対応付けられた特徴量ベクトルを用いて，次式に表す結合確率密度をモデル化する GMM を学習する．

$$P\left(\boldsymbol{Z}_t|\Theta\right) = \sum_{m=1}^{M} \omega_m \mathcal{N}(\boldsymbol{Z}_t; \boldsymbol{\mu}_m^{(Z)}, \Sigma_m^{(Z)}) \tag{6}$$

ここで，$\boldsymbol{Z}_t$ は結合ベクトル $\left[\boldsymbol{X}_t^\top, \boldsymbol{Y}_t^\top\right]^\top$ である．Θ は，混合重み，平均ベクトル，共分散行列からなるモデルパラメータである．

3.2 動的特徴量を考慮した最尤推定 [10]

狭帯域・広帯域特徴量系列ベクトルを各々 $\boldsymbol{X} = \left[\boldsymbol{X}_1^\top, \boldsymbol{X}_2^\top, \cdots, \boldsymbol{X}_T^\top\right]^\top$，$\boldsymbol{Y} = \left[\boldsymbol{Y}_1^\top, \boldsymbol{Y}_2^\top, \cdots, \boldsymbol{Y}_T^\top\right]^\top$ とする．入力された狭帯域特徴量ベクトル系列に対応する広

帯域の静的特徴量ベクトル系列 $\hat{\boldsymbol{y}} = \left[\hat{\boldsymbol{y}}_1^\top, \cdots, \hat{\boldsymbol{y}}_t^\top\right]^\top$ は次式に示す尤度最大化に基づき推定される．

$$\begin{aligned}\hat{\boldsymbol{y}} &= \arg\max P(\boldsymbol{Y}|\boldsymbol{X}, \Theta) \\ &\quad \text{subject to } \boldsymbol{Y} = \boldsymbol{W}\boldsymbol{y}\end{aligned} \tag{7}$$

ここで，$\boldsymbol{W}$ は静的特徴量系列 $\boldsymbol{y}$ を静的・動的特徴量系列に拡張する変換行列である．

本稿では計算量を削減するために，次式に表す近似を用いる．

$$P(\boldsymbol{Y}|\boldsymbol{X}, \Theta) \simeq P(\boldsymbol{m}|\boldsymbol{X}, \Theta) P(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{m}, \Theta) \tag{8}$$

ここで，$\boldsymbol{m}$ は単一分布系列 $[m_1, m_2, \cdots, m_t]$ である．次式に示すように，最尤分布系列 $\hat{\boldsymbol{m}}$ を決定した後で，近似された尤度関数を最大にする広帯域の静的特徴量ベクトル系列 $\hat{\boldsymbol{y}}$ を決定する．

$$\hat{\boldsymbol{m}} = \arg\max P(\boldsymbol{m}|\boldsymbol{X}, \Theta) \tag{9}$$

$$\begin{aligned}\hat{\boldsymbol{y}} &= \arg\max P(\hat{\boldsymbol{m}}|\boldsymbol{X}, \Theta) P(\boldsymbol{Y}|\boldsymbol{X}, \hat{\boldsymbol{m}}, \Theta) \\ &= \left(\boldsymbol{W}^\top \boldsymbol{D}_{\hat{\boldsymbol{m}}}^{(Y)^{-1}} \boldsymbol{W}\right)^{-1} \boldsymbol{W}^\top \boldsymbol{D}_{\hat{\boldsymbol{m}}}^{(Y)^{-1}} \boldsymbol{E}_{\hat{\boldsymbol{m}}}^{(Y)}\end{aligned} \tag{10}$$

ここで

$$\boldsymbol{E}_{\hat{\boldsymbol{m}}}^{(Y)} = \left[\boldsymbol{E}_{\hat{m}_1,1}^{(Y)}, \cdots, \boldsymbol{E}_{\hat{m}_t,t}^{(Y)}, \cdots, \boldsymbol{E}_{\hat{m}_T,T}^{(Y)}\right] \tag{11}$$

$$\boldsymbol{D}_{\hat{\boldsymbol{m}}}^{(Y)^{-1}} = \mathrm{diag}\left[\boldsymbol{D}_{\hat{m}_1}^{(Y)^{-1}}, \cdots, \boldsymbol{D}_{\hat{m}_t}^{(Y)^{-1}}, \cdots, \boldsymbol{D}_{\hat{m}_T}^{(Y)^{-1}}\right] \tag{12}$$

であり，

$$\boldsymbol{E}_{\hat{m}_t,t} = \boldsymbol{\mu}_{\hat{m}_t}^{(Y)} + \boldsymbol{\Sigma}_{\hat{m}_t}^{(YX)} \left(\boldsymbol{\Sigma}_{\hat{m}_t}^{(XX)}\right)^{-1} (\boldsymbol{X}_t - \boldsymbol{\mu}_{\hat{m}_t}^{(X)}) \tag{13}$$

$$\boldsymbol{D}_{\hat{m}_t} = \boldsymbol{\Sigma}_{\hat{m}_t}^{(YY)} - \boldsymbol{\Sigma}_{\hat{m}_t}^{(YX)} \left(\boldsymbol{\Sigma}_{\hat{m}_t}^{(XX)}\right)^{-1} \boldsymbol{\Sigma}_{\hat{m}_t}^{(XY)} \tag{14}$$

である．なお，計算量をさらに削減するため，本稿では $\boldsymbol{D}_{\hat{m}_t}$ の対角成分のみを考慮する．

3.3 系列内変動を考慮した最尤推定 [10]

系列内変動 (GV: Global Variance) とは系列内における静的特徴量の分散を表す．

$$\boldsymbol{v}(\boldsymbol{y}) = \left[v^{(1)}, v^{(2)}, \cdots, v^{(D)}\right]^\top \tag{15}$$

$$v^{(d)} = \frac{1}{T} \sum_{t=1}^{T} \left\{ (y_t(d) - \frac{1}{T} \sum_{\tau=1}^{T} y_\tau(d))^2 \right\} \tag{16}$$

ここで，$y_t^{(d)}$ はフレーム t における広帯域静的特徴量の d 次元目の成分である．本稿では発話単位で GV を計算する．

広帯域静的動的特徴量ベクトル系列と広帯域静的特徴量の GV に対する確率からなる次式の尤度関数を最大にする広帯域静的特徴量 $\boldsymbol{y}$ を推定する．

$$\mathcal{L} = \alpha \log P(\boldsymbol{Y}|\boldsymbol{X}, \hat{\boldsymbol{m}}, \Theta) + \log P(\boldsymbol{v}(\boldsymbol{y})|\Theta_v) \tag{17}$$

ここで，$P(\boldsymbol{v}(\boldsymbol{y})|\Theta_v)$ は正規分布によってモデル化される．モデルパラメータ Θ_v は GV $\boldsymbol{v}(\boldsymbol{y})$ に関する平均ベクトル $\boldsymbol{u}^{(v)}$ と共分散行列 $\boldsymbol{\Sigma}^{(vv)}$ からなり，あらかじめ学習データから推定される．定数 α は二つの尤度間の重みを調整するパラメータを表し，本稿では特徴量ベクトルの次元数比である $\frac{1}{27}$ に定める．広帯域静的特徴量 $\boldsymbol{y}$ に関して対数尤度 L を最大化するために，次式に表す一次の導関数を用いた最急降下法を用いる．

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \boldsymbol{y}} &= \alpha \left(-\boldsymbol{W}^\top \boldsymbol{D}_{\hat{\boldsymbol{m}}}^{(Y)^{-1}} \boldsymbol{W}\boldsymbol{y} + \boldsymbol{W}^\top \boldsymbol{D}_{\hat{\boldsymbol{m}}}^{(Y)^{-1}} \boldsymbol{E}_{\hat{\boldsymbol{m}}}^{(Y)}\right) \\ &\quad + \left[\boldsymbol{v}_1'^\top, \boldsymbol{v}_2'^\top, \cdots, \boldsymbol{v}_t'^\top, \cdots, \boldsymbol{v}_T'^\top\right]^\top\end{aligned} \tag{18}$$

$$\boldsymbol{v}_t' = [\boldsymbol{v}_t'(1), \boldsymbol{v}_t'(2), \cdots, \boldsymbol{v}_t'(d), \cdots, \boldsymbol{v}_T'(D)]^\top \tag{19}$$

$$\boldsymbol{v}_t'(d) = -\frac{2}{T} \boldsymbol{p}_v^{(d)^\top} (\boldsymbol{v}(\boldsymbol{y}) - \boldsymbol{\mu}_v) (y_t(d) - \bar{y}(d)) \tag{20}$$

$$\bar{y}(d) = \frac{1}{T} \sum_{\tau=1}^{T} y_\tau(d) \tag{21}$$

ここで，ベクトル $\boldsymbol{p}_v^{(d)}$ は共分散行列 $\boldsymbol{\Sigma}^{(vv)}$ の逆行列における d 番目の列ベクトルを表す．

4. 帯域拡張処理

帯域拡張の処理フローを Fig. 1 に示す．狭帯域特徴量からすべての帯域における広帯域特徴量の推定が行われるが，低域信号は入力から得られるため，必ずしも推定された信号を使わなくてもよい．本稿では，帯域拡張音声の低域信号には入力された狭帯域音声を用いる．

Step 1 狭帯域音声信号から，特徴量として F_0，メルケプストラム (mcep)，非周期成分 (ap) [12] を抽出する．

Step 2 mcep 変換用 GMM および ap 変換用 GMM を用いて，狭帯域音声の mcep および ap を広帯域音声の特徴量へ変換をする．

Step 3 変換はスペクトル包絡だけであり，パワーは狭帯域音声信号のものを用いる．抽出された F_0 と変換非周期成分を用いて，STRAIGHT 混合励振源 [12] を生成し，変換 mcep に基づき MLSA フィルタリング [13] を施し，推定広帯域音声を合成する．

Step 4 推定広帯域音声に low pass filter (LPF) と high pass filter (HPF) を施すことで低域信号と高域信号に分ける。

Step 5 入力狭帯域音声に対して up sampling を施し，低域信号を作成する．

Step 6 推定広帯域音声の低域信号と入力低域信号のパワーが等しくなるように，推定広帯域音声のパワーを補正する．

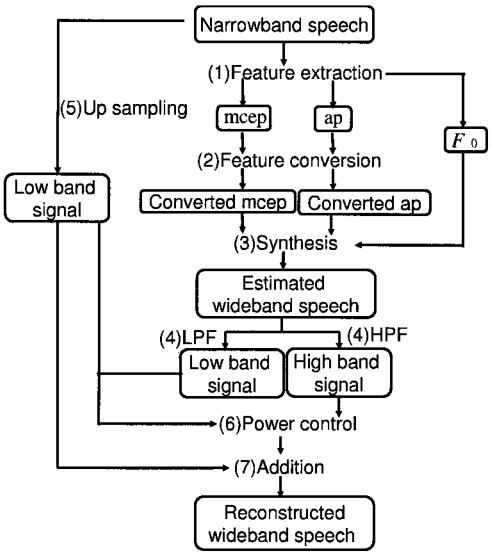

図 1 帯域拡張システム．mcep はメルケプストラム，ap は非周期成分を示す．
Fig. 1 Bandwidth extension system. "mcep" denotes the mel-cepstral and "ap" denotes the aperiodic component.

Step 7 パワー補正が行われた高域信号と低域信号を加算することで帯域拡張音声を得る．

5. 実験による評価

提案法の有効性を検証するために，従来の最小二乗誤差推定に基づく帯域拡張 (MMSE) と，動的特徴量を考慮した最尤推定に基づく帯域拡張 (ML) 及び 動的特徴量と系列内変動を考慮した最尤推定に基づく帯域拡張 (MLGV) を比較する．

5.1 実験条件

広帯域音声として，16 kHz でサンプリングされた日本人話者 4 名の自然音声を用いる．広帯域音声を 8 kHz へダウンサンプリングし，さらに狭帯域音声コーデックである EVRC (Enhanced Variable Rate Codec [14]，5.2 の実験にて使用) および AMR-NB (Adaptive Multi Rate-NarrowBand [15]，5.3 の実験にて使用) を施した音声を狭帯域音声として用いた．学習データとして ATR 音素バランス文サブセット A の 50 文を使用し，評価データとして ATR 音素バランス文サブセット B の 50 文を使用する．

スペクトル特徴量として，狭帯域の場合は，メルケプストラム分析 [13] により得られる 1〜16 次の mcep 係数を用い，広帯域の場合は，STRAIGHT 分析により得られる 1〜24 次の mcep 係数を用いる．また，STRAIGHT 分析 [12] により抽出された各周波数における非周期成分 (ap) に対して，狭帯域の場合は 3 つの帯域 (0〜1，1〜2，2〜4 kHz) で平均したもの，広帯域の場合は 5 つの帯域 (0〜1，1〜2，2〜4, 4〜6, 6〜8 kHz) で平均したものを用いる．

表 1 客観評価実験結果: メルケプストラムひずみ [dB]
Table 1 Result of objective evalustion: Mel cepstral distortion [dB]

	MMY	MYI	FKN	FYM	average
MMSE [dB]	5.37	5.55	5.67	5.54	5.59
ML [dB]	5.25	5.51	5.61	5.49	5.47
MLGV [dB]	5.26	5.48	5.66	5.51	5.48
target [dB]	4.64	4.81	5.36	5.38	5.06

フレームシフトは 5 ms である．mcep 変換用 GMM の混合数は 64，ap 変換用 GMM の混合数 4 とする．本稿では特定話者モデルを評価する．

5.2 従来手法との比較

5.2.1 客観評価

スペクトルひずみ尺度を用いて，従来手法と提案手法における帯域拡張音声の比較評価を行う．スペクトルひずみ尺度として以下のメルケプストラムひずみを用いる．

$$\text{メルケプストラムひずみ [dB]} = \frac{10}{\ln 10}\sqrt{2\sum_{d=1}^{39}(mc_d^{(X)} - mc_d^{(Y)})} \tag{22}$$

ここで，$mc_d^{(X)}$ は帯域拡張音声の d 次元のメルケプストラム係数，$mc_d^{(Y)}$ は広帯域自然音声の d 次元のメルケプストラム係数を表す．

表 1 は各手法における評価話者 4 名 (MMY, MYI, FKN, FYM) のメルケプストラムひずみを表 1 に示す．次元数は 39 とする．なお, target は広帯域自然音声の低域信号を EVRC 出力に置換したものを表し，帯域拡張音声の理想値と考えられる．ML および MLGV のメルケプストラムひずみは MMSE より小さくなっている．このことから，静的特徴量のみでなく動的特徴量も考慮することで，より適切な遷移を持つ広帯域特徴量系列を実現でき，スペクトル変換精度が向上すると言える。

Fig. 2 に狭帯域音声，帯域拡張音声，広帯域自然音声のスペクトル系列の一例を示す．提案手法によって，狭帯域音声のみから広帯域自然音声の高域信号を復元できているのが見て分かる．

5.2.2 主観評価

提案手法と従来手法の帯域拡張音声の音質を 5 段階オピニオンスコア (5 : 非常に良い, 4:良い, 3:普通, 2: 悪い, 1:非常に悪い) を用いて評価する．評価音声は，EVRC, MMSE，ML, MLGV, 広帯域自然音声 (Natural) の 5 つである．被験者は日本人成人男女 8 名である．各被験者あたりの評価文数は 120 文である．

主観評価実験の結果を Fig. 3 に示す．有意水準 5 % において，EVRC と従来手法 MMSE の間に有意差は見られない．一方，提案手法 ML は EVRC と MMSE の両方に対して有意な音質の向上が見られる．さらに GV を考慮することで，最も高い音質の帯域拡張音声が得られる．以上のことから，動的特徴量と GV の両方を考慮することで，帯域拡張の性能を大幅に改善できることが分かる．

5.3 広帯域音声符号化方式との比較

提案手法と広帯域音声に対応した符号化方式を比較

するため，5段階オピニオンスコアを用いて音質を評価する．評価音声は，AMR-NB（Adaptive Multi Rate-NarrowBand [15]）のビットレート6.7 kbpsから帯域拡張した音声（MLGV），AMR-WB（Adaptive Multi Rate-WidebandBand [1]）のビットレート6.6 kbps及び8.85 kbpsの音声である．また，広帯域音声のMNRU（Modulated Noise Reference Unit）信号（SNR 25[dB], SNR 28[dB], SNR 31[dB], SNR 34[dB]）も評価する．被験者は日本人成人男女5名である．各被験者あたりの評価文数は140文である．

主観評価実験の結果をFig. 4に示す．提案手法の音質は等価Q値で30 dB程度に相当する．有意水準5 %において，帯域拡張音声は同程度の情報量の広帯域符号化音声よりも音質が高いことが分かる．一方でその音質は8.85 kbpsの多い情報量の音声には及ばない傾向が見られる．以上のことから，帯域拡張は極めて有効な技術といえる．

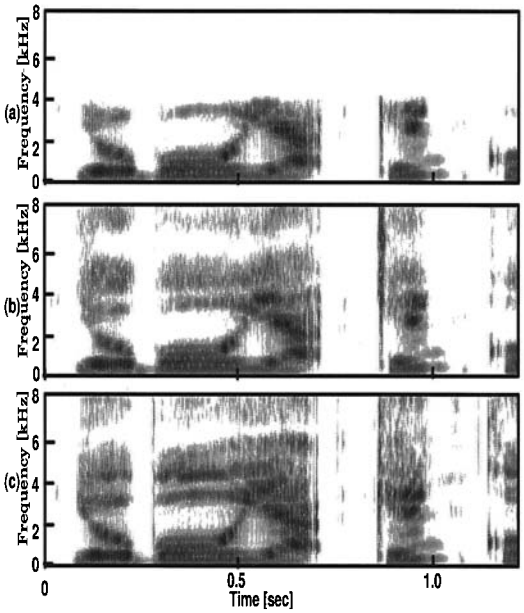

図2　音声スペクトルグラムの一例．(a) 狭帯域音声，(b) 動的特徴量，系列内変動を考慮した帯域拡張音声，(c) 自然音声．（発話内容は「予防や健康」）．

Fig. 2　An example of spectra of narrowband speech, (a) converted speech by the ML using the GV , (b) spectra of natural speech, (c) for a sentence fragment, "/ y o b o: y a k e N k o:/".

6. おわりに

動的特徴量および系列内変動を考慮したGMMに基づく最尤推定による帯域拡張を提案した．提案法の有効性を検証するために，客観評価実験および主観評価実験を行った．その結果，従来手法である最小二乗誤差推定に基づく帯域拡張と比較して大幅に改善されることが分かった．提案法により狭帯域音声の音質が大幅に改善され，また，その音質は同程度の情報量の広帯域音声符号化方式の音声よりも優れていることが分かった．これらの結果から，提案法の有効性が示された．

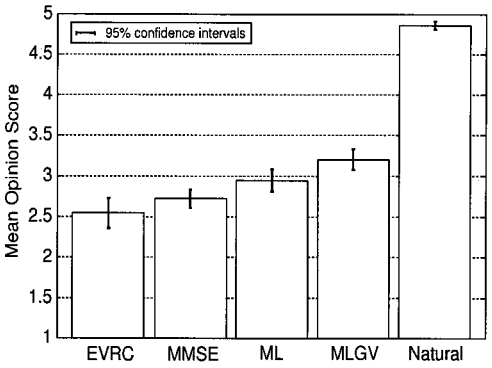

図3　主観評価実験の結果

Fig. 3　Result of subjective evaluation.

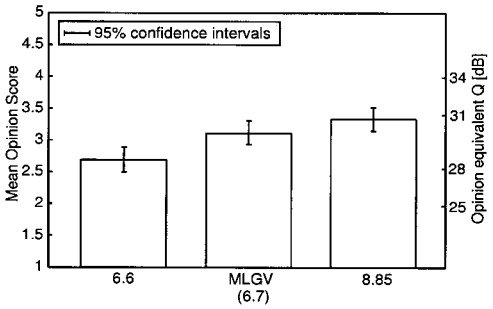

図4　広帯域音声符号化方式との比較

Fig. 4　Result of comparison between our proposed wideband extension and wideband speech codec.

今後の課題として，学習データ外の未知話者への対応や対応した対策，実時間で帯域拡張を可能にする変換アルゴリズムの開発，実環境の雑音に頑健な取り組みといった研究などが挙げられる．

謝辞

本研究の一部は，文部科学省リーディングプロジェクトe-Soceity とKDDI研究所との共同研究により実施したものである．

文　献

[1] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Potola-Pukkila, J. Vainio, H. Mikkola and K.Jarvinen, "The adaptive multirate wideband speech codec (AMR-WB)" *Proc. IEEE*, Vol. 10, No. 8, pp. 620–636, 2002.

[2] Y. Yoshida, and M. Abe, "An Algorithm to Reconstruct Wideband Speech From Narrowband Speech Based on Codebook Mappeing", *Proc. ICSLP94*, pp. 1591–1593,1994.

[3] K.Y. Park and H.S. Kim. "Narrowband to wideband conversion of speech using GMM based transformation", *Proc. ICSLP*, pp. 1847–1850, Istanbul, June, 2000.

[4] Y. Stylianou, O. Cappe, E. Moulines, "Coutinuous proba-

bilistic transform for voice conversion", *IEEE Trans*, Speech and Audion Processing, Vol. 6, No. 2, pp. 131–142, 1998.

[5] M.L. Seltzer, A. Acero, and J. Droppo, "Robust Bandwidth Extension of Noise-corrupted Narrowband Speech" *Proc. ICSLP*, pp. 1509–1512, 2005.

[6] S. Yao and C.F. Chan, "Block-based Bandwidth Extension of Narrowband Speech Signal by using CDHMM", *Proc. ICASSP*, pp. I793–I796, 2005.

[7] S.Y Yao and C.F. Chan, "Speech bandwidth enhancement using state space speech dynamics" , *Proc. ICASSP2006*, pp. I489–I492, 2006.

[8] Y. Agiomyrgiannakis and Y. Stylianou, "Combined Estimation/coding of Highband Spectral Envelopes for Speech Spectrum Expansion" *Proc. ICASSP2004*, pp. 469–472, 2004.

[9] V. Berisha and A. Spanias , "A Scalable Bandwidth Extension Algorithm", *Proc. ICASSP2007*, pp. 601–604, 2007.

[10] T. Toda, A.W. Black, and K. Tokuda, " Spectral conversion based on maximum likelihood estimation considering global variance of converted parameter", *Proc. ICASSP2005*, pp. 9–12, 2005.

[11] A. Kain and M.W. Macon. "Spectral voice conversion for text-to-speech synthesis", *Proc. ICASSP*, Seattle, U.S.A,. pp. 285–288, May 2004.

[12] H. Kawahara, Jo Estill and O. Fujimura, "Aperiodicity extraction and control using mixed mode excitation and group delay manipulation for a high quality speech analysis, modification and synthesis system STRAIGHT", MAVEBA 2001, Sep. 13–15, Firentze Italy, 2001.

[13] 徳田 恵一, 小林 隆夫, 深田 俊明, 今井 聖. メルケプストラムをパラメータをする音声のスペクトル推定. 信学論 (A), Vol. J74-A, No.8, pp.1240–1248, 1991.

[14] T.V. Ramabadran, J.P. Ashley and M.J. McLaughlin, "Background Noise Suppression for Speech Enhancement and Coding", *IEEE Workshop on Speech Coding and Tel.* , Pocono Manor, PA, pp.43–44, 1997.

[15] 3GPP TS 26.077, "Minimum Performance Requirements for Noise Suppresser Application to the AMR Speech Encoder," Mar. 2001.