分散音声認識における実時間周波数特性正規化手法

柘植 覚[†] 黒岩 眞吾[†] 獅々堀 正幹[†]
任 福継[†] 北 研二[††]

本論文では，分散音声認識（DSR: Distributed Speech Recognition）における入力系の周波数特性の差異による認識性能劣化を抑制する周波数特性正規化手法として，複数参照ケプストラムを用いた実時間周波数特性正規化手法を提案する．提案手法は，複数の参照ケプストラムを使用し，周波数特性の正規化を行うバイアスをフレーム同期で計算し，実時間で入力系の周波数特性を正規化する手法である．一般に，DSR で用いられるクライアントではメモリ量，計算量の制限があるため，提案手法ではこれらの増加量を低減させるため，参照ケプストラムを DSR フロントエンドの特徴パラメータ圧縮部で使用される VQ コードブックの組合せで表現した．ETSI Advanced DSR フロントエンドを用いた日本音響学会新聞記事読み上げ音声コーパスの音声認識実験より，提案手法は，ETSI Advanced DSR フロントエンドにおける Blind Equalization と比較し，周波数特性の差異による音声認識精度劣化の抑制に有効であることを確認した．特に，提案手法は MIRS フィルタ条件下で ETSI Advanced DSR フロントエンド（Blind Equalization）の単語誤り率を 10.8%削減することが可能であった．

Real-time Frequency Characteristic Normalization for Distributed Speech Recognition

Satoru Tsuge,[†] Shingo Kuroiwa,[†] Masami Shishibori,[†]
Fuji Ren[†] and Kenji Kita[††]

In this paper, we propose a real-time blind equalization method with multiple references for ETSI standard Distributed Speech Recognition (DSR) front-end. The proposed method compensates for acoustic mismatch caused by input devices. In ETSI advanced DSR front-end, the blind equalization method is introduced to compensate for acoustic mismatch. This method estimates the bias, which compensates for the mismatch, using one reference vector. If the input speech is short or contains many similar phonemes, there is concern that this method might not estimate the accurate bias. On the other hand, the proposed method estimates the bias, which is calculated on frame by frame, using multiple references instead of one reference. Using multiple references, the proposed method estimates the bias more accurately. In addition, we represent the references by combining the VQ centroids used in the data compression process of ETSI standard DSR front-end. This limits increases in memory size and computation costs on the front-end. Experimental results on a Japanese newspaper dictation task indicate that the proposed method gave better performance under acoustic mismatched conditions than the conventional blind equalization method. Especially, we observed a 10.8% improvement in the error rate under the MIRS filter condition.

1. はじめに

携帯電話や PDA（Personal Digital Assistants）などの携帯端末の発達にともない，モバイル環境の普及が急速に進んでいる．一般にこれらの携帯端末は非常に小型であるため，付属デバイスによる複雑なコマンド入力は困難である．この問題を解消する一手法として，音声インタフェースが熱望されている．現在までに音声入力で，登録電話番号に電話をかける機能（単語認識）や電話番号を入力する機能（数字認識）などが携帯電話上で実現されている．しかし，携帯端末のハードウェア（CPU，メモリなど）面の制約により，電子メール文入力などの中・大語彙連続音声認識を端末内で実現することは困難である．そこで，携帯端末を用いた中・大語彙音声認識方法として，通常の電話回線を介し音声をセンタに伝送し，センタで音声認識を行う方式（以下，コーデック方式と表記）が実用化

† 徳島大学工学部
Faculty of Engineering, The University of Tokushima
†† 徳島大学高度情報化基盤センター
Center for Advanced Information Technology, The University of Tokushima

されている．しかし，音声圧縮・復元するコーデックや回線の影響により十分な認識精度が得られないという問題がある[1]．

近年，コーデックによる認識性能劣化の問題を解決する手段として，分散音声認識手法（DSR: Distributed Speech Recognition）が提案された[1),2)]．DSR では，音声認識システムがクライアント（端末）とサーバに分散して設置され，クライアントの処理量を軽減させている．クライアントでは音響分析のみを行い，分析された特徴パラメータをサーバに圧縮し伝送し，サーバ側で音声認識の探索処理を行う．このように DSR 方式では，伝送するデータが圧縮された特徴パラメータであるため伝送速度を低減できる．さらに，伝送路の周波数帯域の制限を受けず音響分析が可能なため，従来より低域，高域の周波数成分を用いることができ，認識精度を向上できる可能性もある．

DSR 方式を広く普及させるためには，音響分析を行うクライアント部と音声認識を行うサーバ部で，圧縮・復元方式，ビットストリーム形式などの共通化が必要である．そのため，欧州電気通信標準化機構（ETSI: the European Telecommunications Standards Institute）は，そのフロントエンドの標準化を進めている．標準化の一環として，ETSI は 2000 年 4 月に雑音に頑健なフロントエンドの研究開発を目的とした『標準 DSR フロントエンド（ETSI ES201）』[3)]，2002 年 10 月には雑音に頑健な『Advanced DSR フロントエンド（ETSI ES202）』を勧告した[4)]．

DSR システムに使用されるターミナルは多種多様であり，個々のターミナルで使用される入力デバイスの周波数特性には差異があることが予想される．このような周波数特性の差異は特徴パラメータの変動となり，上述の DSR フロントエンドで採用されているベクトル量子化（VQ: Vector Quantization）を用いた特徴パラメータ圧縮部において量子化歪みの増加を引き起こす原因となる．この歪みの増加は音声認識性能を劣化させる要因の 1 つとなる．従来より，周波数特性の差異を正規化し，認識性能劣化を抑制する有効な手法としてケプストラム平均減算法（CMS: Cepstral Mean Subtraction）が提案されている[5)]．しかし，ETSI DSR フロントエンドでは，CMS を適用しない特徴パラメータで作成された VQ コードブックを使用しているため☆，特徴パラメータ圧縮前に CMS を適用した場合，量子化歪みを増加させ認識性能を劣化させる可能性がある．

我々はサーバ側での CMS では解決ができない入力系の周波数特性の差異によって生じる量子化歪みに着目し，フロントエンド部でその歪みを減少させる手法に関し研究を行い，入力系の周波数特性を正規化する手法として平均一致化手法を提案した[6)]．この手法はフロントエンド部において，1 発声の特徴パラメータの平均と特徴パラメータ圧縮部で使用される VQ コードブックのセントロイドの平均を一致させ，入力系の周波数特性を正規化する手法である．音声認識実験結果より，提案手法は入力系に周波数特性の差異が生じた場合に起こる認識精度低下を抑制することを示した．しかし，平均一致化手法は 1 発声全体の入力特徴パラメータの平均を計算するため，音声入力から認識結果を出力するまでに処理遅れが生じ，実時間処理ができないという問題がある．

一方，ETSI から勧告されたフロントエンドにおいても，入力系の周波数特性の差異による認識性能の低下の問題は重要視されており，Advanced DSR フロントエンドにおいては，入力系の周波数特性を正規化する手法として Blind Equalization 手法が導入された[7)]．この手法は入力特徴パラメータを 1 つの参照ケプストラムを用い正規化するバイアスをフレーム同期で計算し，正規化する手法である．この手法は，1 つの参照ケプストラムを用い周波数特性を正規化するため，適切な周波数特性の正規化がされず，周波数特性の差異が認識精度の低下を引き起こす場合があると考えられる．

一方，複数の参照ケプストラム（コードワード）を用い，より正確に入力系の周波数特性を正規化する手法として，コードワード依存ケプストラム正規化手法（CDCN: Codeword-Depend Cepstral Normalization）が提案されている[8)]．しかし，この手法はコードワードの選択に最尤推定手法を使用しているため，実時間処理ができない．また，複数のコードワードをフロントエンド部で保持するためにはメモリ量の問題もある．

そこで，本論文では分散音声認識のための入力系の周波数正規化手法として，複数参照ケプストラムを用いた Blind Equalization 手法（BEMR: Blind Equalization technique with Multiple References）を提案する．この手法は，複数参照ケプストラムを用い周波数特性を正規化するバイアスをフレーム同期で計算し，実時間で入力系の周波数特性を正規化する手法である．この手法は，CDCN に類似する手法であるが，実時

☆ 通常 CMS による周波数特性の正規化には 1 発声全体が必要となるため，実時間処理ができない．そのため，ETSI は CMS を適用していない特徴パラメータで VQ コードブックを作成していると予想される．

間処理が可能であるという利点を持つ．さらに，DSR システムでの実時間動作を可能にするため，複数参照ケプストラムを特徴パラメータ圧縮部で使用する VQ コードブックのセントロイドの組合せで表現し，フロントエンド部におけるメモリ量の増加を低減させている．また，参照ケプストラムを選択する際に，特徴パラメータ圧縮部の距離を使用することにより，フロントエンド部における計算量を削減している．本論文では，日本音響学会新聞読み上げコーパスを用い，入力デバイスの周波数特性の変動をシミュレーションした実験を行い，提案手法の有効性を示す．

以下，2 章では従来手法として，平均一致化手法と Advanced DSR フロントエンドで採用された Blind Equalization を紹介する．3 章では，平均一致化手法のバイアス計算を複数の参照ケプストラムを用いるように変更した提案手法を述べる．4 章では，提案手法の評価実験について述べ，実験に関する考察を 5 章で述べる．最後に 6 章で本論文のまとめを述べる．

2. 従 来 手 法

本章では，本論文で提案する複数参照ケプストラムを用いた Blind Equalization 手法の基盤となる平均一致化手法，Advanced DSR フロントエンドで採用された Blind Equalization を紹介する．

ETSI DSR フロントエンドでは特徴パラメータの圧縮に VQ を用いている．入力系の周波数特性の差異は VQ 歪みを増加させ認識性能を劣化させる要因となる．そこで，周波数特性の差異に対しても VQ 歪みを増加させない手法として平均一致化手法を提案した[6]．また，Advanced DSR フロントエンドでは，同様に周波数特性の差異を正規する手法として Blind Equalization 手法が採用された．

2.1 平均一致化手法

平均一致化手法は，認識を行う 1 発声全体の特徴パラメータの平均と VQ コードブック作成データの特徴パラメータの平均を一致させるように認識発声の特徴パラメータを平行移動する手法である．平行移動にはバイアスを減算する方法で行う．

以下に，本手法の手順を示す．

I 前処理：VQ コードブック作成データの平均特徴パラメータの計算

$$\boldsymbol{a}_{train} = \frac{\sum_{s=1}^{S}\sum_{t=1}^{T_s}\boldsymbol{x}_{st}}{\sum_{s=1}^{S}T_s} \quad (1)$$

ここで，$\boldsymbol{a}_{train}$ は VQ コードブック作成データの平均特徴パラメータを示し，$\boldsymbol{x}_{st}$ は発話 s に対する各分析フレームの特徴パラメータを示す．また，S，T_s は VQ コードブック作成データ数，発声 s の総分析フレーム数を示す．

II 認識発声の平均特徴パラメータの計算

$$\boldsymbol{a}_{test} = \frac{\sum_{t=1}^{T}\boldsymbol{x}_t}{T} \quad (2)$$

ここで，$\boldsymbol{a}_{test}$ は各認識発声の平均特徴パラメータを示し，$\boldsymbol{x}_t$ は各分析フレームの特徴パラメータを示す．また，T は認識発声の総分析フレーム数を示す．

III 減算するバイアスの計算

減算するバイアスとして，特徴パラメータの平均と学習データの平均の差を計算する．

$$\boldsymbol{h} = \boldsymbol{a}_{test} - \boldsymbol{a}_{train} \quad (3)$$

IV 周波数特性の正規化

VQ コードブック作成データの平均特徴パラメータと認識発声の平均特徴パラメータの差を以下の式で減算することにより，周波数特性の正規化を行う．

$$\tilde{\boldsymbol{x}}_t = \boldsymbol{x}_t - \boldsymbol{h} \quad (4)$$

ここで，$\tilde{\boldsymbol{x}}_t$ は本手法を適用した後の特徴パラメータを示す．適用した特徴パラメータを特徴パラメータ圧縮部の入力とすることで VQ コードブックとの歪みを減少することが可能である．

実際には，ETSI DSR フロントエンドでは VQ コードブックのみが与えられるため，VQ コードブック作成データの平均特徴パラメータを算出することは困難である．そのため，ETSI で定義されている VQ セントロイドの平均を $\boldsymbol{a}_{train}$ として用いた．

2.2 Blind Equalization

ETSI は周波数特性の差異を正規化するため，Advanced DSR フロントエンドでは実時間で入力系の周波数特性を正規化する Blind Equalization 手法を採用した[4]．この Blind Equalization は，式 (5) から式 (8) を用い周波数特性の正規化を行う．これは，参照ケプストラム（$RC(i)$）に入力ケプストラム（$c(i)$）をフレームごとに近づけるようにすることにより，特徴パラメータ（$c_{eq}(i)$）を正規化している．

$$W = Min(1, Max(0, lnE - 211/64)) \quad (5)$$

$$sSize = 0.0087890625 * W \quad (6)$$

$$c_{eq}(i) = c(i) - b(i) \quad (7)$$

$$b(i) += sSize * (c_{eq}(i) - RC(i)) \quad (8)$$

ここで，E はパワー，i はケプストラムの次元を示す．また，$b(i)$ は正規化に用いるバイアスを示す．参照ケプストラムは文献 4) で勧告されている．

3. 複数参照ケプストラムを用いた Blind Equalization 手法

前章で説明した平均一致化手法では，式 (3) に示したように周波数特性を正規化するバイアスの計算に認識発声全体の平均特徴パラメータが必要になる．そのため，発声終了後でないと手法の適用ができず実時間処理ができない．そこで，本章ではバイアスの計算を実時間で行う手法として複数参照ケプストラムを用いた Blind Equalization 手法を提案する．提案手法は複数の参照ケプストラムを選択しバイアス計算に用い，そのバイアスにより周波数特性の正規化を行う．提案手法では，複数の参照ケプストラムは ETSI DSR フロントエンドの特徴パラメータ圧縮部で使用される VQ コードブックの組合せにより表現することにより，フロントエンドにおける参照ケプストラム保持のためのメモリ量の増加を抑制している．さらに，参照ケプストラム選択時の距離計算量の増加も低減可能である．提案手法は

- 参照ケプストラム作成部（3.1 節）
- 提案手法の適用部（3.2 節）

で構成されており，以下で詳しく説明する．

3.1 参照ケプストラム作成部

周波数特性の正規化を行うバイアス計算の際に使用する参照ケプストラムは以下の手順で作成する．

Step 1 初期参照ケプストラムの作成

初期参照ケプストラムとして，音響モデル学習データに対し LBG アルゴリズムにより N 個のセントロイドを求めた☆．

この初期参照ケプストラム（セントロイド）を

$$\boldsymbol{R}(n) = \{r_{n,1}, r_{n,2}, \ldots, r_{n,M}\}^T, \quad 1 \leq n \leq N \tag{9}$$

とする．ここで，N，M は参照ケプストラム数，特徴パラメータの次元数（本論文では ETSI DSR フロントエンドで分析された特徴パラメータ（MFCC）を用いるため，12 次元である）を示す．

Step 2 参照ケプストラムの量子化

参照ケプストラムの表現方法を**図 1** に示し，以下で説明する．Step 1 で作成した参照ケプストラムの各次元を 32 bit で表現した場合，フロントエンドに 32×12（次元数）$\times N$（参照ケプストラム数）（bit）のメモリ容量が必要となる．

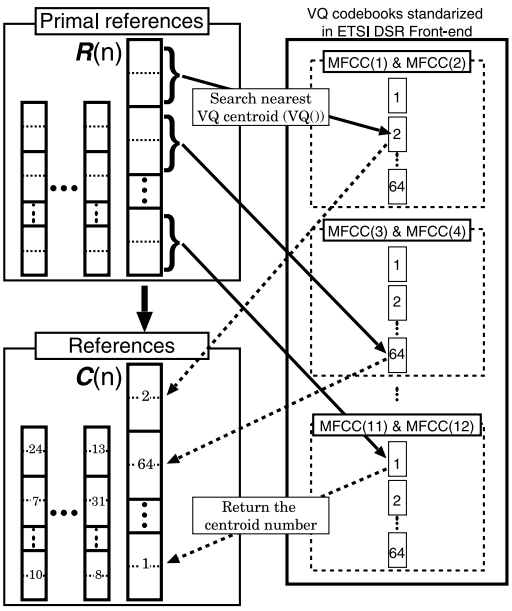

図 1 参照ケプストラムの表現方法
Fig. 1 Represent a reference by combining of VQ centroids.

一般にフロントエンドはメモリ容量が少ないため，必要なメモリ容量を削減する必要がある．そこで，提案手法では参照ケプストラムを DSR フロントエンドのパラメータ圧縮部で使用する VQ コードブックを用いて表現（量子化）する．これにより，参照ケプストラムは VQ コードブックのコード番号のみで表現でき，メモリ量の増加を抑制できる．

ETSI DSR フロントエンドでは特徴パラメータ圧縮部に分割 VQ が採用されており，特徴パラメータ 2 次元ごとに 6 bit の VQ コードブックを持っている☆☆（特徴パラメータの次元は 12 次元なので合計 6 つのコードブックを持っている）．参照ケプストラム（$\boldsymbol{C}(n)$）は，Step 1 で作成した初期参照ケプストラムとこのコードブックにおける最近傍のセントロイドを計算しそのセントロイド番号（コードワード）を返す関数（$\mathrm{VQ}_{i,j}()$）を用い

$$\begin{aligned}&\boldsymbol{C}(n)\\ &= \{\mathrm{VQ}_{1,2}(r_{n,1}, r_{n,2}), \mathrm{VQ}_{3,4}(r_{n,3}, r_{n,4}),\\ &\quad \ldots, \mathrm{VQ}_{M-1,M}(r_{n,M-1}, r_{n,M})\}^T,\\ &\quad 1 \leq n \leq N\end{aligned} \tag{10}$$

と表現する．ETSI DSR フロントエンドの特徴

☆ 本論文では，正規化後の特徴パラメータと音響モデルとのミスマッチを軽減するために初期参照ケプストラム作成データに音響モデル学習データを用いた．

☆☆ ただし，ETSI Advanced 標準 DSR フロントエンドでは MFCC11，12 次元のコードブックサイズは 5 bit となっている．

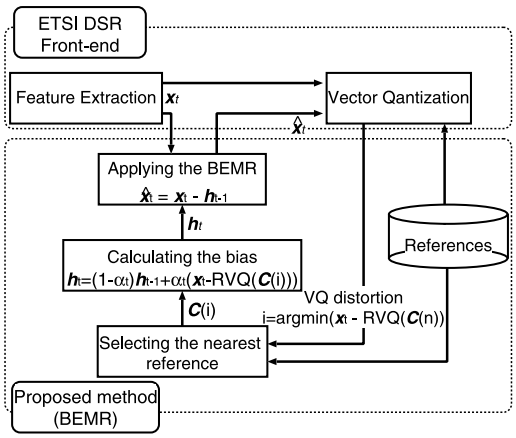

図 2 提案手法の流れ
Fig. 2 Block diagram of the proposed method.

パラメータ圧縮部の VQ コードブックで参照ケプストラムを表現した場合，6（量子化 bit 数）× 6（コードブック数）×N（参照ケプストラム数）で表現が可能であり，量子化を行わない場合と比較し約 90%のメモリ削減が可能である．

3.2 提案手法の適用部

本節では，前節で説明した参照ケプストラムを用い，周波数特性を正規化する方法（提案手法の適用方法）を説明する．提案手法の流れを**図 2** に示し，以下で各部分を説明する．

（A） 提案手法の適用

前フレームで計算されたバイアスを用いて，特徴パラメータ圧縮部に送る特徴パラメータを以下の式で正規化する．

$$\hat{\boldsymbol{x}}_t = \boldsymbol{x}_t - \boldsymbol{h}_{t-1}, \quad 1 \leq t \leq T \qquad (11)$$

ここで，$\boldsymbol{x}_t$ は特徴パラメータ，$\boldsymbol{h}_{t-1}$ は前フレームで計算されたバイアスを示す．t は現時点のフレーム番号とする．このバイアスは以下の（C）で計算され，周波数特性の正規化に使用する．また，$\hat{\boldsymbol{x}}_t$ は特徴パラメータ圧縮部に送られる特徴パラメータである．ただし，$\boldsymbol{h}_0$ は学習データ全体の特徴パラメータの平均と VQ コードブックのセントロイドの平均の差とする．

（B） 参照ケプストラムの選択

特徴パラメータ圧縮部で計算される情報を基に入力特徴パラメータに対し，最近傍の参照ケプストラムを以下の式で計算する．

$$i = \underset{j}{argmin}(|\boldsymbol{x}_t - \mathrm{RVQ}(\boldsymbol{C}(j))|) \qquad (12)$$

ここで，RVQ() はセントロイド番号からそのセントロイドベクトルを返す関数，i は最近傍の参照ケプストラム番号とする．

式 (12) では，最近傍の参照ケプストラムを計算するために入力特徴パラメータと参照ケプストラムとの距離を計算する必要がある．提案手法では，参照ケプストラムは特徴パラメータ圧縮部で使用される VQ セントロイドの組合せで表現されるため，距離計算には特徴パラメータ圧縮部で使用されている VQ（図 2 の Vector Quantization 部分）が利用可能である☆．そのため，提案手法の距離計算部を新たにフロントエンド部に作成する必要がなく，フロントエンド部のコンパクト化が可能である．

（C） 次フレームで使用されるバイアスの計算

（B）で選択された参照ケプストラム（$\boldsymbol{C}(i)$）を用い，次フレームで使用するバイアスを

$$\boldsymbol{h}_t = (1-\alpha_t)\cdot\boldsymbol{h}_{t-1} + \alpha_t \cdot (\boldsymbol{x}_t - \mathrm{RVQ}(\boldsymbol{C}(i))) \qquad (13)$$

で計算する．ここで，α_t は更新係数を示す．この更新係数（α_t）は，現在までのバイアスの平均として，次式で計算される．

$$\alpha_t = \frac{1}{t} \qquad (14)$$

提案手法では，現在までに計算されたバイアスの平均となるようにバイアスを更新するが，Blind Equalization 手法では入力フレームのパワーを使用しバイアスの更新を行っている（式 (8)）．

4. 音声認識実験

提案手法の有効性を確認するため，日本音響学会新聞記事読み上げ音声コーパス（JNAS）[9] を用い，音声認識実験を行った．音声データは 8 kHz にダウンサンプリングを行い，実験に使用した．

4.1 実 験 条 件

音響モデルの学習には，男性話者が発声した音素バランス文（話者：103 名，発声数：5,168 発声）を使用した．テストセットとして，男性話者 23 名が発声した 330 文（総単語数 5,160）を用いた．

音響モデル学習に用いた特徴パラメータは，ETSI Advanced DSR フロントエンドで分析をした MFCC 12 次元，その一次回帰係数，対数パワーの一次回帰

☆ DSR フロントエンドにおける特徴パラメータ圧縮部では分割 VQ が採用されている．提案手法では，分割 VQ で使用されているセントロイドの組合せで参照ケプストラムを表現している．そのため，参照ケプストラムと入力特徴パラメータとの距離計算は，参照ケプストラムを構成する各 VQ のセントロイドと入力特徴パラメータとの距離計算の組合せとして計算ができる．

係数の合計 25 次元である．

音響モデルは，各特徴量で学習を行った木構造クラスタリングにより状態共有した 3 状態 16 混合の音素環境依存 HMM（43 音素）の混合連続分布 HMM を用いた．総状態数は各特徴量ともに約 1,000 状態である．文献 10) で，VQ を行わない特徴パラメータで音響モデルを学習することにより，伝送される特徴パラメータ（VQ された特徴パラメータ）の認識性能が向上することを報告した．そのため本論文でも，音響モデルは圧縮（VQ）を行っていない特徴パラメータで学習した．認識時の特徴パラメータは，VQ により圧縮された特徴パラメータを用いた．

デコーダには Julius（Ver.3.1p2）[11]，言語モデルは *tri*-gram，辞書単語数は約 20,000 単語を用い連続音声認識実験を行った．評価には単語誤り率（WER: Word Error Rate）を用いた．各実験の WER は，テストセットに対し最も WER が低くなるようデコード時の最適なパスの広さの設定を各々の条件で行った結果より計算した．

また，周波数特性の差異の影響を検討するため，テストセットに対し，ITU で定義されている[12]

- G712
- MIRS

のフィルタを適用し，人工的に乗算性雑音を加えた音声を作成した．作成した音声データを用い，乗算性雑音に対する提案手法の有効性をシミュレーションした．これらのフィルタは Aurora2 データベースで使用されているものを使用した[13),☆]．

本論文で提案した周波数特性正規化手法の評価にあたっては，ETSI Advanced DSR フロントエンドにおける Blind Equalization 手法と置き換え，実験を行った．提案手法の参照ケプストラム数は 16 とし，参照ケプストラム選択時には 2 乗距離を使用した．なお，参照ケプストラム数に関しては，5.1 節で詳細な検討を行う．

4.2 認識実験結果

表 1 に音声認識実験結果を示す．表中の “BEMR” は提案手法を示し，“ES202” は ETSI が勧告した Advanced DSR フロントエンド（ES202 050 v1.1.1）を示す．また，“ES202 w/o BE” は周波数特性正規化手法である Blind Equalization 手法を除いた Advanced DSR フロントエンドを示す．“BRM” は以前に提案し 2 章で説明した平均一致化手法[6] を示す．ただし，この手法は実時間処理での周波数特性正規化は実現されていない．

表 1 音声認識実験結果（WER: Word Error Rate（in %））
Table 1 Recognition results on Japanese speech corpus using proposed methods (WER: Word Error Rate (in %)).

	Filter		
	clean	G712	MIRS
BEMR	15.3	15.5	15.7
ES202 w/o BE	16.5	20.9	30.2
ES202	15.8	16.9	17.6
BRM	14.2	14.5	14.5

表 1 より，提案手法である BEMR は ES202 w/o BE の WER をすべての環境下で削減できていることが分かる．特に ES202 w/o BE の場合，学習データと評価データ間に周波数特性の差異がある場合（G712, MIRS），差異がない場合（clean）と比較して WER が増加していることが分かる．しかし，提案手法の場合，差異が生じたとしても WER の増加はほぼみられず，有効に周波数特性の正規化が行えていることが分かる．

また ES202 との比較においても，提案手法はすべての環境下で ES202 の WER を減少させていることが分かる．特に周波数特性の差異が大きい MIRS フィルタの環境下では，10.8%の誤り削減率を示している．これにより，周波数特性を正規化するバイアス計算に複数の参照ケプストラムを使用することが周波数特性の変化による認識性能劣化の抑制に有効であるといえる．

しかし BEMR と BRM との比較では，BRM の WER が BEMR より低いことが分かる．BEMR と BRM の相違はバイアス計算の部分のみであるため，今後はバイアス計算に関してさらなる改良をする必要があると思われる．

5. 考　　察

4.2 節において，提案手法が入力系の周波数特性に差異が生じた場合の認識性能の劣化の抑制に有効であることを示した．提案手法は，量子化された複数の参照ケプストラムを用い周波数特性を正規化するバイアスを計算する．そのため，参照ケプストラムの決定方法が認識精度に大きく影響すると考えられる．そこで本章では，

- 参照ケプストラム数と認識精度との関連（5.1 節）
- 参照ケプストラムの量子化の影響（5.2 節）

☆ Aurora2 データベースでは G712 のフィルタは学習データに使用されており，MIRS フィルタは周波数特性の差異が生じた場合の評価に使用されている．本論文では，学習データにはフィルタリングを行わず，これら 2 つのフィルタを周波数特性の差異が生じた場合として評価した．

について，考察を行う．考察のための音声認識実験は4.1節と同様の実験条件で行った．

5.1 参照ケプストラム数と認識精度との関係

提案手法は，複数の参照ケプストラムを使用し周波数特性を正規化するためのバイアスを計算する．そのため，使用する参照ケプストラム数により認識精度が変動することが考えられる．そこで，バイアス計算に用いる参照ケプストラム数と認識精度との関連性を調べた．**表2**に参照ケプストラム数とWERを示す．

この表より，参照ケプストラム数が16の場合に最もWERが低いことが分かる．参照ケプストラム数が多い場合（32，64），提案手法では最近傍の参照ケプストラムよりバイアス計算が行われているため（式(13)），周波数特性を正規化するバイアスは非常に小さい値になっていると考えられる．そこで，提案手法適応後のDSRフロントエンド部における特徴量圧縮部のVQ歪を調べた．この結果，参照ケプストラム数が多い場合（32，64）において，適応を行わない特徴パラメータと比較するとVQ歪が減少していることが分かった．しかし，WERが最も低い参照ケプストラム数16の場合と比較をすると参照ケプストラム数が多い場合（32，64）のVQ歪は増加していることが分かった．これより，参照ケプストラム数が多くなり正規化するバイアスが小さい値となったため，特徴パラメータの正規化が適切に行われずVQ歪の減少が抑制され，WERが増加したと考えられる．

一方，参照ケプストラムが少ない場合（4，8），Advanced DSRフロントエンドで導入されているBlind Equalization手法と同様に適切な周波数特性の正規化が行われず，WERの増加につながったと考えられる．さらに，参照ケプストラム数が1の場合，参照ケプストラムが少ない場合（4，8）と比較して認識精度の低下が抑制されていることが分かる．しかし，最も適切に参照ケプストラム数を選択した場合よりは認識精度が低下していることが分かる．これにより，適切な参照ケプストラム数を決定することは本提案手法にとって非常に重要な課題であるといえる．今後，最適な参照ケプストラム数を自動的に決定する手法に関してさらなる研究を進めていく予定である．

5.2 参照ケプストラムの量子化の影響

提案手法はターミナル側のメモリ量の削減および計算コストの削減のため，参照ケプストラムを量子化している．この量子化の影響が認識性能に影響を与えるかを調べた．本考察では5.1節の実験で最もWERが低かった，参照ケプストラム数16を用いた．実験結果を**表3**に示す．この結果より，参照ケプストラムの量子化の影響はほとんどないことが分かる．そのため，量子化によるメモリ量の削減および計算コストの削減は有効であるといえる．

表2 参照ケプストラム数と認識精度との関連（WER: Word Error Rate（in %））
Table 2 Recognition results as a function of the number of references (WER: Word Error Rate (in %)).

Number of references	Filter		
	clean	G712	MIRS
1	15.4	16.4	15.8
4	16.5	17.4	16.2
8	16.3	17.2	16.2
16	**15.3**	**15.5**	**15.7**
32	17.7	17.6	17.3
64	17.5	16.7	16.8

表3 参照ケプストラムの量子化による影響
Table 3 Influence of the quantized reference cepstrum.

	Filter		
	clean	G712	MIRS
量子化あり	15.3	15.5	15.7
量子化なし	15.8	16.0	15.9

6. むすび

本論文では，分散音声認識（DSR: Distributed Speech Recognition）における入力系の周波数特性の差異による認識性能劣化を抑制する周波数特性正規化手法として，複数参照ケプストラムを用いた実時間周波数特性正規化手法を提案した．提案手法は，複数の参照ケプストラムを使用し，周波数特性の正規化を行うバイアスをフレーム同期で計算し，実時間で入力系の周波数特性を正規化する手法である．

一般に，DSRで用いられるクライアントではメモリ量，計算量の制限がある．そこで，提案手法では参照ケプストラムをDSRフロントエンドの特徴パラメータ圧縮部で使用されるVQコードブックの組合せで表現し，フロントエンド部における参照ケプストラム保持のためのメモリ量増加，参照ケプストラム選択の計算量増加を抑制した．

ETSI Advanced DSRフロントエンドを用いた日本音響学会新聞記事読み上げ音声コーパスの音声認識実験より，提案手法は，ETSI Advanced DSRフロントエンドにおけるBlind Equalizationと比較し，周波数特性の差異による音声認識精度劣化の抑制に有効であることを確認した．実際，提案手法は周波数特性の差異のシミュレートに使用した全フィルタに対し，ETSI Advanced DSRフロントエンドより高い認識精度を示した．特に，提案手法はMIRSフィルタ条件下でETSI Advanced DSRフロントエンド（Blind Equal-

ization）の単語誤り率を 10.8%削減（17.6%→15.7%）することが可能であった．これより，ETSI Advanced DSR フロントエンド内で採用されている 1 つの参照ケプストラムを用い周波数特性を正規化する Blind Equalization 手法より，複数の参照ケプストラムを選択し使用する提案手法が周波数特性の正規化に有効であるといえる．

しかし，提案手法は平均一致化手法の認識精度より若干低いことが分かった．提案手法と平均一致化手法とは周波数特性の正規化に使用するバイアス計算が異なるだけであるため，今後はバイアス計算方法に関しさらに研究を行う予定である．具体的には，入力特徴パラメータとの距離が近い複数の参照ケプストラムをバイアス計算に用いることを検討する予定である．また，バイアスの更新係数（式 (14) の α_t）は，現在までの平均バイアスとなるように計算されている．そのため，ある程度バイアスの推定が行われた後にはバイアスの更新がほとんどされなくなる．平均一致化手法では，発声開始から約 2 秒間の音声で計算したバイアスを用いた場合，発声全体でバイアスを計算した場合とほぼ同等の認識精度を示すことが分かっている．そこで，今後，計算量のさらなる削減を行うため，バイアスの変動が少なくなった場合，更新係数を一定にする改良を行う予定である．

また，本論文では最適なケプストラム数は実験的に求めたが，最適な参照ケプストラムを自動的に決定する手法に関しても研究を行う予定である．さらに，本提案手法との比較実験として CDCN との比較実験を行う予定である．

謝辞 本研究の一部は文部科学省科学研究費，若手研究（B）15700163，基盤研究（B）17300065，17300036，萌芽研究 17656128 の補助を受けて行った．

参 考 文 献

1) Lilly, B. and Paliwal, K.: Effect of Speech Coders on Speech Recognition Performance, *Proc. ICSLP*, pp.2344–2347 (1996).
2) Pearce, D.: Enabling New Speech Driven Services for Mobile Devices: An overview of the ETSI standards activities for Distributed Speech Recognition Front-ends, *AVIOS* (2000).
3) ETSI ES 201 108 v1.1.2 Distributed Speech Recognition; Front-end Feature Extraction Algorithm; Compression Algorithm (2000).
4) ETSI ES 202 050 v1.1.1 STQ; Distributed Speech Recognition; Advanced Front-End Feature Extraction Algorithm; Compression Algorithms (2002).
5) Liu, F., Stern, R., Huang, X. and Acero, A.: Efficient Cepstral Normalization for Robust Speech Recognition, *Proc. DARPA Workshop*, pp.69–74 (1993).
6) 柘植 覚，黒岩眞吾，獅々堀正幹，北 研二：ETSI 標準分散音声認識フロントエンドにおける入力系の周波数特性正規化手法，電気学会論文誌 C，Vol.125, No.1, pp.120–127 (2005).
7) Mauuary, L.: Blind Equalization in the Cepstral Domain for Robust Telephone based Speech Recognition, *EUSPICO*, pp.359–363 (1998).
8) Acero, A. and Stern, R.: Environmental Robustness in Automatic Speech Recognition, *Proc. ICASSP*, Vol.1, pp.849–852 (1990).
9) 鹿野清宏，伊藤克亘，河原達也，武田一哉，山本幹雄：音声認識システム，オーム社出版局 (2001).
10) 柘植 覚，黒岩眞吾：周波数特性の変動に頑健な分散音声認識手法，*SLP-42*, No.13, pp.77–84 (2002).
11) Lee, A., Kawahara, T. and K.Shikano: Julius — An Open Source Real-Time Large Vocabulary Recognition Engine, *Proc. EuroSpeech*, pp.1691–1694 (2001).
12) ITU-T Recommendation G.723.1 Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s (1996).
13) Hirsch, H. and Pearce, D.: The AURORA Experimental Framework for the Performance Evaluations of Speech Recognition Systems under Noisy Conditions, *ISCA ITRW ASR*, pp.181–188 (2000).

(平成 17 年 12 月 2 日受付)
(平成 18 年 11 月 2 日採録)

柘植 覚（正会員）

平成 8 年徳島大学工学部知能情報工学科卒業，平成 10 年同大学大学院工学研究科博士前期課程知能情報工学専攻修了，平成 13 年同大学院工学研究科博士後期課程システム工学専攻修了．平成 12 年徳島大学工学部助手．平成 18 年徳島大学工学部講師．博士（工学）．音声認識，情報検索等の研究に従事．日本音響学会会員．

黒岩　眞吾（正会員）

昭和 61 年電気通信大学電気通信学部通信工学科卒業，昭和 63 年同大学大学院修士課程了，博士（工学）．同年国際電信電話株式会社入社．昭和 63〜平成 13 年同社研究所において電話音声認識システムの研究・開発に従事．平成 13 年徳島大学工学部助教授．音声認識，話者照合，情報検索の研究に従事．電子情報通信学会平成 8 年度学術奨励賞，日本音響学会第 3 回および第 5 回技術開発賞受賞．日本音響学会，電子情報通信学会，人工知能学会各会員．

獅々堀正幹（正会員）

平成 3 年徳島大学工学部情報工学科卒業，平成 5 年同大学大学院博士前期課程修了，平成 7 年同大学院博士後期課程退学．同年同大学工学部知能情報工学科助手．平成 9 年同大学工学部知能情報工学科講師．平成 13 年同大学工学部知能情報工学科助教授．現在，同大学工学部知能情報工学科助教授．博士（工学）．マルチメディア情報検索，自然言語処理の研究に従事．著書『情報検索アルゴリズム』（共立出版），情報処理学会第 45 回全国大会奨励賞受賞．電子情報通信学会，言語処理学会各会員．

任　　福継（正会員）

昭和 57 年北京郵電大学電信工程卒業，昭和 60 年同大学大学院修士課程修了．平成 3 年北海道大学大学院工学研究科博士後期課程修了，博士（工学）．CSK 研究員，広島市立大学情報科学部助教授を経て，平成 13 年徳島大学工学部教授．現在，同大学大学院ソシオテクノサイエンス研究部教授．平成 8〜9 年アメリカニューメキシコ大学訪問教授．平成 18 年北京郵電大学情報工学部教授（兼務）．自然言語処理，感性情報処理，学習システム，人工知能，多言語多機能多メディア知的システムに関する研究に従事．IEEE，ACL，言語処理学会，教育システム情報学会各会員．

北　　研二（正会員）

昭和 56 年早稲田大学理工学部数学科卒業．昭和 58 年沖電気工業（株）入社．昭和 62 年 ATR 自動翻訳電話研究所出向．平成 4 年徳島大学工学部講師．平成 5 年同助教授．平成 12 年同教授．平成 14 年同大学高度情報化基盤センター教授．工学博士．自然言語処理，情報検索等の研究に従事．平成 6 年日本音響学会技術開発賞受賞．著書『確率的言語モデル』（東京大学出版会），『情報検索アルゴリズム』（共著，共立出版）等．