

## コードブックを用いた実時間処理 CMN の実装と評価

追立 真吾<sup>†</sup> 戸田 智基<sup>†</sup> 李 晃伸<sup>††</sup> 川波 弘道<sup>†</sup> 猿渡 洋<sup>†</sup>  
鹿野 清宏<sup>†</sup>

<sup>†</sup> 奈良先端科学技術大学院大学情報科学研究科 〒 630-0101 奈良県生駒市高山町 8916-5

<sup>††</sup> 名古屋工業大学大学院工学研究科 〒 466-8555 愛知県名古屋市昭和区御器所町

E-mail: <sup>†</sup>{shingo-o,sawatari,shikano}@is.naist.jp, <sup>††</sup>{tomoki,ri,kawanami}@nitech.ac.jp

あらまし 人型ロボットや券売機など公共の音声認識システムには話者交替が頻繁に生じるという特徴があるため、高速な話者適応技術が重要である。話者、空間、収録機器等を簡便に一括して正規化する手法としてケプストラム平均正規化 (CMN) が広く用いられている。公共の音声認識システムでは、空間や収録機器の特性が急激な変化をすることは少ないので、CMN は主に話者正規化の効果を果たす。しかしながら、CMN は一般に発話単位でケプストラム平均値を算出するため、発話の入力開始と同時に認識処理を開始できない。そのため公共の音声認識システムでは必須である実時間処理が困難である。また、公共の音声認識システムは発話が短時間である傾向があるが、そのような場合、CMN は音韻の偏りの影響により性能が劣化してしまうという問題点がある。そこで、認識処理中に母音区間を検出し、母音ケプストラムを入力情報として、コードブックにより発話者に近い話者の発話を発話データベースから選択し、近似的に長時間 CM をフレーム同期で求める手法を提案する。音声対話情報案内システム「たけまるくん」により収集した実環境データを用いた実験により、公共の音声認識システムにおけるオンライン/オフライン CMN の効果を明らかにする。結果、提案法は実時間処理であるが、発話単位で CM を算出する CMN に匹敵する性能を示した。また、代表的な従来の実時間処理 CMN と比較しても、提案法が優位であることがわかった。

キーワード 話者クラスタリング コードブック CMN 実時間 短時間発話

## Real-time Cepstrum Mean Normalization Using Codebook

Shingo OIDATE<sup>†</sup>, Tomoki TODA<sup>†</sup>, Akinobu LEE<sup>††</sup>, Hiromichi KAWANAMI<sup>†</sup>, Hiroshi SARUWATARI<sup>†</sup>, and Kiyohiro SHIKANO<sup>†</sup>

<sup>†</sup> Nara Institute of Science and Technology Takayama-cho 8916-5, Ikoma-shi, Nara, 630-0101 Japan

<sup>††</sup> Nagoya Institute of Technology Gokiso-Cho, Showa-Ku, Nagoya, 466-8555 Japan

E-mail: <sup>†</sup>{shingo-o,sawatari,shikano}@is.naist.jp, <sup>††</sup>{tomoki,ri,kawanami}@nitech.ac.jp

**Abstract** A rapid speaker adaptation in a public speech-recognition system such as the humanoid robots and ticket vending machines is important because a speaker frequently exchanges. Cepstral Mean Normalization (CMN) is widely used as a technique for normalizing the transmission characteristics such as the speaker, the spaces, and the collection equipment. Especially, because the transmission characteristics of the space and the collection equipment is almost consistent in a public speech-recognition system, CMN mainly causes the effect of the speaker adaptation. The recognition processing cannot begin at the same time as the input of the utterance beginning because CM (Cepstral Mean) is calculated over an utterance in general. Therefore, a real-time processing is difficult. We propose a technique for in real time approximately obtaining CM by using the codebook. The proposed method is evaluated using the data which were collected with the speech-oriented information system "Takemaru-kun." We also demonstrate the performance of several offline CMN methods. The result shows that the proposed method works better than the conventional methods.

**Key words** speaker clustering, codebook, CMN, real-time, short utterance

## 1. ま え が き

近年、音声認識技術の発展や計算機の高性能化により、様々な音声対話インターフェースが実用化され、注目を集めている。さらに今後、音声認識の応用として、先の万博でも登場した人型ロボットや、駅券売機などのような公共の場で用いられることが考えられる。

公共の音声認識システムの特徴として話者交替が頻繁であり、発話が短時間であるということが挙げられる。話者性は音声認識性能に大きな影響を与える要因の1つであり、古くから話者適応技術が盛んに研究されているが、数発話～数十発話を適応データとして用いたオフラインでの音響モデルの適応が一般的である。しかしながら公共の音声認識システムでは1発話での話者交替も多く、音響モデルでの話者適応は困難である。実際、実環境音声情報案内システム「たけまるくん」[1]では平均発話時間は1.9秒(発話の前後に合計約0.6秒の無音区間を含む)であり、システム利用時の発話回数1回のユーザーが全体の50%以上である[2]。

1発話で有効に話者を正規化できる手法にケプストラム平均正規化(以降、CMN(Cepstral Mean Normalization)と呼ぶ)がある。CMNは簡便な処理でありながら、十分に長い発話からケプストラム平均(以降、CM(Cepstral Mean)と呼ぶ)を計算する場合には、マイクなどの入力デバイスの違いに起因する周波数特性の差異を正規化することによる性能改善効果が認められており、広く用いられている。また、話者性の正規化の効果も報告されている[3]。

短時間発話の場合、発話内容に音韻的偏りが生じるので、発話単位のCMN処理は逆に性能を劣化させることがある。この問題に対し、認識処理結果を用いて音響モデルに対して最尤になるようなCMを推定し、この値を用い、再度CMNを行って認識処理を行う手法(以降、ML-CMN(Maximum Likelihood CMN)と呼ぶ)がある[4]。また、音声/非音声フレームを区別せずに、CMを求める方法に比べて、音声/非音声フレームで別々にCMを求めCMNを行う手法(以降、E-CMN(Exact CMN)と呼ぶ)が提案され性能改善効果が報告されている[5]。さらに認識結果に基づいてCMを音声/非音声フレームで別々に算出し音声区間CMに対してML推定を行い、再度、CMNと認識処理を行う手法(以降、ML-E-CMNと呼ぶ)が提案されている[6]。

上記の発話単位でのCMNは1発話全体の情報を用いるため、発話開始と同時に処理を開始できない。よって発話終了時にはじめて認識処理が開始される。これでは発話者にシステムが応答するまでに大きく時間がかかってしまう。実時間処理のためには音声入力開始と同時に

認識処理の開始可能なフレーム同期処理が必須である。

CMNのフレーム同期処理を目指した従来法として、当該フレームの前後Nフレームを用いてCMを計算し、正規化処理を行う手法(以後、前後NフレームCMNと呼ぶ)[7]や、当該フレームまでのCMと学習データ全体のCMを用いてMAP推定を行う手法(以後、MAP-CMNと呼ぶ)[8]、過去の1つ以上の発話からCMを計算する手法が提案されている。しかし、前後NフレームCMNでは入力開始から正規化の処理を開始するまでに、Nフレームの遅延が生じる。さらにNの値が小さい場合、認識性能が発話の音韻的偏りに悪影響を受けてしまう。MAP-CMNにおいても、発話の初期においてはこの音韻的偏りの問題が生じる。過去の発話からCMを算出する方法は、音韻的偏りの影響を受けにくい、正規化対象音声の話者がCM算出時の話者と異なる場合、認識性能の低下を引き起こすことがしばしばある。

そこで、我々は発話内容に依存せず、かつフレーム同期処理が可能なCMNの適用法を提案した[2]。提案法は認識時の最尤状態系列から母音区間を検出し、母音区間情報と長時間の発話全体CMの対応を表すコードブックによりCMを推定し、正規化処理を行う手法である。この方式では、CM計算区間が音韻的に偏る場合でも、発話中の母音区間から長時間の発話全体CMを推定するため、フレーム同期で効果的に正規化が行えると期待される。

本稿では、前述したオフラインでのCMNの各種手法を「たけまるくん」により収録したデータを用いて、実環境対話システムにおける発話単位での話者正規化の効果を評価する。また、我々の提案手法であるコードブックを用いたフレーム同期処理CMNの実装および評価を行う。

以下、2節では発話単位での各CMNの手法について詳細を述べ、3節でその実験結果を示す。4節ではフレーム同期処理CMNの従来法を述べ、5節でコードブックを用いたフレーム同期処理CMNについて述べ、6節でフレーム同期処理CMNの実験結果を示す。最後に7節でまとめを述べる。

## 2. オフラインCMNの従来手法

### 2.1 CMN

1発話を単位とする一般的なCMNは次式で表すことができる[9]。

$$\hat{C}(t) = C(t) - \frac{1}{T} \sum_{i=1}^T C(i) \quad (1)$$

ここでは $C$ はCMN処理前、 $\hat{C}$ はCMN処理後のケプストラム係数のベクトルであり、 $i, t$ はフレーム番号で、 $T$ は入力発話のフレーム総数である。上式右辺第2項の

CMには、話者、空間、収録機器の違いから生じた音響伝達特性の偏りを代表することが期待される。

この手法ではCMを認識時の当該フレームを含む1発声から求めるため、発声が終了するまで認識処理を開始することができない。さらに、入力が短い音声の場合、CMが発声内容に依存した値となり、CMNにより悪影響が生じるものと考えられる。

## 2.2 ML-CMN

ML-CMNは1段階目の認識処理結果を用いて音響モデルに対して最尤になるようなCMを推定し、この値を用い、再度CMNを行って2段階目の認識処理を行い最終的な認識結果とする手法である。この手法では音韻の偏りを音響モデルにより緩和するため短時間発話からでも精度の高い長時間CMが予測できると考えられる。

CMの最尤推定式を式次式に示す。

$$\hat{C}(t) = X^{-1} \times Y \quad (2)$$

$$X = \sum_{t=1}^T \sum_{n=1}^n \sum_{m=1}^M \Sigma_{n,m}^{-1} \gamma_t(n, m) \quad (3)$$

$$Y = \sum_{t=1}^T \sum_{n=1}^n \sum_{m=1}^M \Sigma_{n,m}^{-1} \gamma_t(n, m) (y_t - \mu_{n,m}) \quad (4)$$

ここで、 $\hat{C}(t)$ が推定後のCMの値、 $n, N$ は状態番号と総状態数、 $m, M$ はミクスチャー番号と混合数、 $\gamma_t(n, m)$ は時刻 $t$ に状態 $n$ の $m$ 番目の正規分布に滞在する確率、 $y_t$ は時刻 $t$ での観測ベクトル、 $\mu_{n,m}$ は状態 $n$ の $m$ 番目の正規分布の平均ベクトル、 $\Sigma_{n,m}$ は状態 $n$ の $m$ 番目の正規分布の共分散行列である。

## 2.3 E-CMN

音声/非音声フレームを区別せずに、CMを求めることは、本来求めたい乗法性ひずみだけではなく、非音声フレームの影響を受けることになり、その影響はCMを求めるデータ範囲に含まれる非音声フレームの長さによって大きく変動する。E-CMNは音声/非音声フレームで別々にCMを求めCMNを行う手法である。音声/非音声フレームを区別せずに、CMを求める方法に比べて、性能改善効果が報告されている[5]。ただし、この手法を用いることにより、CMの算出区間は短くなるので、短時間発話では有効ではないことがある。

## 2.4 ML-E-CMN

短時間発話での性能劣化というE-CMNの問題点と無音区間によって正確に話者性を推定できないML-CMNの問題点を解決するために、音声/非音声フレームに分けたのち、音声区間と非音声区間のそれぞれに対してML-CMNの処理を行う。

## 3. オフライン実験

公共の音声認識システムに対する各CMNの有効性を

表1 音響分析条件

Table 1 Acoustic analysis conditions

サンプリング周波数	16kHz
フレーム周期	10ms
ハミング窓長	25ms
メルフィルタバンク数	24
特徴パラメータ	MFCC 1-12次 (CMN) MFCC 1-12次 対数パワー 1次

検証するため、実環境音声情報案内システム「たけまるくん」[1]の収集データを用いてオフラインでの評価実験を行った。

### 3.1 実験条件

たけまるデータベースの発話内容が聞き取れる発話データのうちの、主観的に大人であると判別したものを学習データ、評価データとして用いた。全発話内容の書き起こしテキストから、出現頻度上位1000位を選び、この発話内容にマッチする男女各1発話の音声データを抽出した。結果得られた2000発話の評価データとして用いた。学習データとして評価データを含まない17000発話を用いた。音響分析条件を表1に示す。音響モデル(PTM, 2000状態, 64混合)は評価データと同様の手法により作成した学習データのMFCCを用いてEM学習により作成した。本実験では、収集した発話データの書き起こしテキストより作成した3-gramを用いた。音声認識エンジンにはJulius[10] ver.3.5を用いた。

ML-CMN, ML-E-CMNについては前段の認識処理が必要となってくるが、これは発話単位CMNを行った認識結果を用いた。E-CMN, ML-E-CMNにおいて、非音声フレームに対しては、学習データ全体のうちに含まれる非音声区間のみから算出した非音声区間CMを用いた。

### 3.2 実験結果

実験結果を図1に示す。ここでwoCMNは学習データ、評価データともにCMNを行わない場合である。公共の音声認識システムにおいてもCMNは有効であるといえる。一方で、E-CMNとML-CMNは通常のCMNに対して優位な性能を示さなかった。E-CMNに関しては、短時間発話ではCMNよりも性能が劣化してしまうことが報告されており[3]、公共の音声認識システムのような短時間発話の多いタスクでは十分な性能が得られないことがわかる。また、ML-CMNではCMの最尤推定時に無音区間の影響により、十分な性能が得られないことがわかる。これに対して、それぞれの問題を解決したML-E-CMNでは0.5%程度の性能向上が見られた。

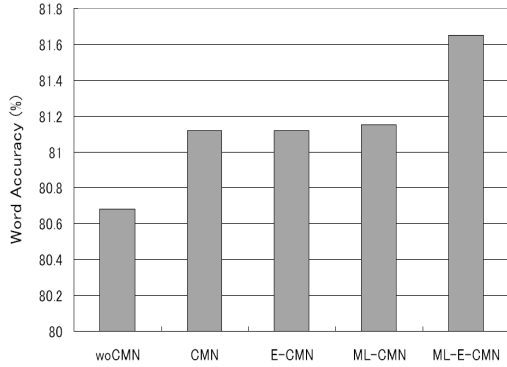


図1 オフラインでの各 CMN の単語正解精度 (%)

Fig. 1 Recognition performance of offline CMN methods

#### 4. オンライン CMN の従来法

従来のフレーム同期処理可能な CMN は大きく 2 種類のアプローチが考えられる。1 つは当該発話の部分的な情報から CM を算出する手法、もう 1 つは当該発話以外の発話データから CM を予測する手法である。

1 番目のアプローチをとる手法として、当該フレームの前後  $N$  フレームを用いて CM を計算し、正規化処理を行う手法 (以降、前後  $N$  フレーム CMN と呼ぶ) [7] が挙げられる。2 番目のアプローチをとる手法として、過去の 1 つ以上の発話から CM を計算する手法が提案されている。さらに 2 つのアプローチを合わせたものに、当該フレームまでの CM と学習データ全体の CM を用いて MAP 推定を行う手法 (以降、MAP-CMN と呼ぶ) [8] が挙げられる。

##### 4.1 前後 $N$ フレーム CMN

式 (1) の右辺第 2 項の CM を対象フレームの前後  $N$  フレームの平均値で代用することでフレーム同期で CMN を行う [7]。

$$\hat{C}(t) = C(t) - \frac{1}{2N+1} \sum_{i=-N}^N C(i) \quad (5)$$

この操作により、 $N$  フレームの遅れで正規化処理が可能となる。しかし、 $N$  フレームの遅延が大きい場合、実際の音声対話システムに導入するのは難しく、 $N$  の値が小さい場合、CM 算出区間の音素の偏りによって性能が劣化してしまう。

##### 4.2 MAP-CMN

MAP-CMN [8] は MAP 推定を用いることにより、発声開始直後の CM の推定精度の向上を図ることで、実時間処理で CMN を行う。CM は式にて計算される。

$$\hat{C}(t) = C(t) - \left( \frac{\tau}{t+\tau} \mu + \frac{1}{t+\tau} \sum_{i=1}^t C(i) \right) \quad (6)$$

ここで、 $t$  は入力発話の当該フレームまでのフレーム数、 $\tau$  は事前知識の確からしさに関する係数である。本稿では事前確率の平均値  $\mu$  として音響モデル用学習データ全体の CM を用いる。

#### 5. 母音区間情報とコードブックを用いたフレーム同期 CMN

入力発話の部分的な特徴を入力情報として用い、発話データベースから声質の近い話者の発話を選択し、長時間 CM を推定する手法を提案する。

母音区間では話者性がよく表れるといわれている。母音区間は短い発話にも含まれており、一般に雑音に頑健である。よって、母音区間から話者性を推定すれば、短時間の発話からでも長時間 CM が推定できる。そこで、母音区間ケプストラムを入力情報とした長時間 CM を推定するためのコードブックを作成する。

まず発話データベースを話者性でクラスタ化を行う。オンライン処理時に入力発話から話者性の情報を取り出して、入力発話の話者性に近いクラスタを選択することで、発話データベースから音韻的偏りが少ない安定した長時間 CM を得ることができる。

さらにフレーム同期処理のために、認識処理過程で得られる最尤状態系列から、逐次、母音区間を確定していき、コードブックから長時間 CM を推定する。提案手法の流れを図 2 に示す。まず、入力音声の特徴量抽出により MFCC に変換した後、予測した CM を用いた CMN を行う。得られた CMN された MFCC で認識処理を行う。その際、フレーム毎にビーム内に存在するパスを全てソートし、第 1 位のパスが存在する状態を用いて、当該フレームでの最尤状態とする。この最尤状態の決定の概念図を図 3 に示す。最尤状態が母音であったとき、各母音ごとに準備されたコードブックに、CMN していない MFCC と最尤状態がどの母音であったかという情報を入力するコードブックの出力に基づき次フレームのための CM を更新する。

この提案手法は母音区間情報からの話者性の推定精度が得られれば、短時間の音韻的偏りの大きい発話初期からでも、音韻的偏りの無い長時間の発話単位 CM を推定できる。よって、局所的な情報を使う前後  $N$  フレーム CMN や発話の音韻的偏りを考慮しない MAP-CMN に比べて、推定精度の向上が期待できる。

##### 5.1 母音ケプストラムと発話単位 CM の対応を表すコードブックの作成

コードブック作成手順を以下に示す。

1. 学習データの各発話に対して書き起こし文を用いて viterbi アルゴリズムにより、セグメンテーションを行い、母音区間を決定する。このとき、各発話毎に発話単

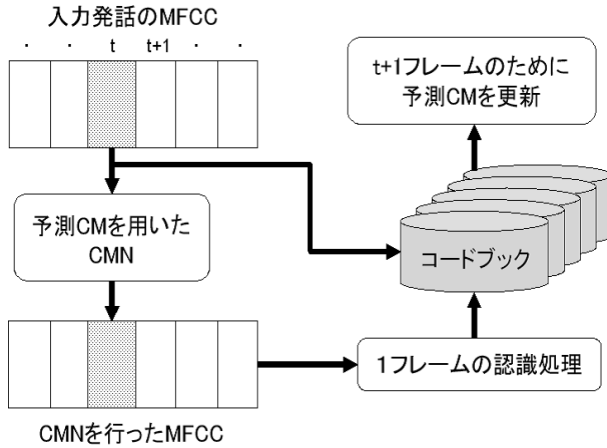


図 2 提案法の流れ

Fig. 2 Block diagram of proposed method

位 CM を算出しておき、母音区間ケプストラムとの対応を保持しておく。

2. 検出した母音区間フレームの各ケプストラムの値をクラスタリングする。音韻性の違いがケプストラムに与える影響を抑えるため、クラスタリングは各母音ごとに行う。クラスタリングは LBG アルゴリズムを用いる。

3. 得られた各セントロイドごとに対応する CM を算出する。これはセントロイドに含まれる母音ケプストラムに対応する発話単位 CM を重み付け平均することで算出する。重みとして用いるのは、同一の発話単位 CM に対応する母音ケプストラムがセントロイドに属している数である。

このように母音ケプストラムでクラスタリングを行うことで、話者性の近いものをクラスタリングでき、セントロイドに対応する CM は話者性の近いものでまとめられるので、近似的に長時間 CM の値が得られることが期待される。

### 5.2 コードブックを用いたフレーム同期での長時間 CM 推定

学習データ全体から CM を求め、これを用いて認識処理を開始する。当該フレームが母音と判断された際には、以下の式を用いて CM を更新する。

$$C_{pre}(n) = C_{cent}(v, m) + V_{in}(n) - V_{cent}(v, m) \quad (7)$$

$$CM(N) = \frac{\tau}{N + \tau} C_{global} + \frac{1}{N + \tau} \sum_{n=1}^N C_{pre}(n) \quad (8)$$

ここで、 $V_{in}(n)$  は入力音声から検出された母音ケプストラム係数ベクトルであり、コードブック中のセントロイドで最も  $V_{in}(n)$  に近いものが  $V_{cent}(v, m)$  である。このセントロイドの CM が  $C_{cent}(v, m)$  であり、フレーム毎に求まる予測された CM が  $C_{pre}(n)$  となる。この値の対

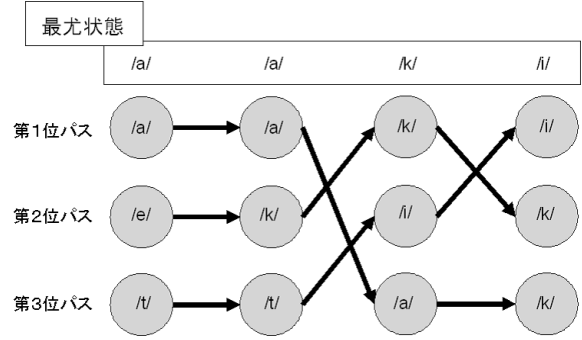


図 3 最尤状態の決定

Fig. 3 decision of ML state sequence

象フレームまでの平均値と学習データ全体の  $CM_{global}$  から MAP 推定により算出した  $CM(N)$  を用いて CMN を行う。 $v$  はどの母音であるかという情報、 $m$  はコードブック中のセントロイドの番号、 $n$  は母音フレームの番号で  $N$  は当該フレームまでの母音フレーム数である。

式 (7) の右辺における検出された母音ケプストラムとコードブック中の母音ケプストラムの差分は、コードブックの量子化ひずみを緩和するために用いる。式 (8) は各母音から推定される話者性のばらつきを抑えるために用いる。

## 6. オンライン実験

提案手法は Julius-3.5 を拡張することにより実装し、その性能をオンライン実験により評価した。

学習データ、評価データ、言語モデル、音響分析条件は 3 節と同様である。音響モデルは発話単位 CMN を行った学習データを用いて EM 学習により作成した。

### 6.1 パラメータの最適化

提案手法における重み係数  $\tau$  とクラス数をパラメータとして変動させ最適化を行った。

クラス数と重み係数をパラメータとしたときの単語正解精度の分布を図 4 で示す。重み係数  $\tau = 10$  のときに安定した性能を示すことがわかる。重み係数  $10$  のとき、クラス数を横軸に、単語正解精度を縦軸にとったものを発話単位で行う通常の CMN や CMN なしと比較したものを図 5 に示す。クラス数 256 で最もよい値が得られる。さらにフレーム同期処理である提案手法が発話単位 CMN に匹敵する性能を示していることもわかる。

### 6.2 従来法との比較

フレーム同期処理可能な従来法と提案法の比較を図 6 に示す。ここで、過去 3 発話とは直前の過去 3 発話の CM を用いて、入力発話全体から減算する手法である。

図より提案手法が従来法と比較して優位であることがわかる。また、過去 3 発話を用いた CMN では CMN を

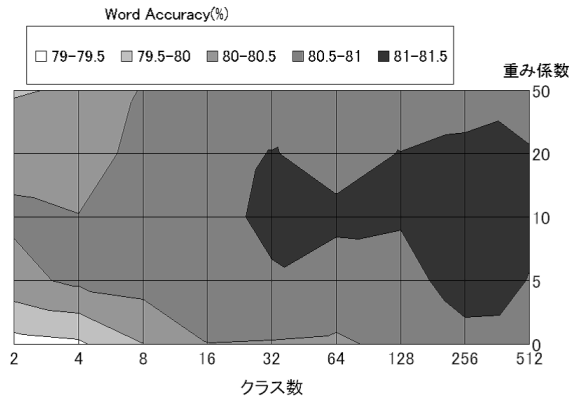


図 4 クラス数と重み係数をパラメータとした単語正解精度の分布  
 Fig.4 Distribution of word accuracy which parameter number of classes and weight coefficient

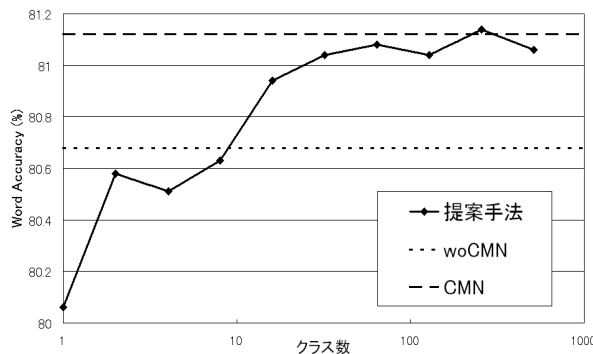


図 5 クラス数と単語正解精度  
 Fig.5 Relation between number of classes and word accuracy

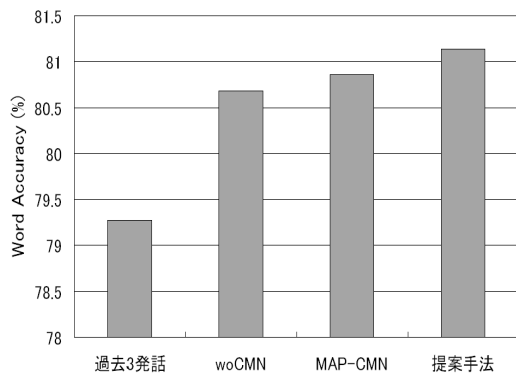


図 6 オンライン CMN の各種手法の性能比較  
 Fig.6 Recognition performance of online CMN methods

行わないものよりも性能が劣化している。これは、話者交替が頻繁に生じる公共の音声認識システムの特徴による影響だと考えられる。

## 7. ま と め

母音区間ケプストラムと発話単位 CM の対応関係を表すコードブックを作成し、コードブックを用いたフレーム同期 CMN を提案した。一般的なオフラインでの CMN に匹敵する性能を示した。なお、2 節で示したようにオフラインでは一般的な CMN よりも ML-E-CMN が優位であることがわかっている。この ML-E-CMN で用いられる CM の値と母音区間ケプストラムの対応関係を表すコードブックを作成することで、提案手法のさらなる性能改善が図れると考えられる。

謝辞 本研究は、文部科学省のリーディングプロジェクト「e-Society 基盤ソフトウェアの総合開発」によって行われたものである。

## 文 献

- [1] 西村竜一, 西原洋平, 鶴見玲典, 李 晃伸, 猿渡 洋, 鹿野清宏, “実環境研究プラットフォームとしての音声情報案内システムの運用,” 電子情報通信学会論文誌, Vol.J87-DII, No.3, pp.789-798, March 2004.
- [2] 追立真吾, 戸田智基, 李 晃伸, 川波弘道, 猿渡 洋, 鹿野清宏, “コードブックを用いた実時間処理 CMN の実装と評価,” SP2005-28, pp.7-12 (2005-06).
- [3] 庄境 誠, 中村 哲, 鹿野清宏, “ケプストラム平均正規化法と HMM 合成法に基づくモデル適応化法 E - CMN / PMC と自動車内音声認識への適用,” 電子情報通信学会論文誌, D-II, No.10, pp.2636-2644, 1997-10.
- [4] M.G.Rahim, B.-H. Juang, “Signal bias removal for robust telephone-based speech recognition in adverse environments,” Proc.Int.Conf.Acoust.Speech, Signal Process., April 1994, pp.445-448.
- [5] W. Huang, B. Rao, “Channel and Noise Compensation for Text Dependent Speaker Verification Over Telephone,” Proc. ICASSP, pp337-340, 1995.
- [6] Y. Minami, “Cepstrum bias adaptation for the switchboard database in unsupervised mode,” The Center for Language and Speech Processing, WS97.
- [7] 黒岩真吾, 内藤正樹, 山本誠一, “ケプストラム移動平均正規化を用いた電話音声の認識,” 音講論, 1-2-3, pp.5-6(1995-9).
- [8] 小坂哲夫, 山本寛樹, 山田雅章, 小森康弘, 八木沼津義, “MAP-CMS と高速 PMC を用いた即時環境適応化の検討,” 音講論, 3-6-8, pp.97-98 (1998-3).
- [9] B. Atal, “Effectiveness of Linear Prediction Characteristics of the Speech Wave for Automatic Speaker Identification and Verification,” Journal of the Acoustical Society of America, 55, 6, pp.1304-1312, June 1974.
- [10] A.Lee, T.Kawahawa, and K.Shikano, “Julius - an Open Source Real-Time normalsize Vocabulary Recognition Engine,” Eurospeech2001, pp.1691-1694, 2001.