

離散混合出力分布型HMMを用いた非定常雑音下の音声認識

小坂 哲夫[†] 加藤 正治[†] 好田 正紀[†]

[†] 山形大学工学部 〒992-8510 山形県米沢市城南4-3-16
E-mail: †{tkosaka,katoh,kohda}@yz.yamagata-u.ac.jp

あらまし 従来、定常的な雑音への対処としてスペクトルサブトラクションやPMCなど種々の手法が提案され一定の成果を収めている。一方非定常雑音はスペクトルの特徴のバリエーションが大きく、大量のデータを集めてモデル化しにくいという問題がある。そのため、未知雑音、非定常雑音への対処は重要な課題の一つとなっている。本研究では、離散混合出力分布型HMM(DMHMM)を用い、未知の非定常雑音が重畳した音声に対するロバストな音声認識手法を提案する。我々はこれまでMAP推定によるDMHMMを用いた雑音下音声認識の検討を行い、その有効性を示してきた。本研究では、missing feature theory(MFT)に基づくDMHMMの離散分布の補正法を提案する。未知の非定常雑音が重畳すると、音響モデルに学習されない特徴が現れるが、閾値を設けることによりDMHMMの離散分布の補正を行うことにより対処する。提案手法の有効性を示すため、従来法である混合連続分布型HMM(CHMM)およびCHMMへMFTを適用した手法の一つであるacoustic backing-off法と比較実験を行なった。その結果3手法の中で最良の結果が得られ、またベースラインと比較して26.0%の認識率の向上を得た。

キーワード 音声認識, 離散HMM, 雑音, MAP推定

Speech Recognition Under Non-Stationary Noise Conditions using Discrete-Mixture HMMs

Tetsuo KOSAKA[†], Masaharu KATOH[†], and Masaki KOHDA[†]

[†] Faculty of Engineering, Yamagata University 4-3-16 Jonan, Yonezawa-city, Yamagata, 992-8510 Japan
E-mail: †{tkosaka,katoh,kohda}@yz.yamagata-u.ac.jp

Abstract Several promising methods of robust speech recognition in stationary noise environments, e.g., spectral subtraction and PMC, have been proposed. Meanwhile, it is difficult to collect a large amount of non-stationary noise for model training because there are many types of non-stationary noise. As stated above, speech recognition under unknown and non-stationary noise environments is one of the most difficult issues. This paper introduces a new method of robust speech recognition under unknown and non-stationary noise environments based on discrete-mixture HMMs (DMHMMs). Recently, we applied the DMHMMs based on MAP estimation to speech recognition in noisy conditions, and found that they were effective for the conditions. In this paper, we propose a compensation method for discrete distributions of DMHMMs based on missing feature theory (MFT). It is more likely that acoustic outlier values appear in unknown and non-stationary noise environments. In order to avoid the influence of the values, discrete distributions are compensated by setting thresholds for discrete probability. The proposed method was compared with both conventional continuous-mixture HMMs and acoustic backing-off method. It showed the best performance, and obtained an average relative improvement of 26.0% over the baseline.

Key words speech recognition, discrete HMMs, noise, MAP estimation

1. まえがき

従来、定常的な雑音への対処としてスペクトルサブトラクション[1]やPMC[2]など種々の手法が提案され一定の成果を収めてきている。しかし実環境での音声認識を考えると、定常的な雑音よりも継続時間の短い環境音(突発性雑音)が問題となる場合が多い。例えば舌打ち、咳払い、ドアの開閉音などは、音声認識に重大な悪影響を与えることが知られている。このような非定常雑音は多様な音源からなり、系統的に大量のサンプルを収集するのは困難である。このため雑音の特徴を予め推定する必要がある手法、たとえば上記のスペクトルサブトラクションやPMCなどを適用するのは困難である。以上のように非定常雑音、未知雑音の問題を解決することが実環境での音声認識の性能向上に必要であるが、現状では有効な方法は見つからない。本研究ではこの非定常雑音、未知雑音の問題の解決を目的とする。

一般に音声認識においては、音響モデルとして混合連続分布型HMM(CHMM)が使用されている。このモデルでは正規分布を仮定することにより、効率よくパラメータの推定を行うことができる。離散分布型HMMの場合、量子化サイズを小さくすると量子化歪みが大きくなり、逆にサイズを大きくすると学習データが不足し、十分にパラメータ推定ができないという問題がある。これに対しCHMMでは、正規分布を仮定することにより、パラメータ数を抑えることができ、学習が行い易い。また正規分布から外れた分布に対しては、混合分布を用いることで対応している。しかし雑音が重畳した場合の音声や、話者のばらつきによっては混合分布でも対応が不可能である場合が想定される。このようなデータに対しては、正規分布の仮定のない、離散分布型HMMが有用ではないかと考えられる。

以上の理由から、我々はこれまでMAP推定による離散混合出力分布型HMMを用いた雑音下音声認識の検討を行い、その有効性を示してきた[3]。このモデルでは、離散分布型HMMの場合問題となるパラメータ推定誤りに対し、入力特徴ベクトルをサブベクトルに分割し量子化する方法[4]と、パラメータ推定の方法としてMAP推定[5]を用いることで対処している。このモデルをベースに本研究では非定常雑音下における音声認識の検討を行った。

非定常雑音への対処法としては、missing feature theory (MFT)に従った手法を提案する。MFTにおいては、音響特徴を雑音の少ない信頼できる区間と、雑音の多い信頼できない区間の2種類に分割し、信頼できない区間をデコードする際の悪影響が全体に及ばないような処理を行う。MFTに従った方法は主に2種類のアプローチが考えられる。一つは信頼できない区間のデータをなんらかの知識を用いて再生する方法であり、もう一つは信頼できない区間を無視するなどして、この区間の処理が他に悪影響を及ぼさないよう扱う方法である。本提案法は後者のアプローチを採用している。いずれのアプローチの場合も、信頼できない区間を検出することが必要であるが、正確に検出することは困難であり、MFTを適用する際の大きな問題点となっている。本研究では、以上の問題を解決するため離

散確率に閾値を設ける方法を提案する。この方法では閾値により自動的に確率の与え方を変えるため、信頼できない区間を明に検出する必要がない。よって上記の問題を回避することが可能である。

本報告では、突発性雑音を重畳した連続音声を用い、音声認識実験による評価を行った。また従来法であるCHMMおよび、CHMMへMFTを適用した手法の一つであるAcoustic backing-off法[6]と比較実験を行ない提案手法の有効性を示す。

2. 離散混合出力分布型HMMのMAP推定

2.1 離散混合出力分布型HMM

離散混合出力分布型HMM(DMHMM)のMAP推定について述べる前に、本節では文献[4]などで提案されているDMHMMについて概説する。

まず入力特徴ベクトル \mathbf{o}_t をサブベクトル数 S に分割し $\mathbf{o}_t = [\mathbf{o}_{t1}, \dots, \mathbf{o}_{tS}]$ とする。サブベクトルの分割法は、例えばケプストラムの1次と2次を1つのサブベクトルとするなど、特徴量の隣接するものを一まとめにするなどの方法をとる。次にコードブック作成用のデータを用いて、各サブベクトルごとコードブックを作成する。 $q_s(\mathbf{o}_{st})$ をサブベクトル s における、入力 \mathbf{o}_{st} に対する量子点とすると、入力 \mathbf{o}_t は以下のように量子化される。

$$q(\mathbf{o}_t) = [q_1(\mathbf{o}_{t1}), \dots, q_s(\mathbf{o}_{st}), \dots, q_S(\mathbf{o}_{tS})] \quad (1)$$

このときDMHMMの出力確率 $b_i(\mathbf{o}_t)$ は以下のように求められる。

$$b_i(\mathbf{o}_t) = \sum_m w_{im} \prod_s \hat{p}_{sim}(q_s(\mathbf{o}_{st})) \quad (2)$$

但し、

$$\sum_m w_{im} = 1.0 \quad (3)$$

ここで \hat{p}_{sim} はサブベクトル s 、状態 i 、混合要素 m における離散確率、 w_{im} は混合分布の重み係数である。上式は、混合内での異なるサブベクトル間の離散確率は互いに独立だが、状態内の異なるサブベクトル間の従属性は、混合要素でモデル化されるという仮定にもとづく。よって状態内でサブベクトル間の相関がないならば混合要素に分ける必要はなくなるが、相関があるならば、混合要素に分けた場合認識性能の向上が期待できる。文献[4]ではDMHMMにおいて混合数を増やした場合、認識率が向上することが示されており、状態内のサブベクトル間には何らかの相関があることが推察される。

2.2 離散出力確率分布のMAP推定

通常の最尤推定では事前分布を定数とし、事前分布の影響を無視するが、MAP推定では事前分布も考慮にいれたパラメータ推定を行う。この方法はCHMMにおける話者適応などに応用され[5]その有効性が示されている。

本節では、本研究で提案する離散混合出力分布型HMMのMAP推定について述べる。 k をコードブックのインデックス、 γ_{imt} を、時刻 t で状態 i 、混合要素 m に存在する確率、とする

と Baum-Welch アルゴリズムにより、離散出力確率の ML 推定値は以下のように求められる。

$$p_{sim}(k) = \frac{\sum_{t=1}^T \gamma_{imt} \delta(q_s(\mathbf{o}_{st}), k)}{\sum_{t=1}^T \gamma_{imt}} \quad (4)$$

$$\delta(q_s(\mathbf{o}_{st}), k) = \begin{cases} 1 & q_s(\mathbf{o}_{st}) = k \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

また事前分布を考慮した離散出力確率の MAP 推定値を以下のように求める。

$$n_{im} = \sum_{t=1}^T \gamma_{imt} \quad (6)$$

このとき、離散出力確率の MAP 推定値 $\hat{p}_{sim}(k)$ は事前分布をディレクレ分布とした場合、

$$\hat{p}_{sim}(k) = \frac{(\nu_{simk} - 1) + n_{im} \cdot p_{sim}(k)}{\sum_{k=1}^K (\nu_{simk} - 1) + n_{im}} \quad (7)$$

で求められる。ここで ν_{simk} は事前分布のパラメータである。

$$\nu_{simk} = \tau \cdot p_{sim}^0(k) + 1 \quad (8)$$

と仮定すると

$$\hat{p}_{sim}(k) = \frac{\tau \cdot p_{sim}^0(k) + n_{im} \cdot p_{sim}(k)}{\tau + n_{im}} \quad (9)$$

となる。ここで τ は事前知識の確からしさに関する係数であるが、今回の実験ではすべての離散確率に対して一定値を用いた。

離散混合出力分布型 HMM に関しては出力分布の平均値だけでなく、混合係数、状態遷移確率についても MAP 推定が可能であるが、これらのパラメータについては、事前分布として使用した CHMM の値をそのまま使用した。

2.3 事前分布のパラメータ設定法

MAP 推定では事前分布のパラメータをどのように定めるかが大きな問題となる。本研究ではパラメータの設定に、一般的に用いられ実績のある混合連続型モデルを利用する。しかし混合連続型モデルの場合、確率密度関数によって確率が与えられているため、そのままでは用いることはできない。そこで各量子点のセントロイドに対応する確率密度を計算して離散化を行う。

各サブベクトル s の各量子点 k におけるセントロイドを $\mathbf{v}_s(k)$ とし、混合連続分布 HMM の出力確率の確率密度関数を $b'_{sim}(\mathbf{v}_s(k))$ と表わすと、事前分布のパラメータ $p_{sim}^0(k)$ は以下のように与えられる。

$$p_{sim}^0(k) = \frac{b'_{sim}(\mathbf{v}_s(k))}{\sum_k b'_{sim}(\mathbf{v}_s(k))} \quad (10)$$

3. 離散確率の閾値の設定

継続時間の短い非定常雑音、例えば舌打ち、咳払い、ドアの開閉音などは、音声認識に悪影響を与えることが知られている。このような非定常雑音は種類も多く大量のサンプルを収集する

のは困難である。対処法としては大別して 1) あくまでサンプルを収集してモデル化を図る方法と、2) モデル化はせず未知雑音として扱う方法、が考えられる。前者の場合は少数サンプルによるパラメータ誤推定の問題があるが、これに対しては、クラスタリングにより特徴の似た雑音をまとめてモデル化するなどの対策が考えられる [8]。

後者の場合は missing feature theory に従い、雑音が乗った区間が他の区間に悪影響を与えないように扱う (例えば [6], [9])。本研究では後者の考えに従い、DMHMM の離散分布に閾値を設ける方法について検討した。

DMHMM において、上式 (2) の $\hat{p}_{sim}(q_s(\mathbf{o}_{st}))$ のいずれかのサブベクトルの確率が 0 またはそれに近い値になると、出力確率も 0 に極めて近い値になる。モデル化されていない未知雑音が重畳した場合、このような状況になりやすいと考えられる。影響を受けるサブベクトルは雑音の性質により、特定のサブベクトルの場合も、すべてのサブベクトルの場合もあり得る。そこで本研究では離散確率に一律に閾値を設け、確率が閾値を下回った場合、閾値に置き換えるという処理を行ない、未知雑音の音声認識への悪影響を除去する。この場合確率モデルの要件を満たすためには、閾値を加えた分、他の確率の値を減少させるという back-off 的な処理が必要であるが、今回は行っていない。閾値の設定は以下の 2 種類を検討し、さらに両者の組み合わせも検討した。

サブベクトルごとの閾値 式 (2) に示す出力確率を以下のように表現する。

$$b_i(\mathbf{o}_t) = \sum_m w_{im} \prod_s p'_{simt} \quad (11)$$

このとき

$$p'_{simt} = \begin{cases} \hat{p}_{sim}(q_s(\mathbf{o}_{st})) & \text{if } \hat{p}_{sim}(q_s(\mathbf{o}_{st})) \geq dth \\ dth & \text{else} \end{cases}$$

混合要素ごとの閾値 出力確率を

$$b_i(\mathbf{o}_t) = \sum_m w_{im} p''_{simt} \quad (12)$$

とすると、

$$p''_{simt} = \begin{cases} \prod_s \hat{p}_{sim}(q_s(\mathbf{o}_{st})) & \text{if } \prod_s \hat{p}_{sim}(q_s(\mathbf{o}_{st})) \geq mth \\ mth & \text{else} \end{cases}$$

サブベクトルごとの閾値を設定する方法では、特定のサブベクトルが影響を受けるときは、その影響を受けたサブベクトルのみを無視することができる。一方混合要素ごとの閾値による方法では、複数のサブベクトルが雑音の影響を受ける場合に有効である。

4. 認識実験条件

4.1 音声資料および分析条件

音声分析条件は表 2 の通りである。今後 ETSI で標準化された分散音声認識 [10] に応用する予定のため、標準化に合わせ c_0

表1 サブベクトルの量子化サイズ

パラメータ	$\log P,$ c_0	$c_1,$ c_2	$c_3,$ c_4	$c_5,$ c_6	$c_7,$ c_8	$c_9,$ c_{10}	$c_{11},$ c_{12}
量子化サイズ	256	64	64	64	64	64	64
パラメータ	$\Delta \log P,$ Δc_0	$\Delta c_1,$ Δc_2	$\Delta c_3,$ Δc_4	$\Delta c_5,$ Δc_6	$\Delta c_7,$ Δc_8	$\Delta c_9,$ Δc_{10}	$\Delta c_{11},$ Δc_{12}
量子化サイズ	256	64	64	64	64	64	64
パラメータ	$\Delta^2 \log P,$ $\Delta^2 c_0$	$\Delta^2 c_1,$ $\Delta^2 c_2$	$\Delta^2 c_3,$ $\Delta^2 c_4$	$\Delta^2 c_5,$ $\Delta^2 c_6$	$\Delta^2 c_7,$ $\Delta^2 c_8$	$\Delta^2 c_9,$ $\Delta^2 c_{10}$	$\Delta^2 c_{11},$ $\Delta^2 c_{12}$
量子化サイズ	256	64	64	64	64	64	64

表2 音声分析条件

標準化周波数	16kHz
量子化	16bit
フレーム長	32msec
分析周期	8msec
分析窓	ハミング窓
高域強調	$1 - z^{-1}$
特徴ベクトル	0~12次のLPCメルケプストラム係数 と対数パワー、および一次と二次の 回帰係数(計42次元)
正規化	発話毎のケプストラム平均正規化

表3 マルチコンディション学習用音声データ

種類	列車 (在来線)	人込み	1500ccクラス 自動車	展示会場 (通路)
∞	786 × 4 文			
20	786 文	786 文	786 文	786 文
15	787 文	787 文	787 文	787 文
10	787 文	787 文	787 文	787 文
5	787 文	787 文	787 文	787 文

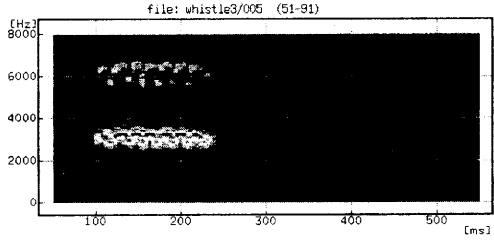


図1 雑音 whistle3 のスペクトログラム

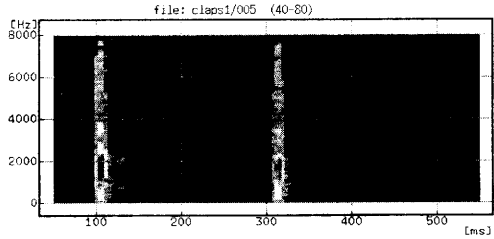


図2 雑音 claps1 のスペクトログラム

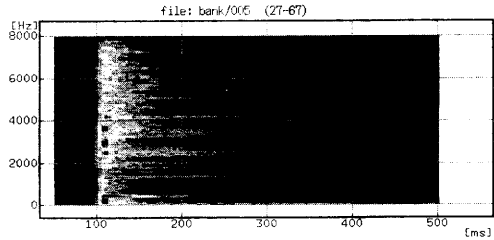


図3 雑音 bank のスペクトログラム

と対数パワーを併用している。ただし、標準化のパラメータはMFCCであり、完全には分析条件が揃っているわけではない。

音響モデルはマルチコンディション学習を行うため、音声データと雑音データを用意する。音声データは日本音響学会の新聞記事読み上げコーパス(ASJ-JNAS)を用いた。学習用およびコードブック作成用として、102名の男性が発声した新聞記事読み上げ文+音素バランス文、計15,732文を使用した。雑音データとしては電子協騒音データベースのうち「列車(在来線)」、「人込み」、「1500ccクラス自動車」、「展示会場(通路)」の4種類を選んで使用した。上記15,732文を20分割し、4種類の雑音×5種類のSNR(5,10,15,20,∞dB)の計20種類の組み合わせで雑音を人工的に重畳しマルチコンディション学習用音声データとした。このデータを表3に示す。

また評価用データとしては、男性10名の新聞読み上げ記事100文に、RWCP実環境音響音声データベース[11]収録の環境音を重畳することにより作成した。このデータベースでは環境音は特徴音、動作音、衝突音に分類されているが、それぞれ1つずつwhistle3(笛)、claps1(拍手複数回)、bank(貯金箱をたたく)を選択した。それぞれのスペクトログラムを図1、図2、図3に示す。重畳に当たっては、環境音を含む1secの区間を切り

出し、それを間隔を空けず繋げたものを音声に重畳した。SNR比は音声区間の平均パワーと雑音区間の最高パワーから計算し0dBとした。以上のように評価データには学習データで重畳された雑音は含まれておらず、雑音に対してオープンな実験と言える。

4.2 評価用認識システム

第1パスでtriphone HM-Net[12]および単語bigramを用いて単語グラフを生成し、第2パスで単語trigramを用いて単語グ

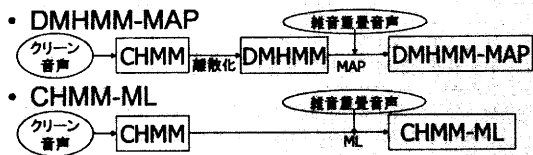


図4 実験の枠組み

ラフをリスコアする2-passデコーダを用いる[13]. 但し今回は第1パスの結果でのみ評価した. 言語モデルは語彙5kで毎日新聞の45ヶ月分を使用して作成した. 音響モデルは各triphone3~6状態, 総状態数2000, 1状態あたり16混合のサブベクトル量子化離散混合分布HM-Netを使用した. 比較として同混合数の混合連続分布HM-Netも使用した.

4.3 コードブックの設計

本研究の実験時のサブベクトルの割り当てを表1に示す. 合わせてコードブックサイズも示した. サブベクトルは基本的には隣接するケプストラム係数2つを1サブベクトルとし, 量子化サイズを64(=6bit)に設定した. また対数パワー($\log P$)と c_0 のペアについては量子化サイズを256(=8bit)とした. これらはETSIの分散音声認識の標準化を参考にした(但し最新の標準[10]では c_{11}, c_{12} のペアについては32(=5bit)に変更されている). ETSIの標準化ではデルタやデルタデルタパラメータは伝送の対象となっていないため規定はないが, DMHMMではこれらのパラメータも量子化する必要があるため, ケプストラム係数に準じて定めた. このためサブベクトル数は合計21となる. ただしデルタやデルタデルタに関しては, サイズ削減の検討も必要であると考えられる. 以上のコードブックの作成には雑音重畳前の学習用データを用い, LBGアルゴリズムを使用して求めた.

5. 実験結果および考察

まず雑音重畳のないデータから混合連続分布HMM(CHMM)をML推定により学習し, それを2.3節に述べた方法で離散化したモデル(DMHMM)を求める. このDMHMMを初期モデルとして雑音重畳データを使ってマルチコンディション学習をMAP推定により行う(DMHMM-MAP). 比較として上記CHMMを初期モデルとし, 学習法としてML推定を用いたマルチコンディション学習を行う(CHMM-ML). 予備実験よりMAPの係数を $\tau = 10.0$ と定めた. 以上の手順を図4に示す.

図5にサブベクトルごとに閾値を設定した場合, 図6に混合要素ごとに閾値を設定した場合, 図7に2種類の閾値を併用した場合のDMHMM-MAPの単語誤り率を示す. また比較としてCHMM-MLの場合の結果も示す.

まずサブベクトルごとに閾値を設定した場合についてみると, DMHMM-MAPにおいて全く閾値処理をしないと, 従来法のCHMM-MLに対し性能が下回る場合も上回る場合もあるが, 閾値を与えることにより大幅な改善が得られる. 最適な閾値は雑音の種類によって異なる. 雑音の種類によっては閾値を大きく

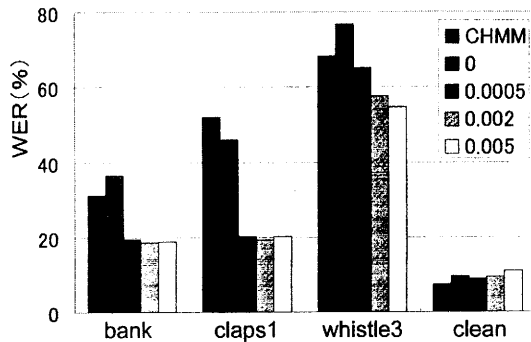


図5 サブベクトルごとに閾値を設定した場合

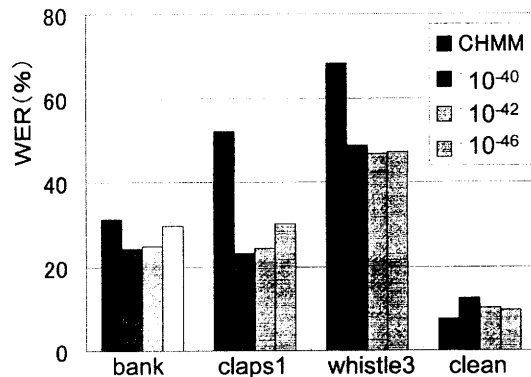


図6 混合要素ごとに閾値を設定した場合

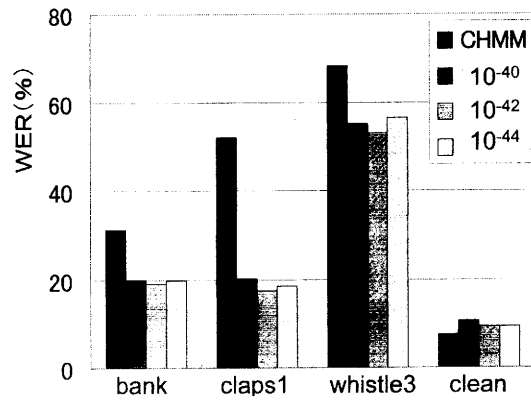


図7 2種類の閾値を併用した場合

する必要がるが, 一方閾値を大きくしすぎると雑音に乗っていない区間に対して悪影響を与える. 例えばwhistle3の場合はさらに閾値を上げると認識率の向上が見込まれるが, 他の雑音では逆に認識率の低下が予想される. 以上の問題を避けるには雑音区間を推定し, 推定区間と他の区間を別の閾値とする必要があると考えられる. また雑音の無い環境ではCHMM-MLが高

表4 提案法とAB法の単語誤り率(%)による比較

	bank	claps1	whistle3	clean	改善率
提案法	19.0	17.5	53.0	9.3	26.0
AB法	24.6	19.6	63.3	6.9	24.8
CHMM-ML	31.3	52.1	68.3	7.5	0.0

い性能を示すため、なんらかの対策が必要である。次に混合要素ごとの閾値の場合と比較すると、bankやclaps1については混合要素ごとの閾値の場合単語誤り率が増加するが、whistle3では逆に性能が向上する。このように閾値の種類によっても効果が異なる。閾値の種類によって効果が異なるため、2種類の閾値を併用することが有効と考えられる。図7に示すように、両者の閾値を併用すると雑音の種類によらず安定した結果が得られる。以上DMHMM-MAPに閾値を与えた場合の結果を示した。

一方連続分布型HMMに閾値を与える方法としてAcoustic backing-off法(AB法)が提案されている[6]。この手法では特徴次元ごとに、観測データの分布 $p(y)$ と非観測データの分布 $p_o(y)$ の混合分布から尤度計算を行う。 ϵ をバックオフ係数、 R_{max} 、 R_{min} をそれぞれ特徴次元ごとのデータの最大値と最小値とすると、以下のように求められる。

$$\log(p_{ab}(y)) = \log\{(1 - \epsilon)p(y) + \epsilon \cdot p_o(y)\} \quad (13)$$

$$p_o(y) = (R_{max} - R_{min})^{-1} \quad (14)$$

このAB法と提案手法の比較を行う。文献[6]では、 R_{max} と R_{min} を学習データの最大・最小値を用いているが、この場合 $p_o(y)$ の値が学習データに依存するため、比較実験においては $R_{max} = \mu + r \cdot \sigma^2$ 、 $R_{min} = \mu - r \cdot \sigma^2$ (但し μ は各特徴次元のデータの平均値、 σ^2 は分散)とし r の値もパラメータとして変化させて実験を行った。実験では $1 < r < 11$ 、 $5.0 \times 10^{-5} < \epsilon < 1.0 \times 10^{-2}$ とパラメータを変化させたところ、 $r=7.0$ 、 $\epsilon = 1.0 \times 10^{-4}$ で最良の結果を得た。この値とDMHMM-MAPの比較を表4に示す。表から分かるようにCHMM-MLからの改善率を比較すると提案法がAB法を上回り、本手法が有効であることが示された。特に3種類の雑音における改善率が大きい。一方雑音のない状況(clean)では認識率が低下し、雑音のない状況での改善が今後の課題として残った。

6. まとめ

MAP推定による離散混合出力分布型HMMを用いた非定常雑音下の音声認識の検討を行った。その結果、離散確率に閾値を設けることにより、認識性能の大幅な改善が得られることが分かった。また閾値はサブベクトルごとに設定する場合と、混合要素ごとに設定する場合とでは性能向上の傾向が異なるため、両者を併用したところ、雑音の種類によらず安定した性能が得られることが分かった。問題点としては雑音の種類によって最適な閾値が異なっている。対策としては非定常雑音に乗った区間を推定するなどが必要と考えられる。またサブベクトルごとの処理が可能というモデルの特性を充分生かし切れていないため、サブベクトルに重みをつけるなどの検討を行いたい。

文 献

- [1] S. Boll: "Suppression of Acoustic Noise Speech Using Spectral Subtraction," IEEE, ASSP-27, No.2, pp.113-120 (1979).
- [2] M.J.Gales and S.Young: "An Improved Approach to the Hidden Markov Model Decomposition of Speech and Noise," Proc. of ICASSP92, pp.233-236 (1992).
- [3] 小坂哲夫, 加藤正治, 好田正紀: "MAP推定を用いた離散混合出力分布型HMMの雑音重畳音声での評価," 信学技報, SP2003-21, pp. 7-12 (2003).
- [4] S. Tsakalidis, V. Digalakis and L. Newmeyer: "Efficient Speech Recognition Using Subvector Quantization and Discrete-Mixture HMMs," Proc. of ICASSP99, pp. 569-572 (1999).
- [5] C.-H. Lee and J.-L. Gauvain: "Speaker Adaptation Based on MAP Estimation of HMM Parameters," Proc. of ICASSP93, pp. 558-561 (1993)
- [6] J. de Veth, B. Cranen and L. Boves: "Acoustic backing-off as an implementation of missing feature theory," Speech Communication, Vol. 34, pp.247-265 (2001).
- [7] S. Takahashi, K. Aikawa and S. Sagayama: "Discrete Mixture HMM," Proc. of ICASSP97, pp. 971-974 (1997).
- [8] 井坂, 山田, 北脇, 浅野: "隠れマルコフ網と逐次状態分割法による環境音モデル化の検討," 日本音響学会講演論文集, 3-Q-13, pp.179-180 (2002.09).
- [9] 山本, 篠田, 嵯峨山: "ガウス分布の補正による突発雑音に頑健な音声認識," 信学技報 SP2002-45, pp.19-24 (2002.06).
- [10] ETSI ES 202 050 V1.1.1: "STQ;Distributed speech recognition; Advanced front-end feature extraction algorithm; Compression algorithms," ETSI standard (2002.10).
- [11] S. Nakamura, K. Hiyane, F. Asano, T. Nishimura, T. Yamada: "Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition," Proc. ICLRE, pp. 965-968 (2000).
- [12] 堀貴明, 加藤正治, 伊藤彰則, 好田正紀: "状態クラスタリングによるHM-Netの構造決定法の検討," 信学論D-II, J81-D-II, No.10, pp.2239-2248 (1998).
- [13] 堀貴明, 岡直生, 加藤正治, 伊藤彰則, 好田正紀: "大語彙連続音声認識のための音素グラフに基づく仮説制限法の検討," 情処論, Vol.40, No.4, pp.1365-1373 (1999.4).
- [14] D. Pearce and H.-G. Hirsch: "The AURORA Experimental Framework for the Performance Evaluation of Speech Recognition Systems under Noisy Conditions," Proc. of IC-SLP2000, Vol.4, pp.29-32 (2000).