社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2004－ＳＬＰ－51　（１）
2004／5／27

複数の雑音重畳モデルの合成による 複数雑音環境に頑健な音響モデルの検討

加藤 裕介† 鈴木 基之† 伊藤 彰則† 牧野 正三†

†東北大学大学院工学研究科 〒980-8579 宮城県仙台市青葉区荒巻字青葉 05

E-mail: †{yusuke,moto,aito,makino}@makino.ecei.tohoku.ac.jp

あらまし 複数雑音環境に頑健なモデルの作成法として，単一の雑音を重畳させた音声で学習したHMMを複数組み合わせる方法を提案する．複数のHMMを組み合わせる方法として，それぞれをマルチパスでつなぐ方法，またそれぞれのHMMを各状態別に分布を混合分布として統合する方法について検討する．提案した２つのモデルでは，従来法と同等かそれよりも良い認識精度が得られた．また複数のHMMを組み合わせるためモデル自体の規模は大きくなる．そこでモデルの規模を小さくするためモデル内の分布を分布間距離などを用いて統合することについても検討する．

キーワード HMM，マルチコンディションモデル，混合分布，マルチパスモデル，分布間距離

An HMM robust to multiple noise conditions by combining multiple noise-adapted HMMs

Yusuke KATO† Motoyuki SUZUKI† Akinori ITO† and Shozo MAKINO†

† Graduate School of Engineering, Tohoku University Aoba 05, Aramaki, Aoba-ku, Sendai 980-8579 Japan

E-mail: †{yusuke,moto,aito,makino}@makino.ecei.tohoku.ac.jp

Abstract This paper describes methods to compose an HMM robust under multiple noise conditions. The methods are based on combination of several HMMs trained under different noise conditions. We propose two combination methods. The first one combines multiple HMMs into a multi-path HMM. The second one combines corresponding states of each HMM into one state by mixing the output probability distributions onto one mixture distribution. The recognition experiment revealed that HMMs composed by the proposed methods shows similar or better results than conventional multi-condition model. One drawback of the model composed by the proposed methods is that it has large number of distributions. To reduce the number of distributions, we examined several methods to unify distributions.

Keyword HMM, multicondition model, Gaussian mixture PDF, multipath model, inter-distribution metric

1. はじめに

近年，音声認識技術の飛躍に伴い，音声認識システムの実用化が進められてきている．しかし現在の音声認識システムの多くは，実環境で環境雑音の影響を受けると大きく認識率が低下してしまう．

音声認識にはさまざまな過程があり，それぞれの過程で雑音対策がなされているが，本報告では音声認識の際に用いる音響モデルの雑音対策を検討する．音響モデルを用いた雑音対策として代表的な方法としてHMM合成法[1][2]がある．この方法はあらかじめ音声HMMと雑音HMMを用意しておき，それらを合成することで雑音に頑健なモデルを得る．しかし合成する際に近似的な処理を行わざるを得ないという問題点がある．そこで本報告では実際に雑音を重畳させた音声を学習データに用いて作成した雑音重畳モデルを用いる．

また実環境を考えた場合，出現する雑音は単一ではなく複数出現すると考えられ，複数の雑音に対して頑健な音声認識システムの構築を考えなければならない．現在，複数の雑音環境に頑健な音響モデルで最も効果のあるものとして，マルチコンディションモデルが提案されている．しかしマルチコンディションモデルは学習に要する時間が多大，また未知の雑音に対してもある程度頑健であるが，新たな雑音への対処の際，一からモデルを構築しなおさなければならないといった問題点がある．

そこで本報告では，新たな雑音への対処がマルチコンディションモデルよりもすばやく行えるように，あらかじめ雑音重畳モデルを作成しておき，それらを合成して1つのモデルを作成する方法を用いて複数の雑音環境に頑健な音響モデルを考える．

モデルの合成方法としては，喜嶋らによって提案されたSN比別マルチパスモデル[3][4]，鈴木らによって提

案された複数のHMM状態の分布を混合分布として統合する方法[5]の2つを用いる．本報告では，これらの方法における合成モデルを雑音重畳モデルにした，雑音重畳マルチパスモデル，雑音重畳マルチミクスチャーモデルを提案する．

2. マルチコンディションモデル

マルチコンディションモデルは雑音重畳モデルの1つである．雑音重畳モデルと異なる点は学習に使用する雑音データが複数であるという点である．

特徴としては，複数の雑音を学習に用いるため複数の雑音に頑健であるという点が挙げられる．また学習に用いなかった雑音に対しても性能は多少劣るがある程度頑健であるという特徴もある．

3. 提案法

3.1. 雑音重畳マルチパスモデル

一般の音声認識に用いるモデルでは，left-to-right型のHMMを用いてモデルを構築するが，雑音重畳マルチパスモデルの場合，各雑音別に雑音重畳モデルを作成し，それらを図1のように並列に並べ，初期状態からの状態遷移を各モデルに対して許すことによりモデルを構築する．雑音環境の異なるモデルを並列に並べているため，マルチコンディションモデルのように複数の雑音環境に頑健なモデルが構築できることが期待される．

またマルチコンディションモデルとは違い，随時新たな雑音に対応したモデルを追加できるという利点がある．

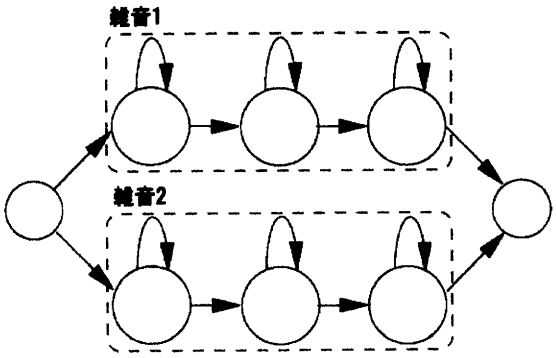


図 1:雑音重畳マルチパスモデルの例

3.2. 雑音重畳マルチミクスチャーモデル

雑音重畳マルチミクスチャーモデルは雑音重畳モデルの状態ごとの分布に重みをつけ，混合分布として結合させたモデルである．これを図2に示す．このとき分布の重み係数の総和は1である．雑音重畳マルチパスモデルと同様に，複数の雑音重畳モデルを用いて作成するため，複数の雑音環境に頑健なモデルが構築できることが期待される．また新たな雑音に対応したモデルも追加できる．

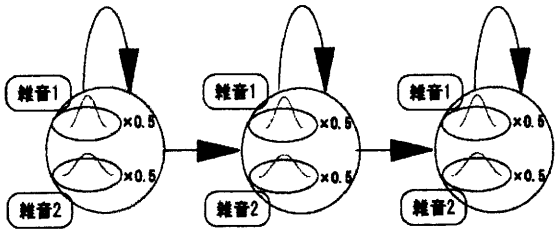


図2:雑音重畳マルチミクスチャーモデルの例

4. 認識実験

4.1. データベース

学習に使用する音声データとして，ATR研究用自然発話音声データベースから約27000文選択した．また雑音データベースは電子協騒音データベースから12種類選択した．SN比を10dBに固定して音声に雑音を重畳させ，学習データとした．

テストデータは同様の音声データベースから学習に用いていない音声を207文選択した．また重畳する雑音は既知の場合と未知の場合で用意した．雑音が既知の場合，学習に用いた雑音と同種類のものから学習に用いていない部分を選んだ．またSN比は10dBに固定した．雑音が未知の場合，雑音データベースから学習に用いていない4種類の雑音を選択し，既知の場合と同様に音声に重畳した．使用した雑音の種類を表1に示す．

表1:雑音の種類

学習に用いた雑音 (雑音が既知の場合)	展示会場(会場通路) 駅 機械工場 板金工場 脱荷シュート 幹線道路 交差点 人込み 列車 空調ファンコイル エレベータホール
学習に用いていない雑音 (雑音が未知の場合)	走行自動車内(2000cc):一般道 走行自動車内(1500cc):高速道 展示会場(ブース内) コーティングルーム

4.2. 音響モデル

本報告の提案法である雑音重畳マルチパスモデルと雑音重畳マルチミクスチャーモデルを各雑音重畳モデルをもとに作成した．また雑音重畳マルチパスモデルは初期状態からの遷移確率を等確率とし，雑音重畳マルチミクスチャーモデルの分布の結合重みもすべて等しくした．従来法であるマルチコンディションモデルも作成した．モデルの条件を表2に，状態数と総分布数を表3に示す．

4.3. 実験条件

認識エンジンとしてJuliusを用いて単語認識実験を行い，マルチコンディションモデル(MC)，雑音重畳マルチパスモデル(MP)，雑音重畳マルチミクスチャーモデル(MM)について比較した．また評価データの雑音と雑音重畳モデルの雑音が一致した場合(matched)についても比較を行った．

言語モデルは音響モデルの学習に用いた音声データと同じものを用いて，単語3-gramを作成した．語彙数は約7000語彙である．また音響分析条件を表4に示す．

表2:音響モデルの詳細

雑音重畳モデル	triphone 16混合 (雑音12種類)
雑音重畳マルチパスモデル	triphone 16混合×12パス
雑音重畳マルチミクスチャーモデル	triphone 192混合
マルチコンディションモデル	triphone {8,16,32,64,128}混合

表3:各モデルの状態数と総分布数

モデル	混合数	状態数	総分布数
マルチコンディションモデル	8	5485	43838
	16	5485	87670
	32	5485	175334
	64	5485	350662
	128	5485	701318
雑音重畳マルチパスモデル	16	37404	600038
雑音重畳マルチミクスチャーモデル	192	5620	1079028

表3:音響分析条件

サンプリング周波数	16[kHz]
特徴量	MFCC+ Δ MFFF(各 12 次元)+Δpow (計25次元)
分析窓	Hamming窓
フレーム長	25[ms]
フレームシフト	10[ms]
CMN	なし

4.4. 実験結果

雑音が既知の場合の実験結果を図3に，未知の場合の実験結果を図4に示す．雑音が既知の場合，従来法のマルチコンディションモデルは混合数16のとき最大の認識精度を示し，66.62[%]となった．雑音重畳マルチパスモデルではほぼマルチコンディションモデルと同等の66.92[%]，雑音重畳マルチミクスチャーモデルではマルチコンディションモデルと比べ約3[%]程高い70.00[%]の認識精度となった．

1つ目の提案法である雑音重畳マルチパスモデルでは各雑音重畳モデルを並列に並べるため，デコード時に入力の雑音とマッチした雑音重畳モデルのパスが最も尤度の高いパスとして選択されやすくなっていると考えられる．しかし完全にマッチしたパスを通るとは限らないため，誤ったパスを通ることもあり，matchedの場合よりも認識精度が低下したと考えられる．

また，2つ目の提案法である雑音重畳マルチミクスチャーモデルでは一度，各雑音別にモデルを構築してから分布を結合するため，マルチコンディションモデルと比べより各雑音に対して詳しい分布を保持できると考えられる．そのため認識精度が向上したと考えられる．

雑音が未知の場合，認識精度は既知の場合と同様の傾向を示し，マルチコンディションモデルは65.25[%]，雑音重畳マルチパスモデルでは65.50[%]，雑音重畳マルチミクスチャーモデルでは68.50[%]となった．各モデルは雑音が既知の場合と比べ若干認識精度が低下するものの，ほぼ変わらない認識精度を示した．このことより提案した2つのモデルでも，未知雑音に対してもさほど認識精度が低下しないことが分かる．

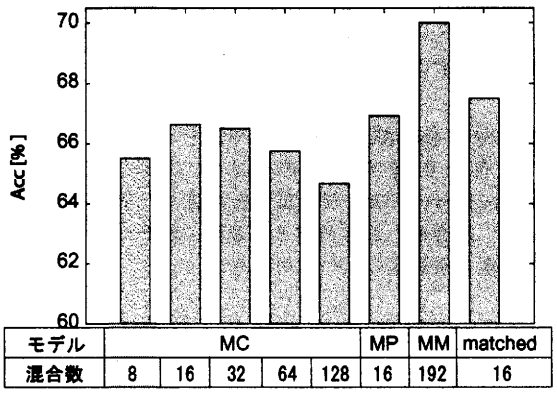
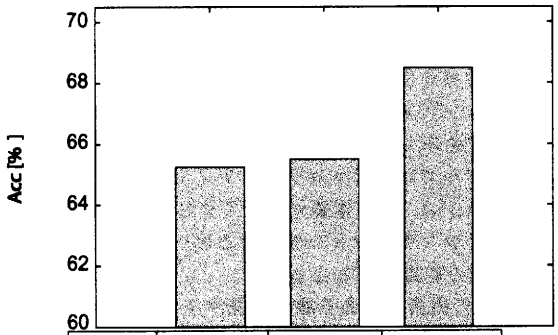


図3:雑音が既知の場合の認識結果



モデル	MC	MP	MM
混合数	16	16	192

図4:雑音が未知の場合の認識結果

5. 分布数の削減

提案した雑音重畳マルチミクスチャーモデルは最も良い認識精度を示したが，表3に示すようにマルチコンディションモデル中で最も良い認識精度を示した混合数16のものと比較するとかなり総分布数が多いことが分かる．そこで雑音重畳マルチミクスチャーモデルの分布数を削減することを考える．

雑音重畳マルチミクスチャーモデルでは先にも述べたように，各雑音重畳モデルを状態別に混合分布として統合している．そのため1つの状態に類似した分布が多数出現してくると考えられる．その類似した分布を統合することで分布数を削減し，認識精度を維持できないか検討する．

5.1. 削減方法

各状態において分布間距離など分布を削減するためのパラメータを設定し，それが最適になる分布を選択し統合する方法を検討する．

5.2. 分布を削減するためのパラメータ

分布を削減するためのパラメータとして，Bhattacharyya距離，混合重みを考慮したBhattacharyya距離，統合前の混合分布と統合後の混合分布との尤度の平均二乗誤差の3つを用いる．

5.2.1. Bhattacharyya 距離

Bhattacharyya距離は2つの分布間の距離の1つである．を表したものである．2つの正規分布間のBhattacharyya距離は次の式で表される．

$$BD(a,b)=\frac{1}{8}(\mu_a-\mu_b)^t\left(\frac{\Sigma_a+\Sigma_b}{2}\right)^{-1}(\mu_a-\mu_b)+\frac{1}{2}\log\left(\frac{\left|\frac{\Sigma_a+\Sigma_b}{2}\right|}{|\Sigma_a|^{\frac{1}{2}}|\Sigma_b|^{\frac{1}{2}}}\right) \quad (1)$$

ここで $BD(a,b)$は正規分布 a, b 間の Bhattacharyya 距離を表す．また μ_a, μ_b, Σ_a, Σ_bはそれれ正規分布 a, b の平均値ベクトル，対角共分散行列を表す．

HMM の各状態において，混合分布を構成する各正規分布間の距離を式(1)によって計算し，距離が最小となる分布から統合していく．

5.2.2. 混合重みを考慮した Bhattacharyya 距離

通常の Bhattacharyya 距離では分布間の距離だけを考えている．しかしモデルの各状態が保持している分布は混合正規分布であり，その混合正規分布の混合重みも重要なパラメータであると考えられる．そこでその混合重みを考慮した Bhattacharyya 距離を定義する．混合重みが大きい分布ほどその分布の持つ情報の重要性が大きくなると考え，より混合重みの小さい分布ほど統合されやすく，分布間距離が小さくなるように(2)式のように定義する．

$$BD'(a,b)=(w_a+w_b)BD(a,b) \quad (2)$$

ここで $BD'(a,b)$は正規分布 a,b の混合重みを考慮した Bhattacharyya 距離である．また w_a, w_bはそれぞれ分布 a,b の混合重みを表す．

この距離が最小となる分布から統合していく．

5.2.3. 尤度の平均二乗誤差

モデルの各状態の混合正規分布全体を考慮して分布数の削減を行うことを考える．統合前と統合後の混合分布どうしの分布間距離を計算するのは困難であるため，統合前と統合後の混合分布の尤度の平均二乗誤差を用いる．まず統合前，統合後の混合分布はそれぞれ(3)(4)式で表される．

$$\phi(x)=\sum_i w_i N(x,\mu_i,\Sigma_i)+w_a N(x,\mu_a,\Sigma_a)+w_b N(x,\mu_b,\Sigma_b) \quad (3)$$

$$\varphi(x)=\sum_i w_i N(x,\mu_i,\Sigma_i)$$
$$+(w_a+w_b)N(x,\mu_c,\Sigma_c) \quad (4)$$

また統合する２つの分布の平均値ベクトル，対角共分散行列，混合重みをそれぞれ μ_a, μ_b, Σ_a, Σ_b, w_a, w_b とし，統合した分布の平均値ベクトル，対角共分散行列を μ_c, Σ_c とする．

(3)(4)式をもとに尤度の平均二乗誤差 L は(5)式で表される．

$$L=\int_{-\infty}^{\infty}(\phi(x)-\varphi(x))^2dx$$
$$=\int_{-\infty}^{\infty}(w_aN(x,\mu_a,\Sigma_a)+w_bN(x,\mu_b,\Sigma_b)$$
$$-(w_a+w_b)N(x,\mu_c,\Sigma_c))^2dx$$
$$=w_a{}^2\prod_i\frac{1}{\sqrt{4\pi\sigma_a^{(i)}}}+w_b{}^2\prod_i\frac{1}{\sqrt{4\pi\sigma_b^{(i)}}}$$
$$+(w_a+w_b)^2\prod_i\frac{1}{\sqrt{4\pi\sigma_c^{(i)}}}$$
$$+2w_aw_b\prod_i\frac{\exp(-\frac{(\mu_a^{(i)}-\mu_b^{(i)})^2}{2(\sigma_a^{(i)}+\sigma_b^{(i)})})}{\sqrt{2\pi(\sigma_a^{(i)}+\sigma_b^{(i)})}}$$
$$-2w_a(w_a+w_b)\prod_i\frac{\exp(-\frac{(\mu_a^{(i)}-\mu_c^{(i)})^2}{2(\sigma_a^{(i)}+\sigma_c^{(i)})})}{\sqrt{2\pi(\sigma_a^{(i)}+\sigma_c^{(i)})}}$$
$$-2w_b(w_a+w_b)\prod_i\frac{\exp(-\frac{(\mu_b^{(i)}-\mu_c^{(i)})^2}{2(\sigma_b^{(i)}+\sigma_c^{(i)})})}{\sqrt{2\pi(\sigma_b^{(i)}+\sigma_c^{(i)})}} \quad (5)$$

ただし $\mu^{(i)}$, $\Sigma^{(i)}$ はそれぞれ平均，対角共分散行列の i 番目の要素である．

5.3. 分布の統合方法

分布の統合方法について述べる．統合前の２つの正規分布の平均値ベクトルをそれぞれ μ_a, μ_b, 対角共分散行列をそれぞれ Σ_a, Σ_b, 混合重みをそれぞれ w_a, w_b とすると，統合した正規分布の平均値ベクトル μ_c, 対角共分散行列 Σ_c はそれぞれ(6)(7)式のように表せる．

$$\mu_c=\frac{1}{w_a+w_b}(w_a\mu_a+w_b\mu_b) \quad (6)$$

$$\Sigma_c=\frac{1}{w_a+w_b}\left(w_a(\Sigma_a+\mu_a{}^t\mu_a)+w_b(\Sigma_b+\mu_b{}^t\mu_b)\right)$$
$$-\mu_c{}^t\mu_c \quad (7)$$

5.4. 分布を削減したモデルの認識実験

先の３つのパラメータを用いて分布数を削減したモデルについて認識実験を行い，認識精度と分布数の関係について調べた．またマルチコンディションモデルとの比較を行った．実験条件は先の認識実験と同様であり，この実験では評価データは雑音が既知の場合のみとした．ただしモデルの分布数は 1079028,100 万,80 万 60 万 40 万,20 万,87670 とした．

5.5. 実験結果

実験結果を図 5 に示す．どのパラメータを用いても，分布を減らすにつれて認識精度が低下してしまうことが分かる．しかし混合重みを考慮した Bhattacharyya 距離は他のパラメータを用いたときよりも認識精度の低下を抑えられた．やはり混合重みを考慮することで，混合重みの大きい重要な分布を統合しにくくできたため認識精度の低下を他のパラメータよりも抑えられたと考えられる．

また全体的に分布数を減らすと認識精度が低下するということから，雑音重畳マルチミクスチャーモデルは冗長な分布はあまりないと考えられる．

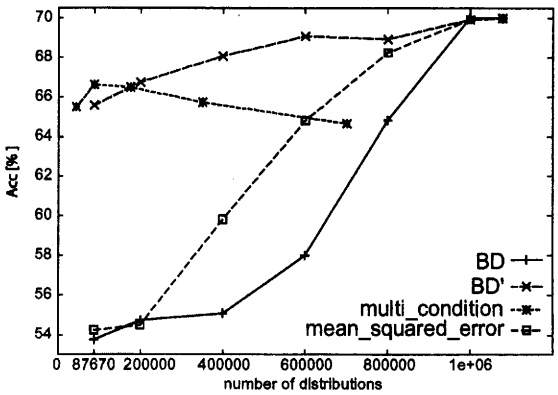


図 5:分布数と認識精度の関係

6. まとめ

複数の雑音環境からそれぞれ雑音重畳モデルを作成し，それらを並列に並べた雑音重畳マルチパスモデルと，各雑音重畳モデルの分布を混合分布として結合した雑音重畳マルチミクスチャーモデルを作成した．またそれらと従来法であるマルチコンディションモデルとで比較を行った．雑音が既知の場合と未知の場合の両方において，雑音重畳マルチパスモデルでは従来法と比較するとほぼ同等の認識精度となるが，雑音重畳マルチミクスチャーモデルの場合は従来法と比べ高い認識精度が得られた．またこれらの提案法は従来法とは違い，新たな雑音に対して比較的容易に対処できるという利点がある．

また雑音重畳マルチミクスチャーモデルの分布数を削減した．削減することで認識精度は低下するが，混合重みを考慮した Bhattacharyya 距離をパラメータとして用いることで，認識精度の低下はある程度抑えられた．

今後，さらに認識精度が低下しないパラメータについて検討する予定である．

文　献

[1] F.Martin，K.Shikano，Y.Minami and Y.Okabe，“Recognition of Noisy Speech by Composition of Hidden Markov Models”，信学技法，SP92-96，pp9-16，1992

[2] M.J.F.Gales，S.J.Young，“Robust Continuous Speech Recognition Using Parallel Model Combination”，IEEE Trans. Speech & Audio Processing，VOL.4，NO.5，pp.352-359，1996

[3] 喜嶋朋令，鈴木基之，伊藤彰則，牧野正三，“マルチパス音韻モデルを用いた非定常雑音に頑健な音声認識の検討”，音響学会 2002 年秋季講演論文集，pp33-34，Sep.2002

[4] 伊田政樹，中村哲，“雑音 GMM の適応化と SN 比別マルチパスモデルを用いた HMM 合成による高速な雑音環境適応化”，電子情報通信学会論文誌，D-Ⅱ.vol.J86-D-Ⅱ,pp-195-203，Feb.2003

[5] 鈴木浩之，全炳河，南角吉彦，宮島千代美，徳田恵一，北村正，“雑音をコンテキストとした音響モデルによる音声認識”，音響学会 2003 年秋季講演論文集，pp25-26，Sep.2003