

# 十分統計量を用いた教師なし話者適応および環境適応

松浪 加奈子<sup>†</sup>, 芳澤 伸一<sup>††</sup> 馬場 朗<sup>†††</sup>  
 李 晃伸<sup>†</sup> 猿渡 洋<sup>†</sup> 鹿野 清宏<sup>†</sup>

音声認識のための音韻モデル適応においては、すべての出現しうる雑音環境や話者に対してモデルをあらかじめ用意するのは困難であり、高精度な話者、環境適応技術が要求される。本研究では、対象話者の任意の1発声文と雑音データのみを入力として環境適応および話者適応を行う音韻モデルの適応アルゴリズムを提案する。本手法は3つのステップから成り立つ(1)雑音を音声データベースに重畳し、全話者の十分統計量を計算する(2)任意の1発声文に対してGMM話者モデルを用いて音声データベースから音響的特徴が近い話者集合を選択する(3)選択された話者集合の(1)の十分統計量を用いて話者適応および環境適応した音韻モデルを構築する。これにより任意の話者の環境発声に対する教師なし適応が可能である。認識実験より、提案法により適応した音韻モデルは、雑音を重畳した音声データベースからEMアルゴリズムを用いて作成した音韻モデルよりも高い認識精度を得られることが示された。さらにこのようにして適応したモデルを初期値として、さらにMLLRによる適応を行った場合、単にMLLRを行う場合よりも高精度な適応が行えることを示す。また、雑音レベルの変動の認識率への影響についても述べる。

## Unsupervised Speaker and Noise Adaptation Based on HMM Sufficient Statistics

KANAKO MATSUNAMI,<sup>†</sup> SHINICHI YOSHIZAWA,<sup>††</sup> AKIRA BABA,<sup>†††</sup>  
 AKINOBU LEE,<sup>†</sup> HIROSHI SARUWATARI<sup>†</sup> and KIYOHIRO SHIKANO<sup>†</sup>

In order to realize a robust and accurate speech recognition in various real environments, noise and speaker adaptation techniques are important, since covering all the possible noises and speakers in a static model is essentially impossible. In this paper, we propose an efficient unsupervised noise and speaker adaptation method based on HMM sufficient statistics. Our method consists of three steps. (1) Noise data of the target environment is added to the training speech database to calculate noise-added HMM sufficient statistics. (2) Given a target speaker, the speakers that are acoustically close to the target speaker are selected from the database, using speaker GMM. (3) Then, the resulting adaptation model is obtained by constructing models from the HMM sufficient statistics of the selected speakers. Recognition experiments show that our adapted models achieve higher accuracy than environment-matched models trained by the E-M algorithm, with only one utterance and environment noise data. Furthermore, we examine the robustness of the adapted models in different SNR conditions. Integration with supervised MLLR method is also investigated.

### 1. はじめに

実環境における大語彙連続音声認識においては、ユーザの発声の特徴に適應する話者適応技術および雑音環境に適應する環境適応技術が求められる<sup>1),6),7)</sup>。しか

し雑音には非常に多くの種類が存在し、それらすべての雑音を網羅するデータをあらかじめ持つことは不可能である。認識システムがおかれている環境の雑音に適應するには、その環境の雑音自体を用いることが望ましい。

また従来よく用いられる教師ありの話者適応および環境適応では、比較的高精度な適応モデルが得られるが、適応には数十文を必要とする。よって、音韻モデルを適応させるには認識を行う前に数十文章発声することが要求され、ユーザに大きな負担をかける。

以上のことから、少ない発声文章数や運用の場での環境雑音をもとに、その環境および話者に音韻モデル

<sup>†</sup> 奈良先端科学技術大学院大学  
Nara Institute of Science and Technology

<sup>††</sup> 松下電器産業株式会社  
Matsushita Electric Industrial Co., Ltd.

<sup>†††</sup> 松下電工株式会社  
Matsushita Electric Works, Ltd.  
現在、三洋電機株式会社  
Presently with SANYO Electric Co., Ltd.

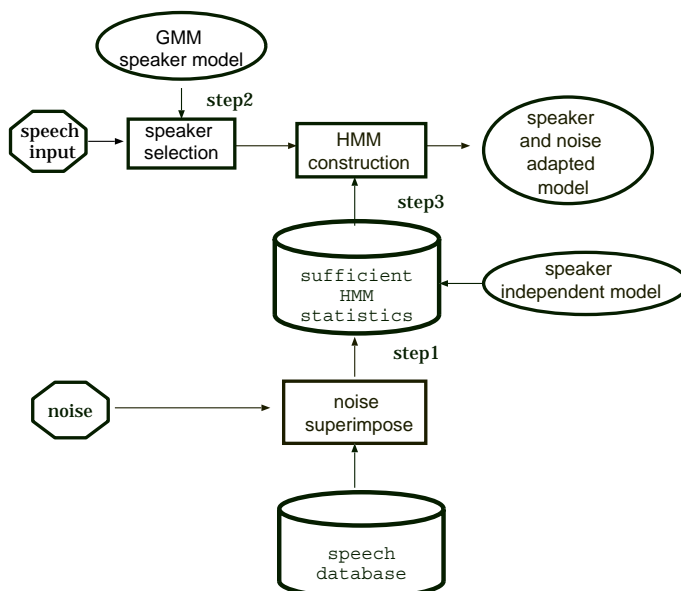


図1 話者適応および環境適応アルゴリズム  
Fig.1 Speaker and noise adaptation algorithm.

を適応させることが必要とされる。

本論文では、ある任意の1発声文と雑音データから話者適応および環境適応を行う高精度な適応アルゴリズムを提案する。音韻モデル学習用の clean 音声データベースに対して環境雑音を重畳するとともに、その重畳後の音声データから十分統計量を算出し、入力話者に対する類似話者を用いた教師なし話者適応を行う。十分統計量を用意することにより、話者に適応した音韻モデルを任意の1文から高速に学習できる。また雑音を重畳させた音声データを用いて十分統計量を計算することにより、実際おかれている環境に適応することができる。

また、実際の環境においては雑音レベルはつねに変動すると考えられる。したがって、本研究では適応後のモデルの雑音レベルの変動に対する頑健性についても評価する。さらに、高精度な適応を目指して、提案手法で適応したモデルを元に MLLR (Maximum Likelihood Linear Regression<sup>5)</sup>) を行う方法についても評価を行う。

以下、2章では提案する適応アルゴリズムについて述べ、3章では提案手法を用いて適応した音韻モデルの評価実験結果を示す。4章では雑音レベルの変動に対する頑健性の実験について述べ、5章では提案手法を用いて適応した音韻モデルを初期モデルとして MLLR<sup>5)</sup> を用いて適応した場合の認識性能を報告する。6章ではまとめを述べる。

## 2. 話者適応および環境適応アルゴリズム

提案する話者適応および環境適応アルゴリズムの概念図を図1に示す。提案法では、任意の1発話文と数十秒の雑音データを入力とする。提案法の特長として2つあげることができる。まず第1に、適応に用いる発話文が任意の1文章であることから入力話者にかかる負担を最小限におさえることができるということである。第2に、認識システム運用の場の環境雑音データを用いて適応を行うので、あらゆる環境雑音に適応することが可能となることである。

提案法における適応手順は以下のとおりである。

- (1) 音声データベースの全話者について雑音重畳データから十分統計量を計算して保存する。
- (2) GMM 話者モデルを用いてテスト話者に音響的特徴が近い話者を  $N$  人選択する。
- (3) 選択された  $N$  人の話者の十分統計量を用いて適応モデルを再構築する。

以上で話者適応および環境適応された音韻モデルが作成される。

また、十分統計量を用いた環境適応法として、MLLR と話者選択を用いた教師なし環境適応<sup>7)</sup>がある。これは clean の音声データを用いて十分統計量をあらかじめ作成しておき、入力話者に音響的特徴が近い話者を選択し話者適応した後、選択された話者の clean の音声データに環境雑音データを重畳し MLLR を用いて

環境適応を行う適応法である．この適応法は任意の 1 発声文・数十秒の環境雑音データで適応を行うことができるが，話者適応を行ったのちに環境適応を行うため，会議室などの雑音の変化が少なくかつ話者の交替が頻繁に起こりうるような環境において異なる話者に適応する場合には，十分統計量による話者適応と 50 文程度の MLLR をオンラインで行うことが必要であり，個々の話者への適応に時間がかかる．これに対して，提案法では環境雑音データを重畳した音声データベースを用いて十分統計量を作成することで環境適応を事前に行う．このため，あらかじめ収集した環境雑音データから十分統計量をオフラインで作成しておくことで，その十分統計量から環境に適応した話者適応モデルを話者ごとに高速に生成することができ，話者の交替に対しても高速に適応することができる．

以下，提案する適応アルゴリズムについて各工程の詳細を述べる．

### 2.1 十分統計量の計算

十分統計量とは，データベースの統計的性質を表すのに十分な統計量のことである．特に HMM においては，各正規分布の平均，分散，EM カウントのことである<sup>8)</sup>．この十分統計量を用いることにより，高速に音韻モデルを学習することができる．本研究では音声データベースとして JNAS<sup>2)</sup> を用いる．これは Clean 環境で収録された読み上げ音声データであり，これに入力雑音を重畳して十分統計量を計算する．十分統計量は，EM アルゴリズムにより不特定話者モデルから 1 回学習を行うことにより作成される．

### 2.2 話者選択

話者選択では任意の 1 発声文と GMM ( Gaussian mixture model ) 話者モデルを用い，テスト話者に音響的特徴が近い話者を音声データベースから選択する．GMM は混合正規分布で表現した話者ごとの出力確率モデルである．また各話者はそれぞれ 1 状態 64 混合で表現されている．話者選択用 GMM は，入力音声の中で雑音レベルの低い部分のみを用いて話者選択を行うことで clean データで作成した GMM をそのまま用いることとする<sup>7)</sup>．具体的には，入力音声の全フレームからパワーが低い半数のフレームを削除した特徴量ファイルを入力ファイルとして話者選択を行う．音響的特徴が近い話者を選択する際の距離尺度として

は尤度を用い，尤度上位 N 人を選択話者として選択する．以前の実験<sup>9)</sup> より，提案法の選択手法は，入力雑音を重畳したデータを用いて作成した GMM による話者選択手法とほぼ同等の性能であることが確認されている．

### 2.3 適応モデルの再構築

2.2 節で選択された N 人の話者の十分統計量を用いて，話者および環境に適応したモデルを再構築する．

## 3. 認識実験による評価

提案した話者適応および環境適応アルゴリズムの評価実験を行った．オフィス環境の居室雑音を対象とした適応実験を行い，提案手法を通常的环境 matched model ( EM アルゴリズムで学習したモデル ) と比較する．また従来法として，MLLR と話者選択を用いた教師なし環境適応アルゴリズム<sup>7)</sup> を用いて適応した音韻モデルの実験結果との比較を行う．

### 3.1 実験条件

音声データベースは日本音響学会による JNAS<sup>2)</sup> の音声データベースを用いる．JNAS データベースは男性 153 人，女性 153 人の計 306 人の話者で構成されており，各話者ごとに 50 文の音素バランス文と約 100 文の新聞読み上げ文を持つ．今回の実験では 306 人中男性 130 人，女性 130 人の計 260 人を学習用話者とし，残る男性 23 人，女性 23 人の計 46 人を評価用話者として用いた．

学習データ数は各話者につき音素バランス文 50 文，新聞読み上げ文 100 文である．評価用データとしては各話者につき 4, 5 文章，計 200 文章を用いる．サンプリング条件は 16 kHz, 16 bit, 特徴量は窓シフト長 10 ms で分析した 12 次元の MFCC ( Mel-frequency cepstrum coefficient ) とデルタ MFCC, デルタパワーを用いる．言語モデルは語彙数 20 k の新聞記事から構築した 3-gram を使用し，デコーダは Julius を用いる．音韻モデルは 43 音素 3 状態 16 混合のモノフォンモデルと 2000 状態 64 混合の PTM ( phonetic tied mixture model )<sup>4)</sup> を使用する．各音韻モデルは不特定話者モデルである．

選択話者数はモノフォンモデルの場合で 20 人，PTM の場合で 40 人とする．これらは clean 環境下での話者選択実験において最適な話者選択数であったためこの値を用いることとした．

話者選択で用いた GMM 話者モデルは，1 人あたり 1 状態 64 混合で構成される．GMM 話者モデルの構築には，各話者につき約 150 文章の音声データ ( 雑音重畳なし ) を用いる．環境適応対象は，居室雑音下

オフラインでの十分統計量の作成には PTM で 1 人あたり約 20 分を要した．また十分統計量のサイズはモノフォンモデルで 114 Mbyte ( 0.44 Mbyte × 260 人 )，PTM ではその約 10 倍であった．十分統計量はハードディスクに蓄積しておけばよいので，話者適応に大きな主記憶量を必要とするものではない．

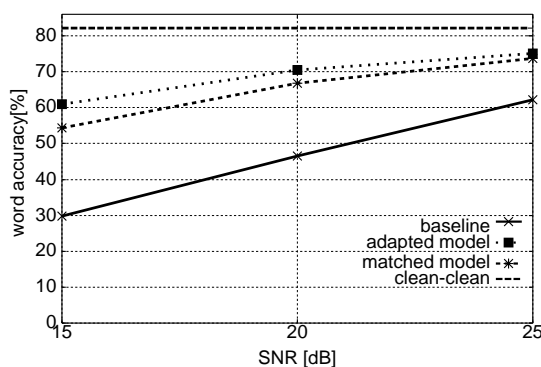


図 2 モノフォンモデルでの適応結果

Fig.2 Adaptation effect on monophone.

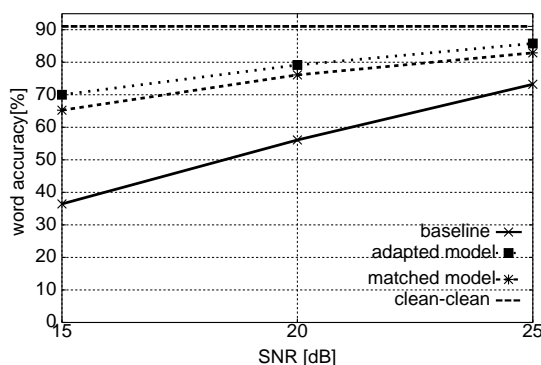


図 3 PTM での適応結果

Fig.3 Adaptation effect on PTM.

での発声であり、マイク収録した 3 分間の居室雑音を評価用データに重畳して入力とする。SNR は 15 dB, 20 dB, 25 dB とする。

### 3.2 実験結果

モノフォンモデルと PTM でのそれぞれの適応後モデルの認識率を図 2 および図 3 に示す。それぞれのグラフにおいて、縦軸は単語正解精度(%)、横軸は SNR(dB)である。baseline は Clean モデル(適応なし・不特定話者モデル)での認識率、adapted model は提案法を用いて話者適応および環境適応した音韻モデルでの認識率、matched model は雑音を重畳した音声データベースより EM アルゴリズムを用いて作成した不特定話者音韻モデルの認識率となっている。また clean-clean は Clean モデルを用いて Clean データを認識した結果である。

実験結果より、adapted model と matched model を比較すると、モノフォンモデルでは 15 dB の場合では約 6%、25 dB では約 1.5% adapted model のほうが高い認識率が得られた。また PTM においては 15 dB の場合では約 4.5%、25 dB では約 3% adapted

表 1 従来法<sup>7)</sup>との比較Table 1 Comparison between propose method and conventional method<sup>7)</sup>.

monophone model		
SNR	matched model (260)	提案法
15 dB	54.4(57.0)	61.0(63.1)
20 dB	66.8(69.0)	70.5(72.2)
25 dB	73.6(75.7)	75.1(77.5)
SNR	matched model (305)	従来法
15 dB	55.3(58.0)	54.2(56.3)
20 dB	67.1(68.9)	67.0(68.8)
25 dB	74.1(76.1)	76.4(78.4)
PTM		
SNR	matched model (260)	提案法
15 dB	65.4(67.6)	70.1(72.2)
20 dB	76.2(77.2)	79.3(80.7)
25 dB	83.0(84.8)	85.9(87.1)
SNR	matched model (305)	従来法
15 dB	69.3(71.2)	63.4(66.4)
20 dB	76.2(78.2)	77.5(79.2)
25 dB	84.6(86.3)	84.1(85.7)

(注) 従来法では学習話者が音声データベースから評価話者を除いた 305 人を用いた。値は単語正解精度(単語正解率)。

model のほうが高い認識率が得られた。このことより、提案法では matched model よりも高い認識率を得られることが分かる。また、雑音レベルが高いほど提案法を用いた適応モデルの方が改善幅が高い。

モノフォンモデルおよび PTM のそれぞれの音韻モデルにおいて、ほぼ同様の改善傾向が見られた。SNR が 15 dB のとき、baseline (Clean) での認識率に対して adapted model では約 30%、matched model では約 25%改善された。また SNR が 25 dB のときは、baseline (Clean) での認識率に対して adapted model では約 12%、matched model では約 10%改善された。

また、従来法<sup>7)</sup>を用いて適応した音韻モデルと、提案法を用いて適応した音韻モデルとの認識率を比較した結果を表 1 に示す。値は単語正解精度である。ただし、従来法では学習用話者は評価用話者を除いた 305 人としており、本論文での matched model および提案法の学習用話者数より多くなっている。

従来法では matched model と同程度の認識率であったが、提案法では学習話者数が少ないにもかかわらず従来法を大きく上回る認識率が得られた。

提案法は、EM アルゴリズムを用いて話者・環境適応を行うことと同等であるため、正確に適応モデルを学習することができ、高い認識率が獲得できたのだと考えられる。それに対して従来法は、環境適応にお

MLLR のクラス数は文献 7) と同じ条件である評価話者平均約 37 個であり、クラス数としては十分な数と考えられる。

表 2 異なる SNR における認識性能 (monophone)

Table 2 Result in various SNR conditions (monophone).

		adapted monophone model			
		SNR	15 dB	20 dB	25 dB
Test data	15 dB	61.0	59.4	59.6	
	20 dB	69.4	70.5	70.5	
	25 dB	74.0	76.0	75.1	
		matched monophone model			
		SNR	15 dB	20 dB	25 dB
Test data	15 dB	54.4	52.7	47.8	
	20 dB	66.0	66.8	63.5	
	25 dB	72.1	73.2	73.6	

表 3 異なる SNR における認識性能 (PTM)

Table 3 Result in various SNR conditions (PTM).

		adapted PTM			
		SNR	15 dB	20 dB	25 dB
Test data	15 dB	70.0	68.2	65.5	
	20 dB	79.7	79.3	77.0	
	25 dB	84.5	85.9	85.9	
		matched PTM			
		SNR	15 dB	20 dB	25 dB
Test data	15 dB	65.4	61.6	58.3	
	20 dB	76.2	76.2	72.8	
	25 dB	81.5	82.6	83.0	

いて MLLR を用いているためパラメータの線形変換による適応という制約が入るため、認識率の改善が低いと考えられる。

#### 4. 異なる SNR に対する頑健性

SNR はユーザや環境によってさまざまに変動する。そのため、評価用データの SNR と異なる SNR の適応モデルを用いても、ある程度の認識性能が確保できることが求められる。そこで、異なる SNR に対する音韻モデルの頑健性を評価するため、前節と同様の条件のもと評価用データの SNR と異なる SNR の音韻モデルを用いた認識実験を行った。音韻モデルとしては、提案法を用いて話者適応および環境適応した adapted model と、雑音を重畳した音声データベースより EM アルゴリズムを用いて作成した matched model を用いる。SNR は音韻モデルおよび評価用データともに 15 dB, 20 dB, 25 dB である。

##### 4.1 実験結果

モノフォンモデルと PTM でのそれぞれの認識実験結果を表 2, 表 3, 図 4, および図 5 に示す。値は単語正解精度 (%) である。グラフの縦軸は単語正解精度 (%), 横軸は SNR (dB) である。adapted は提案法で適応した音韻モデルの実験結果であり、それぞれ評価用データの SNR が 15 dB, 20 dB, 25 dB のときの結

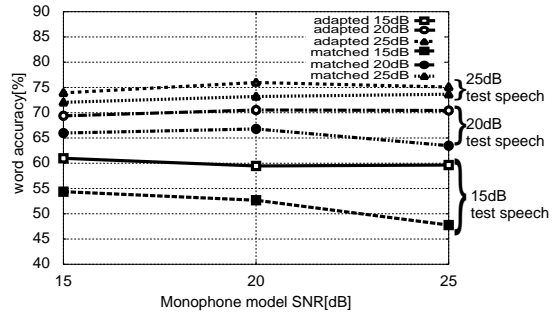


図 4 異なる SNR における認識性能 (monophone)

Fig. 4 Result in various SNR conditions (monophone).

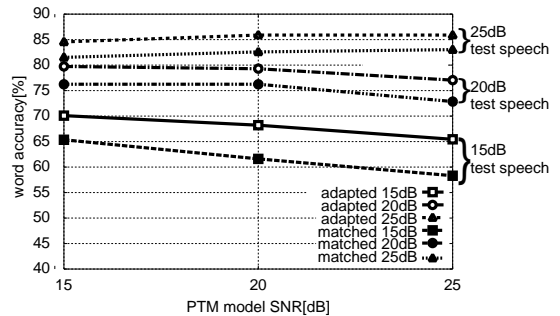


図 5 異なる SNR における認識性能 (PTM)

Fig. 5 Result in various SNR conditions (PTM).

果となっている。同様に matched は matched model の実験結果であり、それぞれ評価用データの SNR が 15 dB, 20 dB, 25 dB のときの結果となっている。

adapted model の場合、15 dB のテストデータを 15 dB の PTM の音韻モデルで認識したときの認識率は 70.0%であったのに対して、25 dB の音韻モデルで認識したときの認識率は 65.5%であり、4.5%の認識率の低下にとどまった。また 25 dB の評価用データを 15 dB の音韻モデルで認識したときの認識率は 84.5%であったのに対して、25 dB の音韻モデルで認識したときの認識率は 85.9%であり、SNR の違いによる認識率の低下はほとんど見られなかった。以上の結果より適応後のモデルは同種の雑音であれば雑音レベルの変動に対して頑健に認識することができることが分かる。

matched model の場合も同様に、15 dB の評価用データを 15 dB の音韻モデルで認識したときの認識率と 25 dB の音韻モデルで認識したときの認識率を比較すると約 7%認識率の低下が見られた。また、25 dB の評価用データを 25 dB の音韻モデルで認識したときの認識率と 15 dB の音韻モデルで認識したときの認識率を比較すると約 1%認識率の低下にとどまった。以

上のことより matched model, adapted model, どちらの音韻モデルにおいても同様に雑音レベルに対して頑健に認識できることが分かった。また、モノフォンモデルおよび PTM のどちらにおいても同様の結果が得られた。

以上の結果より、提案手法は SNR の変動にも頑健で、一度環境適応を行えば十分に環境に適応できることが分かる。また、提案法では話者適応を高速に行うことができ、同じ雑音下であれば話者の変動に対して迅速に対応できるので、提案法は有用であるといえる。

5. MLLR の初期モデルとしての有効性

現在最も一般的な教師ありの話者適応・環境適応法は、MLLR (Maximum Likelihood Linear Regression<sup>5)</sup>) である。MLLR は、適応データに対する尤度を最大化するように線形行列を求め、その行列を用いて HMM の平均および分散ベクトルを変換することにより、音韻モデルを適応させるアルゴリズムである。この MLLR は教師あり適応であり、あらかじめ決められた発話内容を確実に読み上げる必要がある。本研究の提案手法は、教示の必要がない教師なしの適応手法であるが、個人に特化した音声認識システムのような教示が可能なアプリケーションにおいては、さらにこの MLLR を行うことで、さらなる精度の向上が期待できる。また MLLR の観点からは、その適応モデルの認識率は初期モデルに依存することが知られている<sup>1),6),8)</sup>。

提案法で適応した音韻モデルが MLLR の初期モデルとして有効であるかを検証するため、adapted model と matched model を初期モデルとして MLLR を用いて適応したモデルの評価を行った。MLLR で適応する際に用いる適応文章は、各話者につき JNAS データベース上の各話者の全発話文から評価用文章を除いた 10 文および 50 文を使用した。

モノフォンモデルと PTM でのそれぞれの認識実験結果を表 4, 表 5 および図 6 に示す。値は単語正解精度 (%) である。グラフの縦軸は単語正解精度 (%), 横軸は適応文章数である。adapted は SNR が 15 dB, 20 dB, 25 dB の各データについて提案手法で適応したモデルを初期モデルとしたときの実験結果である。同様に matched は matched model の実験結果でそれぞれ SNR が 15 dB, 20 dB, 25 dB のときの結果である。また、表中の clean はそれぞれの clean モデルに MLLR 適応した結果である。

adapted model と matched model を比べると、MLLR 適応前の認識率がそれぞれ 15 dB の場合、モノ

表 4 MLLR の初期モデルとしての有効性の評価 (monophone)  
Table 4 Initial model effects on MLLR (monophone).

SNR	adapted monophone model		
	before	10 utterances	50 utterances
15 dB	61.0	64.5	68.2
20 dB	70.5	73.0	76.5
25 dB	75.1	79.4	82.7
clean	85.4	86.7	86.9
SNR	matched monophone model		
	before	10 utterances	50 utterances
15 dB	54.4	60.0	65.4
20 dB	66.8	71.4	74.3
25 dB	73.6	77.6	80.6
clean	82.2	84.1	87.6

表 5 MLLR の初期モデルとしての有効性の評価 (PTM)  
Table 5 Initial model effects on MLLR (PTM).

SNR	adapted PTM		
	before	10 utterances	50 utterances
15 dB	70.1	72.4	74.2
20 dB	79.3	81.5	83.2
25 dB	85.9	88.5	89.5
clean	93.1	93.5	94.5
SNR	matched PTM		
	before	10 utterances	50 utterances
15 dB	65.4	69.2	71.0
20 dB	76.2	79.3	81.6
25 dB	83.0	86.8	88.7
clean	91.1	91.9	93.2

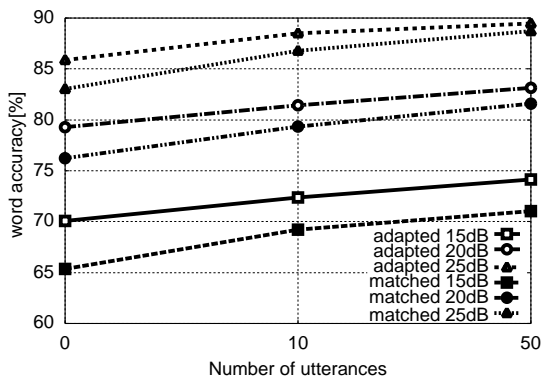


図 6 MLLR の初期モデルとしての有効性の評価 (PTM)  
Fig. 6 Initial model effects on MLLR (PTM).

フォンモデルでは各々 61.0%, 54.4% であったのに対し、50 文章で MLLR 適応した結果は 68.2%, 65.4% となっており、初期モデルとして認識率が高かった adapted model のほうが約 3% 高い認識率を示した。また、25 dB の場合も約 2% 高い認識率を示した。

PTM でも同様に、15 dB の場合は約 3%, 25 dB の場合も約 1%, matched model に比べて adapted model のほうがより高い認識率を示した。

これらのことから、提案法で話者適応および環境適

応した適応モデルは MLLR の初期モデルとして有効であり, MLLR を行うことでさらに精度が向上することが確かめられた。

## 6. おわりに

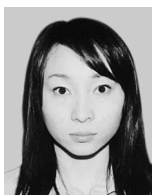
本論文では, 十分統計量を用いた話者適応および環境適応アルゴリズムを提案した。提案法では, ある任意の 1 発声文と数十秒の環境雑音を用いて適応を行うことができる。また, 認識実験を行うことにより, 提案法を用いて適応した音韻モデルが, 雑音重畳した音声データから EM アルゴリズムを用いて作成した音韻モデルよりも高い認識率を得られることを示した。また評価用データと SNR が異なる音韻モデルを用いても, ある程度の認識性能を確保できることを示した。最後に, 現在最も一般的な適応法である MLLR の初期モデルとしての有効性を示した。

## 参 考 文 献

- 1) Gao, Y., Padmanabhan, M. and Picheny, M.: Speaker Adaptation Based on Pre-Clustering Training Speakers, *Proc. EuroSpeech*, pp.2091–2094 (1999).
- 2) Itou, K., Yamamoto, M., Takeda, K., Takezawa, T., Matsuoka, T., Kobayashi, T., Shikano, K. and Itahashi, S.: JNAS: Japanese Speech Corpus for Large Vocabulary Continuous Speech Recognition Research, *The Journal of the Acoustical Society of Japan (E)*, Vol.20, pp.199–206 (1999).
- 3) Kawahara, T., Lee, A., Kobayashi, T., Takeda, K., Minematsu, N., Sagayama, S., Itou, A., Ito, K., Yamamoto, M., Yamada, A., Utsuro, T. and Shikano, K.: Free Software Toolkit for Japanese Large Vocabulary Continuous Speech Recognition, *Proc. ICSLP*, Ob(16)-V-07, pp.IV-476–479 (2000).
- 4) Lee, A., Kawahara, T., Takeda, K. and Shikano, K.: A New Phonetic Tied Mixture Model for Efficient Decoding, *Proc. ICASSP*, pp.1269–1272 (2000).
- 5) Leggetter, C.J. and Woodland, C.: Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density Hidden Markov Models, *Computer Speech and Language*, Vol.9, pp.171–185 (1995).
- 6) Padmanabhan, M., Bahal, L.R., Nahamoo, D. and Picheny, M.A.: Speaker Clustering and Transformation for Speaker Adaptation in Large-Vocabulary Speech Recognition System, *Proc. ICASSP*, pp.701–704 (1995).
- 7) Yamada, M., Baba, A., Yoshizawa, S., Mera, Y., Lee, A., Saruwatari, H. and Shikano, K.: Unsupervised Noisy Environment Adaptation Algorithm Using MLLR and Speaker Selection, *Proc. EuroSpeech*, pp.869–872 (2001).
- 8) Yoshizawa, S., Baba, A., Matsunami, K., Mera, Y., Yamada, M. and Shikano, K.: Unsupervised Speaker Adaptation Based on Sufficient HMM Statistics of Selected Speakers, *Proc. ICASSP*, pp.341–344 (2001).
- 9) 山田, 馬場, 芳澤, 米良, 李, 猿渡, 鹿野: 環境雑音適応アルゴリズムの大語彙連続音声認識による評価, 情報処理学会研究報告, SLP-35-6, pp.31–36 (2001).

(平成 13 年 11 月 19 日受付)

(平成 14 年 4 月 16 日採録)



松浪加奈子

平成 12 年近畿大学工学部電子工学科卒業。平成 14 年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了。同年三洋電機(株)入社。日本音響学会会員。



芳澤 伸一

平成 6 年名古屋大学工学部電気工学科卒業。平成 8 年同大学大学院修士課程修了。平成 11 年同大学院博士課程修了。在学中, ニューラルネットワークによるクラスタリングに関する研究に従事。平成 11 年松下電器産業(株)入社。平成 12~13 年(財)イメージ情報科学研究所に外向, 現在に至る。工学博士。主として音声・音情報処理の研究に従事。電気学会, 計測自動制御学会, 音響学会, 電子情報通信学会各会員。



馬場 朗

平成 6 年九州工業大学工学部電気工学科卒業。平成 8 年九州大学大学院総合理工学研究科博士前期課程修了。同年松下電工(株)入社。平成 12~13 年(財)イメージ情報科学研究所に外向。現在音声認識の研究開発に従事。音響学会, 電子情報通信学会各会員。

**李 晃伸 (正会員)**

平成 8 年京都大学工学部情報工学科卒業。平成 10 年同大学大学院修士課程修了。平成 12 年同大学院情報学研究科博士課程修了。同年より奈良先端科学技術大学院大学情報科学研究科助手。主として音声認識・理解の研究に従事。博士(情報学)。平成 14 年日本音響学会栗屋潔学術奨励賞受賞。日本音響学会, 電子情報通信学会各会員。

**猿渡 洋**

平成 3 年名古屋大学工学部電気工学科卒業。平成 5 年同大学大学院修士課程修了。平成 12 年同大学院博士課程修了。平成 5 年セコム(株)入社。セコム IS 研究所音声情報処理研究室において, 超音波アレー信号処理に関する研究に従事。平成 12 年奈良先端科学技術大学院大学助教授。音響アレー信号処理, ブラインド処理, 音場再生等に関する研究に従事。工学博士。平成 13 年電子情報通信学会論文賞受賞。電子情報通信学会, 日本音響学会, IEEE 各会員。

**鹿野 清宏 (正会員)**

昭和 45 年名古屋大学工学部電気工学科卒業。昭和 47 年同大学大学院修士課程修了。同年電電公社武蔵野電気通信研究所入所。昭和 59~61 年カーネギーメロン大学客員研究員。昭和 61 年~平成 2 年 ATR 自動翻訳電話研究所音声情報処理研究室長。平成 4 年 NTT ヒューマンインタフェース研究所首席研究員。平成 6 年より奈良先端科学技術大学院大学情報科学研究科教授。音情報処理学講座を担当。工学博士。主として音声・音情報処理の研究および研究指導に従事。昭和 50 年電子通信学会米沢賞, 平成 3 年 IEEE SP 1990 Senior Award, 平成 6 年日本音響学会技術開発賞, 平成 12 年情報処理学会山下記念研究賞, 平成 13 年 VR 学会論文賞。IEEE, ISCA, 音響学会, 電子情報通信学会, VR 学会各会員。