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

2003－ＳＬＰ－49　(17)
2003／12／18

大人・子供に適応した音声情報案内のためのユーザ自動識別

西村　竜一†　中村　敬介†　李　晃伸†　猿渡　洋†　鹿野　清宏†

† 奈良先端科学技術大学院大学 情報科学研究科　〒630–0192 奈良県生駒市高山町 8916–5
E-mail: †{ryuich-n,keisu-na,ri,sawatari,shikano}@is.aist-nara.ac.jp

あらまし　本報告では，音声インタフェースにおけるユーザ年齢層に応じた柔軟な対話処理の実現を目指して，話者の大人・子供識別手法を検討する．これまでの大人ユーザをターゲットとする音声認識では子供発話の認識は困難であった．しかし，家庭や公共施設への音声インタフェースの導入を考えると子供の存在は無視できない．子供発話を扱うための音声認識と音声インタフェースの改良が求められる．提案手法では，大人・子供に適応した音声情報案内の実装に必要な話者識別手段として，音声認識結果の対数尤度から求める音響的特徴と言語的特徴を併用した統計学習に基づく識別手法を実装する．二値分類アルゴリズムである SVM（Support Vector Machine）を識別に用いた実験では 91.8%の識別率を得た．これは音響的特徴のみを含む GMM（Gaussian Mixture Model）の尤度比較を使った識別結果から 5.4%の識別率改善である．本研究ではフィールドテストをすすめている生駒市コミュニティセンターの音声情報案内システム「たけまるくん」をプラットホームとしており，実験にはそのフィールドテスト収集発話を用いた．また，子供収集発話を音声認識モデル構築に含めることで子供認識精度の向上を試みており，その結果も報告する．
キーワード　公共型音声情報案内システム，子供発話認識，大人・子供ユーザ識別，音声認識スコア

Adult and Child Discrimination for Flexible Spoken Guidance System

Ryuichi NISIMURA†, Keisuke NAKAMURA†, Akinobu LEE†,

Hiroshi SARUWATARI†, and Kiyohiro SHIKANO†

† Graduate School of Information Science, Nara Institute of Science and Technology
8916–5 Takayama-cho, Ikoma-shi, Nara, 630–0192, JAPAN
E-mail: †{ryuich-n,keisu-na,ri,sawatari,shikano}@is.aist-nara.ac.jp

Abstract　This paper describes necessities of flexible spoken dialogues to both adult and child users. The conventional speech recognition program, which is developed on adult utterances, can not recognize child utterances correctly. It becomes impossible to disregard the increase of child users when the system is installed in a home or a public place. To realize the flexibility according to the user's age group, an automatic approach discriminating speakers between adult and child users is necessary. We propose a novel discrimination method on the basis of a statistical learning. As for parameter vectors in the algorithm, acoustic and linguistic properties extracted from speech recognition logarithm likelihood are adopted. Although GMM-based recognition uses only acoustic properties, this method can also consider linguistic properties. In the experiments with the SVM-based screening, we obtained 91.8% discrimination rate to the actual users' utterances. 5.4% improvement is shown as comparison with the GMM-based recognition. Our research platform "Takemaru-kun" system is a real world spoken guidance system located at the Ikoma-city Community Center. The system aims at a long-term field test of a speech interface and collecting actual users' utterance. To improve child speech recognition precisions, collected utterances are applied in training recognition models. Evaluation results of child speech recognition accuracy are also described in this paper.
Key words　Public spoken guidance system, Child speech recognition, Adult and child discrimination, Speech recognition scores

1. はじめに

音声インタフェースが次世代のマンマシンインタフェースとして注目されている．その中核を成すのが近年大きな進歩を遂げた大語彙連続音声認識である．しかし，現在の大語彙連続音声認識において高い認識精度を得るには認識タスクに適した音声認識モデル（音響モデルおよび言語モデル）が必要である．例えば，従来のほとんどの音声インタフェースシステムは大人をユーザの対象として設計されてきた [1]．それらシステムが持つ大人音声から作成されたモデルベースの音声認識では，子供の声を正しく認識することは困難である [2] [3]．しかし，家庭や公共施設などに音声インタフェースが導入されることを考えると，ユーザに子供も多く含まれるのは当然であり，その数は無視できない．今後，子供にも対応した音声認識モデルの導入が求められる．

また，対話状況によって変化する認識タスクをシステムが正確に把握し，状況に応じて使用する音声認識モデルを切り換えることで，はじめて高精度認識は可能となる．ユーザに応じた柔軟な対話戦略を実現する一つのアプローチにユーザモデルがある．ユーザモデルは，ユーザの持つシステムに対する習熟度，タスクドメインに関する習熟度，性急度などの性質をモデル化したものであり，不特定ユーザ向けの音声対話では，その性質に応じた対話戦略の決定に用いられてきた [4]．このようにユーザモデルは応答精度向上をもたらす．同様に個々のユーザの性質を音声認識モデルの選択に反映できれば認識精度の向上も期待できる．

本研究の目的は，大人・子供の両ユーザに柔軟な順応力を備えた音声インタフェースの開発である．そこで，ユーザモデルのアプローチを大人・子供別の音声認識モデルと対話戦略の導入によって音声インタフェース上に実装し，主に子供ユーザの音声インタフェースに対する利便性向上を目指す．本報告では，その実現に必要となる音声認識スコアをパラメータとする統計に基づいた大人・子供の話者識別手法を提案し，評価する．また，フィールドテスト収集発話を用いた子供音声認識性能の改善についても報告する．

2. 音声情報案内システム「たけまるくん」

本研究は，著者らが開発した生駒市北コミュニティセンターの音声情報案内システム「たけまるくん」[5] [6] をプラットホームとして実装を予定している（図 1）．たけまるくんは館内や周辺情報などの案内システムであり，現在，同センターにおいて長期間にわたるフィールドテストを実施，ユーザの発話を収集している．なお，本研究で使用するデータは 2002 年 11 月から 2003 年 3 月までのセンター開館日 125 日間のフィールドテストで収集した実際のユーザ発話である（計 46,754 個）．本研究では，その中から雑音などが比較的少なく明瞭，さらに音声のみからの人の主観によって話者の年齢層が判別できた 25,498 発話を用いる．発話の分類結果を表 1 に示す．使用した年齢層クラスは，a) 幼児，b) 低学年子供（小学校 3 年生ぐらいまで），c) 高学年子供（中学生ぐらいまで），d) 成人，e) 高齢者の 5 クラスである．本研究では，a) ～ c) を子供と見なす．この表の 68.2%が子供による発話であり，子供発話を扱うことの重要性がわかる．

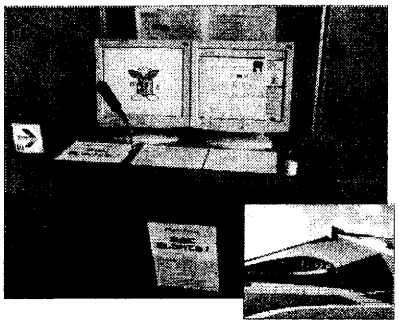

図 1 音声情報案内システム「たけまるくん」
Fig. 1 Speech-oriented guidance system "Takemaru-kun".

表 1 収集データの年齢層と性別ごとの分類結果
Table 1 Age and gender classification of collected data.

年齢層		男性	女性	性別不明	合計
a)	幼児	76	1421	585	2082
b)	低学年子供	1920	7961	2843	12724
c)	高学年子供	934	1154	498	2586
d)	大人	5520	2496	70	8086
e)	高齢者	8	12	0	20
合計		8458	13044	3996	25498

図 2 は，収集発話の発話内容をトピックごとに分類した結果である．図中の "Guidance" はセンターや周辺などの問合わせ，"Takemaru" はたけまるくんエージェントに関する質問，"Greeting" は挨拶，"News&Time" はニュース，天気予報や時間の問合わせ，"Other" はその他の発話が占める割合を示している．"Unclear" は叫び声などの不明瞭発話である．この図より大人と子供で発話内容の傾向に違いがあることが確認できる．大人はシステム本来の目的である "Guidance" の発話が多い．一方，子供はエージェントに関心を示し，名前や年齢などのたけまるくんのキャラクタ設定に関する質問が多い．

この結果，大人と子供ではタスクドメインに対する知識レベルが異なっており，音声情報案内システムを利用する際の興味の対象も異なることが明らかになった．つまり，大人・子供別の対話戦略を適用し，これら発話傾向の違いを考慮することで，ユーザに適した応答が実現できる可能性が高い．また，発話傾向の違いは発話中の使用単語や言い回し表現などコンテクストにも影響し，大人・子供の識別における言語的特徴の有用性を示唆している．

3. 子供発話の大語彙連続音声認識

従来の大人をターゲットとした音声認識では，子供発話に対して十分な精度を得ることができなかった．しかし，子供発話から，子供用に音声認識モデル（音響モデルと言語モデル）を学習することで，ある程度の精度回復を見込める．この確認の

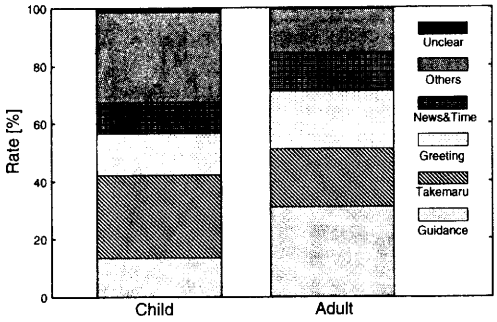

図 2　年齢層ごとの発話トピックの割合

Fig. 2　Topic population for different age groups.

表 2　音響モデル学習データ

Table 2　Training data for acoustic modeling.

大人	収集発話（男女，大人） 4,197 文
子供	収集発話（男女，子供） 15,922 文
JNAS	JNAS 読み上げ音声（男女） 40,086 文

表 3　言語モデル学習データ

Table 3　Training data for language modeling.

大人	収集発話書き起こし（大人） 7,606 文，単語数 40k，異なり単語数 1.7k
子供	収集発話書き起こし（子供） 16,892 文，単語数 89k，異なり単語数 3.4k
Web	生駒市関連 Web ページテキスト [5] 1,080,272 文，単語数 31,265k，異なり単語数 218.7k
QA	人手で収集した想定質問文テキスト [5] 6,488 文，単語数 56k，異なり単語数 3.2k

ために Julius [7] による大語彙音声認識実験を行った．

まず，大人・子供の認識対象ごとに音響モデル（PTM [8] triphone HMM，性別非依存）とバックオフ単語 3-gram 言語モデルを収集発話から作成した．

3.1　音響モデル

音響モデルの学習データの諸元を表 2 に示す．フィールドテストの収集発話は音韻バランスが取れていない．そのため，必要なデータを確保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] も学習に含めた．大人モデルの学習には JNAS と大人収集発話，子供モデルには JNAS と子供収集発話を用いた．

さらに，大人・子供別の依存性を高めるため，MAP（Maximum *A Posteriori*）推定 [10] もしくは MLLR（Maximum Likelihood Linear Regression）適応 [11] による適応音響モデルも作成した．適応元モデルは作成した大人・子供別音響モデルである．各々に大人・子供の収集発話（表 2）で適応し，依存性を高めている．収集発話には収録系の雑音なども含まれているので，環境適応も同時に施されたと考えることができる．

3.2　言語モデル

言語モデルの学習に用いたテキストデータの諸元を表 3 に示す．まず，Web と QA テキストからベースとなる言語モデルを作成する（語彙数: 40k）．次に大人・子供の書き起こしテキストから 2 つの言語モデルを作成し，年齢層依存モデルとする．年齢層依存モデルを先ほど作成したベース言語モデルにモデル融合ツール [12] を用いて重み 0.8 で融合し，大人・子供各々の年齢層適応モデルとする．最後に，モデルの高精度化のため，別途人手で作成した異なり単語数 441 のネットワーク文法を適用，年齢層適応モデルが持つ N-gram 確率を強化した [5]．

元気ですか？
高山サイエンスプラザはどこですか？
市役所はどこにありますか？
バスの時刻を教えてください。
図書館の利用時間教えてください。

図 3　大人用テストセットの例

Fig. 3　Example utterances of adult test set data.

あなたは誰？
図書館はどこですか？
何ができんの？
男の子ですか？女の子ですか？
んなとこ聞いてない。
トイレはどこでしゅか？

図 4　子供用テストセットの例

Fig. 4　Example utterances of child test set data.

3.3　テストセット

テストセットには学習に含まない大人 500（男 280，女 220），子供 500（男 149，女 351）の収集発話を用いた．これら発話は，発話内容の偏りを防ぐため，書き起こしテキストをキーにしてソートを行い，同じ発話内容のデータを間引いて除いた後（大人 2,833 個，子供 5,599 個）から，その発話内容の出現頻度に基づいて上位 500 個のデータを選んだ．テストセットに含まれる発話内容の例を図 3 および図 4 に示す．

3.4　音声認識実験

大語彙連続音声認識実験の結果として単語正解率を表 4 に示す．大人・子供発話ともに各々対応した音声認識モデルを用いることで最も高い認識率を得た．大人の音声認識モデルで子供発話を認識すると，最も良い MAP 適応音響モデルの場合でも 64.0%であり，既存の大人モデルベースのシステムで子供発話の認識が困難であることを再確認した．子供モデルを用いることで 18.5%（MAP 適応音響モデル使用時）の向上を得た．

音響モデルの適応手法間の比較に関しては，MAP 適応モデルが高い精度を示している．また，適応なし音響モデルの子供モデルと比べ，MAP および MLLR 適応後は大人発話に対して認識率低下，子供発話には認識率向上を得ている．つまり，適応することで子供発話への依存性を高めることができた．

この結果，話者年齢層ごとの音声認識モデルの選択は認識精度向上に有用であると言える．

表 4　年齢層別音声認識モデルを用いた単語正解率 (%)
Table 4　Word correct rate with age group dependent models. (%)

音声認識モデルタイプ	音響モデルの適応	テストセット	
		大人 500 発話	子供 500 発話
大人モデル	適応なし	92.5	(62.7)
	MAP	93.0	(64.0)
	MLLR	93.1	(62.2)
子供モデル	適応なし	(88.1)	72.8
	MAP	(82.6)	82.5
	MLLR	(83.7)	80.8

4. 音声認識スコアに基づく話者年齢層識別

ここからは提案手法である音声認識スコアに基づく大人・子供の話者識別手法について述べる.

提案手法では，音声認識結果から得られる音響的特徴および言語的特徴をパラメータとする話者識別を行う．従来，話者の認識および識別には GMM（Gaussian Mixture Model; 混合正規分布）に基づく尤度比較が用いられてきた [13] [14]．その識別には発話の持つ音響的特徴のみが用いられており，言語的特徴は考慮されなかった．これは発話のコンテクストに依存しない任意発話で識別できる方が利便性は高いと判断されたためである．これまで実験に提供されてきたデータは読み上げ音声であり，発話内容にユーザの性質が反映されていなかったこともその理由である．しかし，2 節で述べたように音声インタフェースに対する発話内容には大人と子供で異なる傾向が含まれる．このため，言語的特徴も考慮することで識別精度の向上が見込まれる．また，本手法は，大人・子供の発話傾向の違いを考慮しても発話のコンテクストを限定するものでなく，利便性が失われることもない．

本手法では，音響的特徴（Acoustic Property; 以下，AP）として Julius の第 2 パスが出力する音声認識スコアのフレーム平均音響対数尤度，言語的特徴（Linguistic Property; 以下，LP）として単語平均言語対数尤度を利用する．

$$AP = \frac{\text{音響対数尤度}}{\text{入力音声のフレーム数}} \tag{1}$$

$$LP = \frac{\text{言語対数尤度}}{\text{出力単語列の単語数}} \tag{2}$$

また，音声認識モデルや収録系に変更があってもその違いを吸収できるように，音響的特徴，言語的特徴は各々，大人音声認識モデルを用いた音声認識結果から求めたもの（AP_{adult}, LP_{adult}）から子供音声認識モデルによるもの（AP_{child}, LP_{child}）を引いた差を用いる．

図 5 は，適応なし音響モデル使用時の音響的特徴 $AP_{adult} - AP_{child}$ の頻度分布である．認識に使用した大人，子供の各音響モデルおよび言語モデルは 3 節作成のものである．各グラフは，子供女性，子供男性，大人女性，大人男性の表 1 の収集発話を入力した際のものである．大人男性の分布は明確に区別されるが，大人女性に関しては子供の分布と重なる部分が多い．

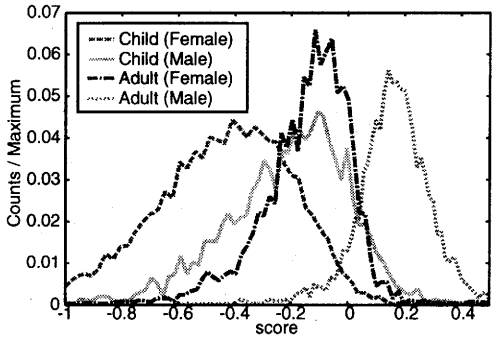

図 5　$AP_{adult} - AP_{child}$ の頻度分布 （適応なし音響モデル使用）
Fig. 5　Distributions of $AP_{adult} - AP_{child}$ without acoustic model adaptation.

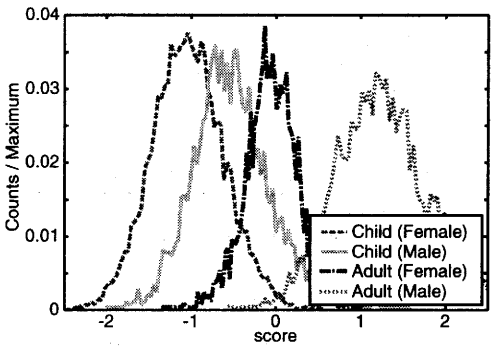

図 6　$AP_{adult} - AP_{child}$ の頻度分布 （MAP 推定音響モデル使用）
Fig. 6　Distributions of $AP_{adult} - AP_{child}$ using MAP adapted acoustic models.

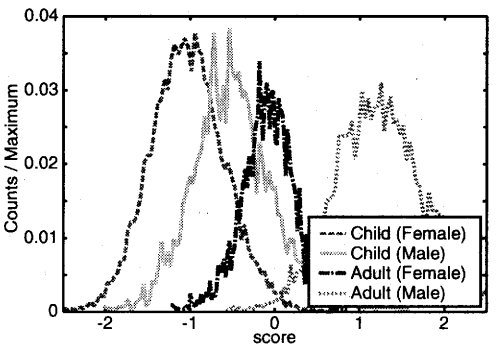

図 7　$AP_{adult} - AP_{child}$ の頻度分布 （MLLR 適応音響モデル使用）
Fig. 7　Distributions of $AP_{adult} - AP_{child}$ using MLLR adapted acoustic models.

これは大人女性の音響的特徴が子供に比較的似ているために生じた結果と思われる．次に MAP もしくは MLLR 適応した音響モデル使用時の結果を見る（図 6，図 7）．大人・子供各々に適応され，音響モデル学習に用いた JNAS 大人（男女）の特徴が薄らぎ，大人女性の分布を子供から分離することができた．

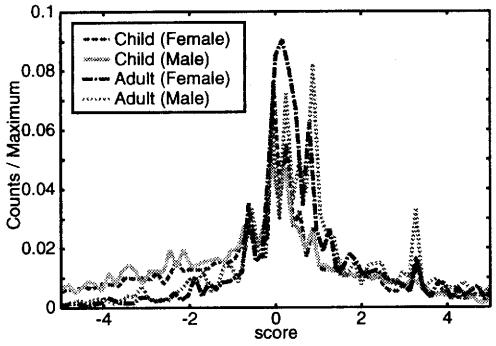

図 8 $LP_{adult} - LP_{child}$ の頻度分布
Fig. 8 Distributions of $LP_{adult} - LP_{child}$.

同様に言語的特徴 $LP_{adult} - LP_{child}$ の頻度分布を図 8 に示す（音響モデルには MAP 適応モデルを使用）．音響的特徴ほど大人と子供で分布傾向の違いは確認できない．しかし，分布の中心以外では大人と子供で若干の傾向違いがあり，音響的特徴との組合わせることにより識別精度の向上が見込まれる．

識別には統計アプローチに基づく二値分類器である SVM (Support Vector Machine) [15] [16] を用いた．SVM は主に自然言語処理分野で広く使われるアルゴリズムであり信頼性は高く [17]，小玉ら [18] によると複数の異なる音声認識プログラム出力の混合に SVM を用いることで他手法より高い精度向上を得ることができる．そこで，本研究では，大人・子供別の音声認識モデルを用いて並列に音声認識を行い，その結果得られる AP と LP で構成されるベクトルを SVM に与えるパラメータとした識別を行った．なお，SVM のカーネル関数には予備実験で最も良い結果を得た Gaussian 関数を用いる．

5. 評価実験

5.1 識別実験

提案手法を用いてテストセット（大人 500 発話，子供 500 発話）の識別実験を行った．SVM の学習に用いた音声は，表 1 の収集発話から大人，子供各 8,180 発話である．

実験結果を表 5 に示す．本手法を用いることで 91.8%の識別率を得た．全体として特徴量に言語的特徴を加えた方が音響的特徴のみの場合より高い識別精度を示した．MAP または MLLR 適応した音響モデルの影響が確認でき，MAP 適応時が最も精度が良い．音響的特徴，言語的特徴において大人・子供音声認識モデル間の差ではなく実測値を用いると若干の精度低下を起こす結果となった．

また，比較として従来法である GMM による識別実験も行う．学習に用いた音声は，テストセットを除いた表 1 の全発話である．SVM を用いた識別は大人・子供の二値分類だったが，GMM では a) ～ e) クラスの発話から各々64 混合のモデル計 5 個を作成した．GMM に使用した音声分析パラメータは，音声認識のものと同じ 16bit，16kHz 音声を窓シフト長 10ms で分析した 12 次元の MFCC と Δ MFCC，Δ Power である．

表 5 大人・子供識別率 (%)
Table 5 Discrimination rate of adult and child. (%)

特徴量タイプ	音響モデルの適応		
	適応なし	MAP	MLLR
音響的特徴のみ 1. $AP_{adult} - AP_{child}$	84.4	90.3	89.7
音響的特徴と言語的特徴 1. $AP_{adult} - AP_{child}$ 2. $LP_{adult} - LP_{child}$	87.1	91.8	91.3
音響的特徴と言語的特徴 （実測値を使用） 1. AP_{adult} 2. AP_{child} 3. LP_{adult} 4. LP_{child}	86.4	90.9	90.0
GMM 識別（ベースライン）			86.4

表 6 GMM による年齢層分類結果
Table 6 Age group classification results by GMM.

正解		分類結果				
		a)	b)	c)	d)	e)
a)	幼児	45	8	1	-	-
b)	低学年子供	65	182	114	2	-
c)	高学年子供	1	22	43	17	-
d)	大人	2	17	97	382	-
e)	高齢者	-	-	1	1	-

GMM を用いたテストセットの分類結果を表 6 に示す．5 モデル間で入力発話に対する尤度を比較して最も高い尤度を得たクラスに分類している．表中の各数字は，分類された発話の個数である．そして，a) ～ c) に分類された発話を子供，d)，e) に分類された発話を大人の発話と識別する．この結果，大人・子供の識別では正しく識別された発話は 86.4%（表 5 中ベースライン）であった．表 5 の提案手法はこの結果と比較して最大 5.4%の改善であり，その有効性が確認できたと言える．

5.2 モデル選択を伴う音声認識率

最後に，提案手法の音声認識精度に対する影響を調べる．大人・子供テストセット計 1000 発話に対して提案手法の SVM によって大人・子供を識別する．そして，大人もしくは子供音声認識モデルが出力した認識出力（表 4）を識別結果に従い選択し，認識率を算出した．単語正解精度を表 7 に示す．選択方法が “SVM で選択” が，提案法によって認識出力を選んだ時の結果である．“正解を選択” は，提案法には依らずテストセットの年齢層ラベルをそのまま用いて認識出力を選択した場合の精度である．また，“選択なし’ は，大人・子供別音声認識モデルを伴う並列デコーディングをせず，大人と子供全ての学習データ（表 2，表 3）から作成した年齢層非依存の音声認識モデルを用いて認識した際の結果である．

表 7 からわかるように，選択方法間で精度の違いは非常に小さいものになった．前節の識別実験で最も高い識別率を得た MAP 適応音響モデルを使用時，提案手法は 84.8%，正解を選択は 85.1%であり，若干の劣化が見られる．これは識別誤りによって話者に適した音声認識モデルが選択されなかったことに

表 7　モデル選択を伴う音声認識実験結果
Table 7　Speech recognition results with selected models.

選択方法	音響モデルの適応	単語正解精度 (%)
SVM で選択	適応なし	78.7
	MAP	84.8
	MLLR	84.0
正解を選択	適応なし	79.4
	MAP	85.1
	MLLR	84.3
選択なし	適応なし	79.1
年齢層非依存	MAP	84.5
モデルを使用	MLLR	84.0

起因する．一方，年齢層非依存モデルを使う選択なしとの比較では 0.3%ではあるが改善の傾向が見られる．

5.3 評価実験のまとめ

今回の実験では，年齢層に非依存な音声認識モデル使用時と比べ，提案手法による音声認識精度に劣化は見られなかった．また，話者の大人・子供識別では GMM から大きな改善を得ることができた．この結果，提案手法の有効性は示された．しかし，テストセットの年齢層ラベルを用いた選択時の正解精度から考えると，識別誤りが原因の精度劣化が若干見られる．

6. ま と め

本報告では，大人・子供の両ユーザに柔軟な順応力を備えた音声インタフェースを検討し，その実装に必要な音声認識スコアに基づく話者の大人・子供識別法を提案した．本手法は，統計学習のパラメータに音響的特徴と言語的特徴を併用することにより高い識別性能を有する．実験では，音声情報案内システム「たけまるくん」のフィールドテストで収集した実際のユーザ発話を用いた評価を行った．その結果，GMM の尤度比較に基づく識別法より 5.4%の識別率改善を得た．提案手法による音声認識精度への影響は少なく，改善はほとんど見られなかった．

また，子供発話の音声認識性能を調査し，既存の大人発話ベース音声認識の性能不足を示した．子供発話から音声認識モデルを構築することにより一定の精度向上を得ることはできたが，その認識精度は不十分であり，さらなる改良を必要とする．

大人と子供ではタスクドメインに対する知識レベルが異なっており，音声情報案内システムを利用する際の興味の対象も異なる．この違いに即した対話戦略の実装が今後の課題として求められる．また，識別手法の改良として，SVM 以外の統計学習アルゴリズムや音声認識結果の対数尤度以外の学習パラメータの導入を検討したい．最後に，たけまるくんへの大人・子供の両ユーザに柔軟な順応力を備えた音声インタフェースの実装とフィールドテストを通じた評価を予定している．

謝辞　本研究では，「生駒市北コミュニティセンター ISTA はばたき」にてデータ収集を行った．関係各位に深く感謝する．

文　献

[1] L. Bell, J. Gustafson: "Child and Adult Speaker Adaptation during Error Resolution in a Publicly Available Spoken Dialogue System", *Proc. 8th European Conference on Speech Communication and Technology (EUROSPEECH2003)*, pp.613–pp.616, 2003

[2] 小川 厚徳, 山口 義和, 松永 昭一: "小学生音声データベースの構築とそれを用いた子供音声認識の一検討", 電子情報通信学会技術研究報告, SP2002-36, 2000

[3] K. Shobaki, J.P. Hosom, R.A. Cole: "The OGI Kid's Speech Corpus and Recognizers" *Proc. 6th International Conferences on Spoken Language Processing (ICSLP2000)*, Vol.4, pp.258–261, 2000

[4] K. Komatani, S. Ueno, T. Kawahara, H.G. Okuno: "User Modeling in Spoken Dialogue Systems for Flexible Guidance Generation", *Proc. 8th European Conf. on Speech Communication and Technology (EUROSPEECH2003)*, pp.745–748, 2003

[5] 西村 竜一, 西原 洋平, 鶴身 玲典, 李 晃伸, 猿渡 洋, 鹿野 清宏: "実環境研究プラットホ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, D-II（採録決定）

[6] R. Nisimura, Y. Nishihara, R. Tsurumi, A. Lee, H. Saruwatari, K. Shikano: "Takemaru-kun: Speech-Oriented Information System for Real World Research Platform", *Proc. First International Workshop on Language Understanding and Agents for Real World Interaction*, pp.70–78, 2003

[7] A. Lee, T. Kawahara, K. Shikano: "Julius — An Open Source Real-Time Large Vocabulary Recognition Engine," *Proc. 7th European Conference on Speech Communication and Technology (EUROSPEECH2001)*, pp.1691–1694, 2001

[8] 李 晃伸, 河原達也, 武田一哉, 鹿野清宏, "Phonetic Tied-Mixture モデルを用いた大語彙連続音声認識", 電子情報通信学会論文誌, Vol.J83-D-II, No.12, pp.2517–2525, 2000

[9] K. Itou, M. Yamamoto, K. Takeda, T. Takezawa, T. Matsuoka, T. Kobayashi, K. Shikano, S. Itahashi: "JNAS: Japanese speech corpus for large vocabulary continuous speech recognition research", *The Journal of the Acoustical Society of Japan (E)*, Vol.20, No.3, pp.199–206, 1999

[10] J.L. Gauvain, C.H. Lee: "Maximum A Posteriori Estimation for Multivariate Gaussian Mixture Observation of Markov Chains", *IEEE Trans. on Speech and Audio Processing*, Vol.2, No.2, pp.291–298, 1994

[11] C.J. Leggetter, P.C. Woodland: "Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous-Density Hidden Markov Models", *Computer Speech and Language*, Vol.9, pp.171–185, 1995

[12] 長友 健太郎, 西村 竜一, 小松 久美子, 黒田 由香, 李 晃伸, 猿渡 洋, 鹿野 清宏: "相補的バックオフを用いた言語モデル融合ツールの構築", 情報処理学会論文誌, Vol.43, No.9, pp.2884–2893, 2002

[13] D.A. Reynolds, R.C. Rose: "Robust text-independent speaker identification using Gaussian mixture speaker models", *IEEE Trans. on Speech and Audio Processing*, vol.3, no.1, pp.72–83, January 1995.

[14] 峯松 信明, 広瀬 啓吉, 関口 真理子: "話者認識技術を利用した主観的高齢話者の同定とそれに基づく主観的年代の推定", 情報処理学会論文誌, Vol.43, No.7, pp.2186–2196, 2003

[15] V.N. Vapnik: *The Nature of Statistical Learning Theory*, Springer, 1995

[16] V. Wan, S. Renals: "SVMSVM: Support Vector Machine Speaker Verification Methodology", *Proc. 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2003)*, Vol.2, pp.221–224, 2003

[17] T. Kudo, Y. Matsumoto: "Fast Methods for Kernel-based Text Analysis", *Proc. 41st Annual Meeting of the Association for Computational Linguistics (ACL2003)*, pp.24–pp.31, 2003

[18] 小玉 康広, 渡辺 友裕, 宇津呂 武仁, 西崎 博光, 中川 聖一: "機械学習を用いた複数の大語彙連続音声認識モデルの出力の混合", 情報処理学会研究報告, 2003-SLP-45-16, 2003