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

2007－ＳＬＰ－69 (35)
2007／12／20

巨視的な時間発展系に基づく逐次モデル適応

-モデルの逐次更新における学習データの発話数に関する考察-

渡部 晋治† 中村 篤†

† 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所 〒619-0237 京都府相楽郡精華町光台
E-mail: †{watanabe,ats}@cslab.kecl.ntt.co.jp

あらまし 我々は，音響的特徴の発話単位での変化に追従するために，音響モデルパラメータ事後分布の時間的変動を記述する，巨視的な時間発展系に基づく適応法について研究を進めている．本稿では，提案法の逐次適応における実用上の課題として，逐次更新単位の長さについて調べた．まず予備的考察として，日本語話し言葉コーパスの同一講演音声内の音響的特徴が著しく時間変化することを実験的に示す．また本実験をもとに，対象発話全てを使って音響的特徴の時不変なミスマッチを緩和するバッチ適応と，時変なミスマッチを緩和する逐次適応，両手法の効果について考察する．最後に，逐次更新単位の長さに関して実験を行い，実用上のオンライン性と推定精度のトレードオフについて考察を行う．

キーワード 音声認識，音響モデル，逐次適応，巨視的な時間発展系，逐次更新における発話数

Acoustic model adaptation based on a macroscopic time evolution system

A study on the number of adaptation utterances in an incremental update of an acoustic model

Shinji WATANABE† and Atsushi NAKAMURA†

† NTT Communication Science Laboratory, NTT Coporation 2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0237 Japan
E-mail: †{watanabe,ats}@cslab.kecl.ntt.co.jp

Abstract Recently, we proposed an incremental adaptation method based on a macroscopic time evolution system, which models the dynamics of posterior distributions of acoustic model parameters to adjust time-variant acoustic characteristics during utterances. This paper examines recognition performance of the proposed incremental adaptation in terms of the length of incremental adaptation unit, which is critical for practical use in speech recognition. A preliminary experiment with the corpus of spontaneous Japanese reveals that acoustic characteristics are temporally changing even along the same lecture. Based on this experiment, we compare the relationship between batch and incremental adaptation methods, which can respectively mitigate time-invariant and time-variant mismatches of acoustic characteristics, in terms of the recognition performance of the lectures Finally, we discuss the practical trade-off between online property and estimation accuracy in terms of the length of incremental adaptation unit experimentally.

Key words speech recognition, acoustic model, incremental adaptation, macroscopic time evolution system, number of utterances in incremetal update

1. はじめに

一般に音声認識が対象とする音響的特徴は，発話を重ねるにつれ，話者交代や，外部ノイズ環境の変化，発話スタイルの変動等により，時々刻々と変化していく．特に会議や講演などの実環境音声認識タスクでは，上記の時間変動が度々生じ，学習データとのミスマッチが大きくなる．これは認識性能低下の大きな要因となっている．このような音響的特徴の時間的変動に

対処するために，音声認識用モデルを認識対象発話の時間変化に追従させて適応を行う，逐次適応の研究が広く行われている (例えば, [1]〜[5] など). 逐次適応においては，環境の時間変化に素早く追従すること (迅速性) や，過去に推定されたモデルと適応データとを共に用いて時間的ミスマッチを着実に緩和させること (安定性) の 2 点をバランスよく満たすことが重要である．我々は，音響的特徴の時間変化に迅速かつ安定的に追従するために，複数発話のまとまり（チャンク）を単位として，音響モデルパラメータ事後分布の時間的変動をチャンク単位の時系列モデルとして定式化することにより，巨視的な時間発展系に基づく適応法を提案した [6], [7].

提案法においては，チャンク単位の音響モデルの時間発展に対して線形動的システムを採用することにより，事後分布の逐次更新式が解析的に導出される. [6] では，これらの導出と共に，解析解がチャンク内適応データの統計量に基づいて更新される，拡張されたカルマンフィルタの解として解釈できることを示した．また，カルマンフィルタの予測更新アルゴリズムと同様の働きから，逐次適応における迅速性及び安定性が満たされることを理論的・実験的に示した. [7] では，提案法と従来の音響モデル適応法との関係について理論的・実験的に考察し，提案法が，従来の間接型・直接型適応法，及びそれらをカスケードにつなぐ組み合わせ手法を内包する枠組みであることを示した．ここで間接型適応法は，バイアス適応や MLLR 適応に代表される手法であり，音響的特徴のミスマッチを表現する変換パラメータ (平行移動やアフィン変換など) を少ないパラメータで効率よく表現し，そのパラメータを推定する枠組みである [8]〜[11]. また直接型適応法は，MAP 法に代表される，音響モデルパラメータを直接推定対象とする枠組みであり，ベイズ的手法を用いて，推定時のデータスパースネス問題に対処をしている [12].

このように，従来研究 [6], [7] では，提案法のカルマンフィルタ理論との整合性及び従来適応手法との関連性の観点から，理論的・実験的考察を行ってきた．一方で，逐次適応の実用上の課題として，逐次更新の単位をどのように設定するかという問題がある．例えば，仮に適応の処理時間が十分無視できるくらい高速であったとしても，認識過程において原理的に更新単位分の遅延が必ず生じるため，オンライン性の観点からは，更新単位をできるだけ短くするのが望ましいといえる．また，発話環境の急激な変化に対応できるという点でも更新単位を短くできた方が良い．しかし，提案法の実装において必要とされる変換パラメータの推定は，適応データ量が十分多ければ高い性能向上を示すが，適応データ量が少ない場合は推定精度が低くなり，逐次適応が十分機能しない．このように，変換パラメータの推定精度の観点から言えば，適応対象の音響的特徴が時不変であれば，その更新単位ができるだけ長い方が望ましい．

我々は，このような逐次更新単位の長さに依存するオンライン性と推定精度のトレードオフについて調べるため，日本語話し言葉コーパス (CSJ [13]) 中の特に講演時間の長い (発話数の多い) 講演に対して実験的考察を行う．本稿ではまず認識対象となる同一講演音声内の音響的特徴の時間変化を調べ，話者や講演内容が同一の場合でも，音響的特徴は時間的に著しく変化することを示す．また，このように時間的変化の激しい音声においては，音響的特徴のミスマッチの時変性が大きいと考えられるため，逐次適応が有効だと考えられる．従って，対象発話全てを使って音響的特徴の時不変なミスマッチを緩和するバッチ適応と，時変なミスマッチを緩和する逐次適応，両手法の効果について考察する．最後に，逐次更新単位の長さに関して実験を行い，実用上のオンライン性と推定精度のトレードオフについて考察を行う．

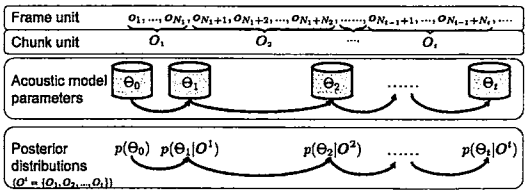

図 1　モデルパラメータの事後分布 $p(\theta_t|\boldsymbol{O}^t)$ のチャンク毎の更新.

2. 巨視的な時間発展系に基づくモデル適応

はじめに本節では，巨視的な時間発展系に基づくモデル適応について概説を行う (詳しくは [6]). 本適応法では，図 1 にあるように，特徴量ベクトルの時系列からなる適応データを，発話等の単位 (チャンク) でまとめられた部分時系列の連鎖ととらえ (式 (1))，各部分時系列の入力毎にモデルパラメータの事後分布を更新する．

$$\boldsymbol{O} = \{\underbrace{o_1, ..., o_{N_1}}_{\boldsymbol{O}_1}, ... \underbrace{o_{N_t+1}, ..., o_{N_t+N_{t+1}}}_{\boldsymbol{O}_t}, ...\} \qquad (1)$$

音響モデルのガウス分布平均パラメータ μ に注目すると，時刻 $t+1$ での事後分布 $p(\mu_{t+1}|\boldsymbol{O}^{t+1})$ は，次のような漸化式で表現される．

$$p(\mu_{t+1}|\boldsymbol{O}^{t+1}) \propto \underbrace{p(\boldsymbol{O}_{t+1}|\mu_{t+1})}_{1)} \int \underbrace{p(\mu_{t+1}|\mu_t)}_{2)} \underbrace{p(\mu_t|\boldsymbol{O}^t)}_{3)} d\mu_t \qquad (2)$$

ここでは音響モデル中の全ガウス分布の平均パラメータを推定の対象としているが，式中では簡単のため，ガウス分布インデックスは省略する. [6] では右辺の 3 つの分布は 1) 出力確率分布, 2) 離散確率過程, 3) 時刻 t での事後分布を意味している．以降ではこれらの 3 分布に対し，それぞれ具体系を当てはめていく．

1) 出力確率分布：連続分布 HMM

出力確率分布には通常の HMM, GMM で表現される連続分布 HMM を用いる．このとき部分時系列の適応データ $\boldsymbol{O}_{t+1}$ は次のような分布から出力される(注1).

(注1)：潜在変数が存在するため，正確には出力分布ではなく補助関数形式で分布が表現されている．

$$p(\boldsymbol{O}_{t+1}|\mu_{t+1}) = \prod_{n|o_n \in \boldsymbol{O}_{t+1}} (\mathcal{N}(o_n|\mu_{t+1}, \Sigma))^{\zeta_n} \qquad (3)$$

ここで $\mathcal{N}(\cdot|\mu, \Sigma)$ は平均パラメータ μ, 共分散行列パラメータ Σ のガウス分布である. ζ_n は注目するガウス分布に割り当てられた o_n の占有確率値である. Σ は初期モデルの共分散行列パラメータであり, 更新されないため時刻 t には依存しないとし, 状態遷移確率および混合重み因子は $p(\mu|\boldsymbol{O})$ の推定に関係ないため省いた. また HMM や GMM の潜在変数も式の上では無視したが, これらは EM アルゴリズムを用いることによって対処可能である.

2) 離散確率過程：線形動的システム

離散確率過程には線形動的システムを与える. このとき, μ のダイナミクスはアフィン変換で次のように表現される.

$$\mu_{t+1} = A\mu_t + \nu + \varepsilon_t$$

ここで ε_t は平均 $\mathbf{0}$, 共分散行列 U_t のガウシアンノイズである. これは, 通常の MLLR 法などで用いられるアフィン変換に加えて, μ が確率的に揺らいでいるシステムだといえる. このとき, 確率的ダイナミクスの分布具体形は

$$p(\mu_{t+1}|\mu_t) = \mathcal{N}(\mu_{t+1}|A\mu_t + \nu, U_t) \qquad (4)$$

として与えられる.

3) 時刻 t での事後分布：共役分布

最後に時刻 t での事後分布 $p(\mu_t|\boldsymbol{O}^t)$ に対しては共役分布であるガウス分布を仮定し, その平均ベクトルパラメータが $\hat{\mu}_t$, 共分散行列パラメータが $\hat{Q}_t$ で表現されるとすると関数形は

$$p(\mu_t|\boldsymbol{O}^t) = \mathcal{N}(\mu_t|\hat{\mu}_t, \hat{Q}_t) \qquad (5)$$

となる.

以上を具体系として与えることにより, 式 (3), (4), 及び (5) は全てガウス分布で表すことができる. これらを式 (2) に代入することにより次のような解析解を導出することができる.

$$p(\mu_{t+1}|\boldsymbol{O}^{t+1}) = \mathcal{N}(\mu_{t+1}|\hat{\mu}_{t+1}, \hat{Q}_{t+1}) \qquad (6)$$

ここで

$$\begin{cases} \hat{Q}_{t+1} &= ((U + A\hat{Q}_t A')^{-1} + \zeta_{t+1}\Sigma^{-1})^{-1} \\ \hat{K}_{t+1} &= \hat{Q}_{t+1}\zeta_{t+1}\Sigma^{-1} \\ \hat{\mu}_{t+1} &= A\hat{\mu}_t + \nu \\ & \quad + \hat{K}_{t+1}(\mathcal{M}_{t+1}/\zeta_{t+1} - A\hat{\mu}_t - \nu) \end{cases} \qquad (7)$$

である. $'$ は行列の転置を表す. 事後占有確率値の和 $\zeta_{t+1} = \sum_{n|o_n \in \boldsymbol{O}_{t+1}} \zeta_n$ および 1 次統計量 $\mathcal{M}_{t+1} = \sum_{n|o_n \in \boldsymbol{O}_{t+1}} \zeta_n o_n$ は部分時系列の適応データ $\boldsymbol{O}_{t+1}$ が与えられた際の十分統計量である.

以上により, 本時間発展系の逐次更新は, $p(\mu_{t+1}|\boldsymbol{O}^{t+1})$ の分布パラメータ $\hat{Q}_{t+1}$, $\hat{\mu}_{t+1}$ が式 (7) によって更新されることにより実現される. [6] においては, 本解が通常のカルマンフィルタ解と同様の性質を持ち, カルマンフィルタ理論の予測更新アルゴリズムを内包していることを示し, また [7] では, 間接型・直接型適応法を包含する枠組みであることも示した. 図 2(注2)は従来のカルマンフィルタに基づく音声認識もあわせた相関関係を示している.

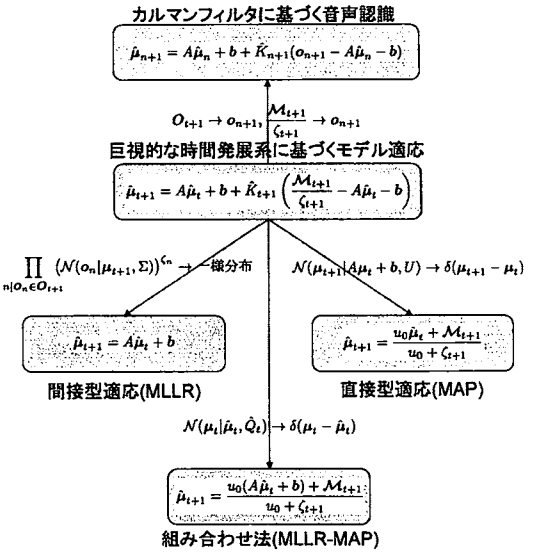

図 2 本手法と従来型適応法及び一般的なカルマンフィルタに基づく音声認識の相関図.

2.1 実　　装

提案法の実装の概略について図 3 に示す. 更新式 (7) を実装するに当たって, システムノイズの分散 U, 変換パラメータ A, b 及び十分統計量 $\zeta_{t+1}, \mathcal{M}_{t+1}$ を与える必要がある. 本稿では, システムノイズパラメータは $U = (u_0)^{-1}\Sigma$ と仮定して, u_0 を調整パラメータとする. 変換パラメータの推定には $\boldsymbol{O}_{t+1}$ を用いる. その際, 変換パラメータの推定に関しては通常の MLLR 推定アルゴリズム [10], [11] を, もしくは $A = I$ としてバイアス推定アルゴリズムを用いる [8], [9]. また, この推定の際に出力される十分統計量をそのまま式 (7) に用いる.

このように本実装においては, 変換パラメータはチャンク内の適応データ $\boldsymbol{O}_{t+1}$ を用いて推定する. 従って, 提案法の逐次適応におけるチャンクあたりの発話量は, 変換パラメータの推定精度, 及びそれがもたらす認識性能に対して影響を与えると考えられるため, 実験的に考察する必要がある.

3. 実　　験

3.1 実験環境

読み上げ音声から講演音声への音響モデルの教師無しの逐次適応を行った. 図 4 はその実験手順, 表 1 は, 音声認識実験に用いられた分析条件, 音響モデル及び言語モデル情報を示す. まず, IPA 準拠の読み上げ音声データベースの男性話

(注2)：図中間接型適応への極限では, 出力確率分布の影響を無視する無情報分布として, 定義域が無限大の多次元一様分布を採用している.

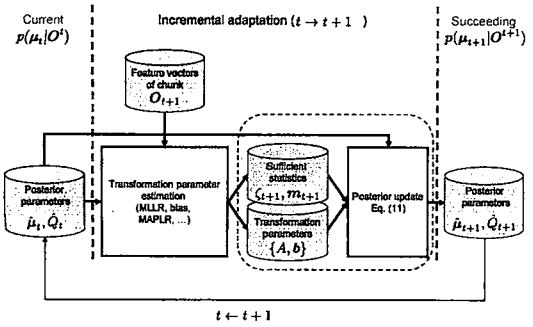

図 3 巨視的な時間発展系に基づく逐次適応法の実装の概略図.

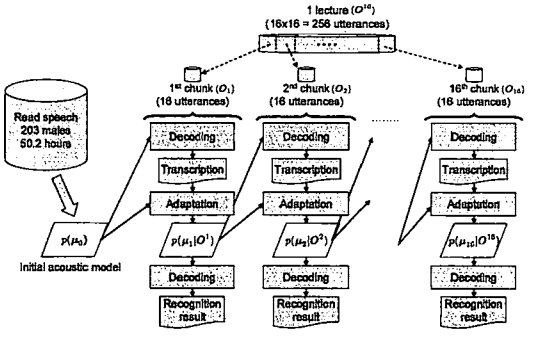

図 4 読み上げ音声から講演音声への教師無し逐次適応の実験手順. 図中では逐次適応の更新単位（チャンク）を 16 発話とし，これを 16 ステップ (計 256 発話分) 繰り返した.

者 203 人分 50.2 時間を用い，総状態数 2,000, 状態あたり混合数 16 の不特定話者音響モデルを構築した (特徴量は 12 次元 MFCC+Energy+Δ+ΔΔ). これを初期音響モデルとし，1 講演あたり 256 以上の発話数を含む比較的長時間の CSJ10 講演を対象として，教師無しの逐次適応を行った. 教師無し逐次適応の手続きは，図 4 にあるように，各チャンクごとの適応対象発話に対する 1) ラベル情報を取得するための認識, 2) 得られたラベル情報を元にした適応学習, 3) 適応された音響モデルを用いた認識，によって構成される. 発話単位はトランスクリプションからの情報をそのまま用いた. また，言語モデルは CSJ のトランスクリプションから作成し，語彙サイズ 3 万語のトライグラムを用いた.

3.2 音響的特徴の時間変化

はじめに，音響的特徴の時間変化について考察を行う. 一般に，音響的特徴の時間変動要因としては，話者交代や収録環境の変化などが考えられる. 本実験で対象としている講演音声の適応タスクには上記のような変動要因は存在しないが，同一講演内においても音響的特徴は，発声内容の変化に伴う発話様式の変化 (抑揚や強調など) や，長時間発声による疲れに起因する発音の怠けなど様々な変動要因が想定される. 本節では，音声認識に影響を与える音響的特徴の指標として，発話様式の違いを表す発話速度，及び音声認識率の音響モデルの性能を表す音

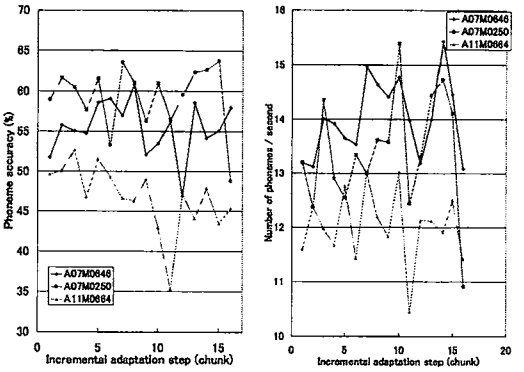

図 5 不特定話者音響モデルを用いた音素認識率の時間変化（左図），及び 1 秒あたりの発声音素数の時間変化（右図）.

素認識率，両指標を用いて時間的変化を測定した. ここで，発話速度はトランスクリプションから 1 秒当たりの音素数を算出して求め，音素認識率は不特定話者 (初期) 音響モデルのみから音素認識実験を行うことにより音素正解精度として求めた. なお本実験では，更新単位（チャンク）を 16 発話とし，これを 16 ステップ (計 256 発話分) 繰り返した.

図 5 は，逐次適応で評価対象とする 10 講演のうちの 3 講演 (講演 ID:A07M0646, A07M0250, A11M0664) に対する音素認識率・発話速度を示す. 図より，音素認識率・発話速度共に同一講演内でもチャンクごとに非常に大きくばらついており，音響的特徴は時間的に変動しているということがわかる. このことから，時変な音響的ミスマッチの緩和を目的とする逐次適応が，講演のような発声時間が長いタスクで有効である可能性が示唆される.

3.3 講演音声に対するバッチ適応と逐次適応の効果

本実験では，本適応タスクに存在する話者・発話スタイルといった時不変なミスマッチ，及び前節の実験結果で示した講演内の時変のミスマッチに対する，バッチ適応と逐次適応の認識性能の効果について考察を行う. バッチ適応では 256 発話をまとめて適応データに用い，逐次適応では前節同様チャンクを 16 発話として逐次更新を行った. 認識性能は 10 講演の平均で算出し，不特定話者音響モデルの単語誤り率 (平均 33.2%) からの誤

表 1 分析条件，音響モデル及び言語モデル情報

Sampling rate/quantization	16 kHz / 16 bit
Feature vector	12 order MFCC with energy
(39 dimensions)	+Δ+ΔΔ
Window	Hamming
Frame size/shift	25/10 ms
Number of temporal HMM states	3 (left to right)
Number of phoneme categories	43
Number of context-dependent HMM states	2,000
Number of mixture components	16
Language model	Standard trigram (made by CSJ transcription)
Vocabulary size	30, 000
Perplexity (OOV rate)	82.2 (2.1 %)

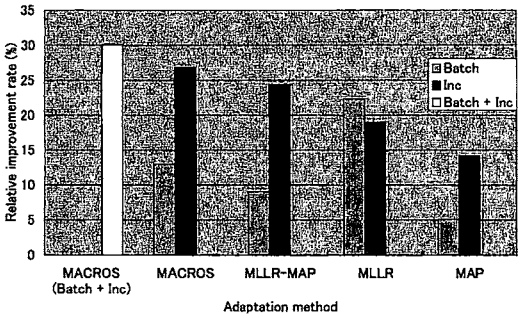

図 6　教師ラベル無し逐次適応における提案法 (MACROS) と従来法 (MLLR-MAP, MLLR, MAP) のバッチ，逐次適応での比較.

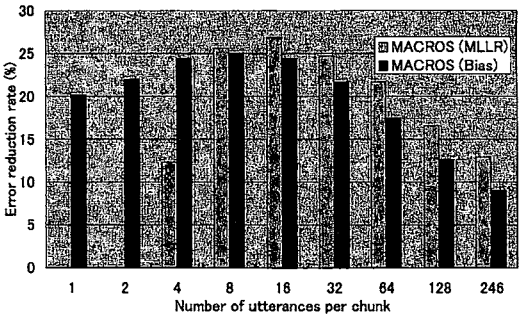

図 7　チャンク中の発話数を変化した際の，逐次適応による誤り改善率.

り改善率で評価した．図 6 は逐次適応の全てのステップでの性能の平均を取ったものであり，あわせて全データをまとめて適応させたバッチ適応の結果も示している．提案法 (MACROS) の実装に関しては，変換パラメータの推定法に MLLR 推定法を採用し，システムノイズパラメータの設定に関しては，[6] での実験結果を参考にして $u_0 = 10$ とした．図 6 には，比較のため，MLLR 法や MAP 法，及びそれらをカスケードに組み合わせた MLLR-MAP 法も加えた．

まずバッチ適応での実験結果について考察する．バッチ適応では MLLR 法の方が MAP 法よりも改善率が高いことがわかる．一般に，教師有り適応では，今回のケースのように十分データが多い場合は MAP 法の方が MLLR 法よりも性能改善が高いと言われている [14]．しかし教師無し適応の場合は，図 4 にある認識器によるラベル付けの段階で，ラベルに認識誤りが含まれるため，ラベルの信頼性が低下する．そのような場合，間接的に変換パラメータを推定する MLLR 法の方が，直接モデルパラメータを推定する MAP 法よりも，認識誤りの影響が緩和されるため，MAP 法よりも性能改善が大きくなる．結果として MLLR-MAP や MACROS 法も，MAP 法の影響に引きずられて，バッチ適応では MLLR 法よりも改善率が悪くなっている(注3)．

一方，逐次適応においては，MAP 法の時間的変動に対する緩和効果がもたらす安定性が強く働くため，特に，MLLR-MAP 法や MACROS 法は MLLR 法も含めたバッチ適応法に比べて大きな改善率を示すことがわかる．このように，MACROS 法やその近似的極限である MLLR-MAP 法の時系列モデル (線形回帰) は，音響的特徴の時間変化に有効である．

最後に，MLLR 法でバッチ適応を行った音響モデルを初期モデルとして，MACROS 法で逐次適応を行ったのが図 6 の MACROS(Batch+Inc) の結果である．これにより更なる性能改善が得られているが，これはバッチ適応により話者性等に起因する音響的特徴の時不変ミスマッチを，逐次適応により発話様式の変化等に起因する音響特徴の時変ミスマッチを，両方緩和したことによる効果と考えられる．MACROS(Batch+Inc) により得られた単語正解精度は 76.8 % であった．評価セットが違うため厳密な比較はできないが，この精度は CSJ 音声を音響モデルの学習データとして，最尤学習により同一モデル規模の音響モデル・言語モデルで得られる不特定話者の認識率 (76.9 % [16]) におおよそ匹敵している(注4)．

3.4　チャンク中の発話数の変化と認識性能の関係

前節の実験より MACROS 法の逐次適応は，講演音声内の音響的特徴の時間変化に対して適切にモデル化できることを示したが，一方で逐次更新単位の長さは固定 (チャンクあたり 16 発話) して議論を進めた．しかし，逐次更新単位の長さは実用上，オンライン性と推定精度のトレードオフの関係にあり，また変換パラメータの推定手法に応じてそのトレードオフの傾向は変化すると考えられる．従って本実験では，チャンク中の発話数を変化した際の逐次適応による誤り改善率について考察を行った．その際，前節に引き続き変換パラメータの推定法に MLLR 推定法を採用した MACROS(MLLR) と共に，バイアス推定法を採用した MACROS(Bias) 2 種類の逐次適応を用意して傾向を調べた．

図 7 は，発話単位を変えた場合の MACORS(MLLR), MACROS(Bias) 両手法の改善率を表す．なお 1 発話，2 発話の場合の MACROS(MLLR) は，適応データが過度にスパースであるために適応データ量見合いの MLLR 推定の推定精度が悪く，認識性能を大幅に劣化させるため，図からは除外してある．8 発話以降バイアス適応よりも MLLR 推定法をベースにした逐次適応のほうが性能が良く，認識性能の観点からは，16 発話更新の MACROS(MLLR) が最も性能が高くなる．一方で，バイアス推定に基づく MACROS(Bias) 適応のほうがスパースデータに対して頑健である．以上の結果から，実用上は，認識性能を重視する場合は MACROS(MLLR), オンライン性

(注3)：u_0 などの初期値を調節することにより,MACROS 法及び MLLR-MAP 法における MAP 法の影響を弱めて，MLLR 法と同程度の認識性能をもたらすことも可能であるが，本実験では議論の簡略化のためにバッチ適応においても $u_0 = 10$ として実験を行った．また，信頼度の高い発話を抽出しそれを用いた教師なし話者適応を行うことによっても改善は可能であり，講演音声においてその効果は示されている [15].

(注4)：ただし，音響モデル・言語モデルのサイズを全共分散モデルの適用や大語彙化によって大規模化し，識別学習や教師無し適応学習などの学習手法で高精度化した際の，特定話者認識率 (約 85% [17], [18]) と比べると，認識性能に依然改善の余地があるといえる.

を重視する場合はMACROS(Bias)と用途に応じて使い分ける必要があるといえる．

4. むすび

本研究では巨視的な時間発展に基づく適応法の逐次更新単位の長さに関して実験を行い，その実用性について考察を行った．本実験では，変換パラメータの推定法にMLLR推定法，バイアス推定法を採用した2種類の実装に対して検証を行い，両推定法の適応データ量に応じた特長が，そのまま逐次適応の結果に現れた．具体的には，適応データが十分多い場合はMLLR推定法は高精度であるが少量の場合はバイアス推定法の方が頑健という性質から，更新単位が短い場合はバイアス推定法が，更新単位が長い場合はMLLR推定法が適しているのがわかった．逐次適応の実用上は，短い更新単位で高い認識性能をえることが求められるため，今後は，変換パラメータの推定法についてベイズ推定などのスパース推定手法をMLLR法に適応するなど(例えば[19])モデル化・学習両面から，更なる提案法の改善をすすめる．

また，今回の評価系は同一話者同一講演内の発話様式の変化である．この場合に想定される音響的特徴の時間的変化は，発声内容の変化に伴う発話様式の変化(抑揚や強調など)や，長時間発声の疲れに起因する発音の怠けなどが要因として考えられ，発話やチャンクといった時間スケールで見た場合比較的なだらかな(連続的な)変化であると考えられる．一方，会議や放送音声のような状況では，話者交代や収録環境の変化等が存在し，それらは上記発話様式と比べて不連続な音響的特徴の時間変化だといえる．今後は，会議や放送音声なども評価の対象として考え，連続・不連続な音響的特徴の時間変化に対する，提案法の枠組みによる逐次追従適応の効果について検証していきたい．

謝辞

チャンク中の発話数の変化と認識性能の関係について認識実験にご協力頂いた，東京大学西亀健太氏に感謝する．

文献

[1] T. Matsuoka and C.-H. Lee. A study of on-line Bayesian adaptation for HMM-based speech recognition. In *Proc. EUROSPEECH'93*, pp. 815–818, 1993.

[2] G. Zavaliagkos, R. Schwartz, and J. Makhoul. Batch, incremental and instantaneous adaptation techniques for speech recognition. In *Proc. ICASSP1995*, Vol. 1, pp. 676–679, 1995.

[3] Q. Huo and C.-H. Lee. On-line adaptive learning of the continuous density hidden Markov model based on approximate recursive Bayes estimate. *IEEE Transactions on Speech and Audio Processing*, Vol. 5, pp. 161–172, 1997.

[4] C. J. Leggetter and P. C. Woodland. Flexible speaker adaptation using maximum likelihood linear regression. In *Proc. ARPA Spoken Language Technology Workshop*, pp. 104–109, 1995.

[5] V. Digalakis. Online adaptation of hidden Markov models using incremental estimation algorithms. *IEEE Transactions on Speech and Audio Processing*, Vol. 7, pp. 253–261, 1999.

[6] 渡部晋治，中村篤．確率分布の巨視的な時間発展システムに基づく逐次モデル適応．秋季音響学会講演論文集，2-2-10, pp. 71–72, 2006.

[7] 渡部晋治，中村篤．巨視的な時間発展系に基づくモデル適応と従来型適応との関係の考察．秋季音響学会講演論文集，2-3-12, 2007.

[8] K. Shinoda and T. Watanabe. Speaker adaptation with autonomous control using tree structure. In *Proc. EUROSPEECH95*, pp. 1143–1146, 1995.

[9] M. Rahim and B.-H. Juang. Signal bias removal by maximum likelihood estimation for robust telephone speech recognition. *IEEE Transactions on Speech and Audio Processing*, Vol. 4, pp. 19–30, 1996.

[10] C. J. Leggetter and 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.

[11] V. Digalakis, D. Ritischev, and L. Neumeyer. Speaker adaptation using constrained reestimation of Gaussian mixtures. *IEEE Transactions on Speech and Audio Processing*, Vol. 3, pp. 357–366, 1995.

[12] J.-L. Gauvain and C.-H. Lee. Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains. *IEEE Transactions on Speech and Audio Processing*, Vol. 2, pp. 291–298, 1994.

[13] S. Furui, K. Maekawa, and M H. Isahara. A Japanese national project on spontaneous speech corpus and processing technology. In *Proc. ASR2000*, pp. 244–248, 2000.

[14] C.-H. Lee and Q. Huo. On adaptive decision rules and decision parameter adaptation for automatic speech recognition. In *Proceedings of the IEEE*, Vol. 88, pp. 1241–1269, 2000.

[15] 緒方淳，有木康雄．音素事後確率に基づく信頼度を用いた音響モデルの教師なし適応化．電子情報通信学会技術研究報告，SP2001-105, pp. 19–24, 2001.

[16] 中村篤，大庭隆伸，渡部晋治，石塚健太郎，堀貴明，Mike Schuster, Erik McDermott, 南泰浩．音声認識システム solon の日本語話し言葉コーパス (公開版 ver.1.0) による評価．電子情報通信学会技術研究報告 SP2005-106, pp. 7–12, 2005.

[17] 草間隆，奥山洋平，加藤正治，小坂哲夫，好田正紀．繰り返し教師なし適応による講演音声認識．秋季音響学会講演論文集，2-3-14, 2007.

[18] 大庭隆伸，渡部晋治，石塚 健太郎雅清，Erik McDermott, 南泰浩，中村篤．音声認識システム solon における日本語講演音声への教師なし適応に関する評価．春季音響学会講演論文集，1-9-11, 2007.

[19] C. Chesta, O. Siohan, and C.-H. Lee. Maximum a posteriori linear regression for hidden Markov model adaptation. In *Proc. Eurospeech1999*, Vol. 1, pp. 211–214, 1999.