

最尤推定に基づく線形変換を用いた声道長正規化

六井淳[†] 中井満[†] 下平博[†] 嵯峨山茂樹^{††}

[†] 北陸先端科学技術大学院大学 情報科学研究科, 石川県
^{††} 東京大学大学院 情報理工学研究所システム情報学専攻, 東京都

E-mail: †{rokui,mit,sim}@jaist.ac.jp, ††sagayama@hil.t.u-tokyo.ac.jp

あらまし 声道長正規化 (Vocal Tract Length Normalization: VTLN) は音声認識において最も盛んに研究されている話者適応技術の1つである。本報告では最尤推定法を用いて周波数軸の線形伸縮係数を推定し、これによって音響モデルのパラメータを適応する話者適応法を提案する。現在、HMM (隠れマルコフモデル) のパラメータ推定に用いられるEMアルゴリズムに基づくケプストラム領域での声道長正規化手法が提案されている。これら声道長正規化手法の多くは周波数領域において端点固定の非線形伸縮となり、HMMパラメータ推定はケプストラム領域で行われていた。これに対し、提案手法は唯一つの伸縮係数を持つ線形な周波数伸縮を仮定しており、テイラー展開の一次項近似によりメルケプストラム領域でモデル化される手法である。提案手法は不特定話者孤立単語認識実験を通して認識性能の向上が確認された。

キーワード 声道長正規化、線形変換、最尤推定、話者適応、話者正規化

Vocal Tract Length Normalization Using Linear Transformation based on Maximum Likelihood Estimation

Jun ROKUI[†], Nakai MITSURU[†], Hiroshi SHIMODAIRA[†], and Shigeki SAGAYAMA^{††}

[†] Japan Advanced Institute of Science and Technology, Hokuriku. Dept of Information Science. 1-1
Asahidai Tatunokuchi, Ishikawa 923-1292 Japan

^{††} The University of Tokyo. Graduate School of Information Science and Technology. 7-3-1 Hongo
Campus Bunkyo Tokyo 113-8654 Japan

E-mail: †{rokui,mit,sim}@jaist.ac.jp, ††sagayama@hil.t.u-tokyo.ac.jp

Abstract Vocal tract length normalization (VTLN) is one of the popular speaker adaptation techniques for speech recognition. The present study proposes a new VTLN algorithm in which expectation-maximization (EM) based parameter adaptation of HMM to vocal tract length is achieved in the mel-cepstral domain by utilizing a linear transformation model. Compared to other existing approaches based on bi-linear transformation for VTLN where a specific non-linear frequency warping function is employed in the spectrum domain and parameter adaptation of HMM is carried out in the cepstral domain, the proposed approach assumes a linear frequency warping with a single scaling factor and equivalent operation is modeled in the mel-cepstral domain by using a first order Taylor series approximation. The proposed scheme demonstrates significant improvement of recognition performance in a speaker independent word recognition task.

Key words Vocal Tract Length Normalization, Linear Transformation, Maximum Likelihood Estimation, Speaker Adaptation, Speaker Normalization.

1. はじめに

現在の音声認識技術において性能劣化要因として話者性の違いや雑音、回線などの使用条件の変化などが挙げられる。本報告では話者性の違いによる性能劣化に着目する。

話者性の違いに対し音声認識における頑健性を保つため、様々な状況における多数の話者の発声データを用いてモデルを学習することで話者性の違いに対応しようとする不特定話者認識システムの研究が行われている。しかしながら、不特定話者認識システムの構築は話者による音声の多様性のため与えられた全ての特徴空間に適應するモデルは作成することができないという問題がある。この問題に対し、話者適應や話者正規化の研究が行われている。

最近の話者適應化研究では MLLR(Maximum Likelihood Linear Regression) [2] などの線形変換行列作成の際に尤度を利用し、話者性の違いに適應させるという手法が提案されている。また、VFS(Vector Field Smoothing) [10], [11] は特徴量空間の部分的変動に対して効果的な手法である。

モデルパラメータを再推定によって学習するという話者適應化に対して、話者正規化は特徴量抽出の段階で話者性の違いを取り除く手法である。話者正規化に効果的な手法として VTLN(Vocal Tract Length Normalization) [3], [9] が挙げられる。VTLN は比較的少量の音素データから声道長を推定する手法である。しかし、実環境下では声道長の推定精度が低下するといった問題が指摘されている。この問題に対し、ML-VTLN(Maximum Likelihood Vocal Tract Length Normalization) [5]~[7] が提案されている。ML-VTLN は予め用意した複数の声道長パラメータの中から対象話者に最適なパラメータを選択する手法である。この手法では用意したパラメータの数だけ尤度計算が必要であり、計算量が多い。近年、全域通過フィルタを用いることでケプストラム領域でのモデルパラメータ適用を行おう VTLN-R(Vocal Tract Length Normalization using Rapid Maximum Likelihood Estimation) [1], [8] が提案された。これら声道長正規化研究の多くは周波数領域で非線形な伸縮となる端点固定の変換に基づくものである。

一方で、最近の研究では周波数領域からケプストラム領域への線形変換によって声道長を線形に正規化し、その推定には最尤推定を用いる手法が提案されている [7]。本研究では声道長正規化を1つのパラメータによって構成される周波数領域での線形変換

によって行う手法を提案する。パラメータ選択の必要はなく、計算量は少ない。

2. 周波数線形伸縮による話者適應法

2.1 周波数領域での線形変換

周波数軸の線形伸縮は一般に伸縮係数 ν を用いて、 $\hat{\omega} = \nu\omega$ の形で定式化される ($\omega, \hat{\omega}$ は変換前と後の角周波数を意味する)。一方、離散化された周波数軸における表現を考えると、周波数パワースペクトルはベクトルの形で表現されるので、周波数変換を行う前のパワースペクトルを \mathbf{p} 、変換後のそれを $\hat{\mathbf{p}}$ で記すと、両者は以下のような線形変換で表せる。

$$\hat{\mathbf{p}} = \mathbf{W}\mathbf{p} \quad (1)$$

ここで、変換行列 \mathbf{W} は次式で与えられる。

$$w_{i,j} = \begin{cases} 1, & \text{if } i = \lfloor \nu \cdot j \rfloor, \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

$$\mathbf{W}_{N \times N} = \begin{bmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N,1} & \dots & w_{N,j} & \dots & w_{N,N} \end{bmatrix} \quad (3)$$

($1 \leq i, j \leq N$)。 $\lfloor \cdot \rfloor$ は小数点第1位以下を繰り下げた整数値である。大人から子供、男性から女性などのホルマント位置が周波数領域の低域にある話者からより高域にある話者への適應の場合は伸縮係数 $\nu > 1$ となる。逆に、ホルマント位置が周波数領域の高域にある話者からより低域にある話者へは $\nu < 1$ となり、伸縮パラメータ ν により一意に求まる伸縮行列 \mathbf{W} によって周波数パワースペクトルが伸縮される。

次に周波数領域での変換行列 \mathbf{W} に基づき、メルケプストラム領域への変換方法を示す。パワースペクトル \mathbf{p} からメル周波数スペクトル \mathbf{s} への変換は次のように表される。

$$\mathbf{s} = \mathbf{F}\mathbf{p} \quad (4)$$

$$\mathbf{F}_{n_c \times N} = \begin{bmatrix} f_{1,1} & \dots & f_{1,N_1} & 0 & \dots & 0 \\ 0 & \dots & f_{2,1} & \dots & f_{2,N_2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \dots & \dots & \dots & f_{n_c, N_{n_c}} \end{bmatrix} \quad (5)$$

\mathbf{F} はフィルタバンク、 $f_{i,j}$ はフィルターの重みである (チャンネル幅: $N_{n_c} \gg N_1$)。 n_c はフィルタバンクのチャンネル数である。

本研究の定式化では線形周波数軸上で元のパワースペクトル \mathbf{p} は式 (4) の変換の逆行列、すなわちメル周波数スペクトル \mathbf{s} からは計算することができない。これは、フィルタバンク F は正規行列ではないため、逆行列が存在しないためである。そこで、本研究では F の一般化逆行列 F^- を用いる。

$$\mathbf{p} = F^- \mathbf{s} \quad (6)$$

行列 F^- は $F \cdot F^- = I$ となるような行列である。 I は恒等行列である。結果として、変換後のメル周波数スペクトル $\hat{\mathbf{s}}$ とメル周波数スペクトル \mathbf{s} の関係は次のように表される。

$$\hat{\mathbf{s}} = T \mathbf{s} \quad (7)$$

$$T = F \cdot W \cdot F^- \quad (8)$$

メル周波数スペクトル \mathbf{s} は DCT 変換行列を C とした場合、メルケプストラム \mathbf{c} を用いることで $\mathbf{s} = \exp(C^{-1} \mathbf{c})$ と表すことができる。これにより初期 MFCC パラメータ集合 \mathbf{c} と変換 MFCC パラメータ集合 $\hat{\mathbf{c}}$ との間には次のような関係が成り立つ。

$$\hat{\mathbf{c}} = C \log \{ T \exp(C^{-1} \mathbf{c}) \} \quad (9)$$

2.2 HMM 適用のための拡張

MFCC の変換は式 (9) のように与えられるため、MFCC パラメータ平均 $\mu_{\mathbf{c}}$ 、共分散行列 $\Sigma_{\mathbf{c}}$ を与えると、変換後の MFCC パラメータ平均 $\hat{\mu}_{\mathbf{c}}$ 、共分散行列 $\hat{\Sigma}_{\mathbf{c}}$ が求まる。この時、変換自体が線形でないために、次のような問題が生じる。

$$E[\hat{\mathbf{c}}] = E\{C \cdot \log\{T \cdot \exp(C^{-1} \cdot \mathbf{c})\}\} \quad (10)$$

$$\neq C \cdot \log\{T \cdot \exp(C^{-1} \cdot E\{\mathbf{c}\})\} \quad (11)$$

すなわち、点から点への変換を表す式 (9) は平均ベクトルの変換に用いることができない。

$$\hat{\mu}_{\mathbf{c}} \neq C \cdot \log\{T \cdot \exp(C^{-1} \cdot \mu_{\mathbf{c}})\} \quad (12)$$

本研究ではこの問題を解決するために変換方式を以下に示すように線形近似する。

式 (9) を $\exp(x) \approx (1+x)$ という近似を用いることで変換 $\hat{\mathbf{c}}$ は次式のように与えられる。

$$\hat{\mathbf{c}} = C \log \{ T \exp(C^{-1} \mathbf{c}) \} \quad (13)$$

$$\approx C \log \{ T(e + C^{-1} \mathbf{c}) \} \quad (14)$$

$$\hat{\mathbf{c}} \triangleq C \log \{ T(e + C^{-1} \mathbf{c}) \}$$

$$= C \log \left\{ \begin{bmatrix} \sum_j t_{1,j} \\ \dots \\ \sum_j t_{n_c,j} \end{bmatrix} + T C^{-1} \mathbf{c} \right\} \quad (15)$$

$$= C \log \left\{ \begin{bmatrix} \sum_j t_{1,j} & & 0 \\ & \dots & \\ 0 & & \sum_j t_{n_c,j} \end{bmatrix} \cdot \left(e + \begin{bmatrix} \sum_j t_{1,j} & & 0 \\ & \dots & \\ 0 & & \sum_j t_{n_c,j} \end{bmatrix} T C^{-1} \mathbf{c} \right) \right\} \quad (16)$$

$t_{i,j}$ は変換行列 T の要素であり、 \mathbf{q} , B は次式で与えられる。

$$B = C \begin{bmatrix} \frac{1}{\sum_j t_{1,j}} & & 0 \\ & \dots & \\ 0 & & \frac{1}{\sum_j t_{n_c,j}} \end{bmatrix} T C^{-1} \quad (17)$$

$$\mathbf{q} = C \begin{bmatrix} \log \{ \sum_j t_{1,j} \} \\ \dots \\ \log \{ \sum_j t_{n_c,j} \} \end{bmatrix} \quad (18)$$

線形近似により、平均値 $\mu_{\mathbf{c}}$ 、分散 $\Sigma_{\mathbf{c}}$ に関する変換も次のように簡潔になる

$$\hat{\mu}_{\mathbf{c}} \approx \mathbf{q} + B \cdot \mu_{\mathbf{c}} \quad (19)$$

$$\approx C \cdot \log(T \cdot \exp(C^{-1} \cdot \mu_{\mathbf{c}})) \quad (20)$$

$$\hat{\Sigma}_{\mathbf{c}} \approx B \cdot \Sigma_{\mathbf{c}} \cdot B^t \quad (21)$$

さらに、 i 番目フレームの Δ ケプストラム $\Delta \mathbf{c}^{(i)}$ と Δ^2 ケプストラム $\Delta^2 \mathbf{c}^{(i)}$ の変換に関しても次のような効果が得られる。

$$\Delta \hat{\mathbf{c}}^{(i)} = \hat{\mathbf{c}}^{(i)} - \hat{\mathbf{c}}^{(i-1)} \quad (22)$$

$$\approx \mathbf{q} + B \cdot \mathbf{c}^{(i)} - \mathbf{q} - B \cdot \mathbf{c}^{(i-1)} \quad (23)$$

$$\approx B \cdot (\mathbf{c}^{(i)} - \mathbf{c}^{(i-1)}) \quad (24)$$

$$\approx B \cdot \Delta \mathbf{c}^{(i)} \quad (25)$$

$$\Delta^2 \hat{\mathbf{c}}^{(i)} \approx B \cdot \Delta^2 \mathbf{c}^{(i)} \quad (26)$$

平均値 $\mu_{\mathbf{c}}$ 、分散 $\Sigma_{\mathbf{c}}$ に関する変換も次のように簡潔になる

$$\hat{\mu}_{\Delta \mathbf{c}} \approx B \cdot \mu_{\Delta \mathbf{c}} \quad (27)$$

$$\hat{\mu}_{\Delta^2 \mathbf{c}} \approx B \cdot \mu_{\Delta^2 \mathbf{c}} \quad (28)$$

$$\hat{\Sigma}_{\Delta \mathbf{c}} \approx B \cdot \Sigma_{\Delta \mathbf{c}} \cdot B^t \quad (29)$$

$$\hat{\Sigma}_{\Delta^2 \mathbf{c}} \approx B \cdot \Sigma_{\Delta^2 \mathbf{c}} \cdot B^t \quad (30)$$

これにより、 Δ ケプストラム $\Delta \mathbf{c}^{(i)}$ と Δ^2 ケプストラム $\Delta^2 \mathbf{c}^{(i)}$ への拡張も容易に行うことができる。

3. 伸縮係数の尤度推定法

本章では周波数領域での伸縮行列 W を構成する伸縮係数 ν の推定方法について述べる。最適な ν^* を求めるために、次のような基準を設ける。

$$\nu^* = \arg \max_{\nu} P(O|\Theta) \quad (31)$$

O は観測系列、 $P(O|\Theta)$ は出現確率であり、 $\Theta \equiv (\theta, \nu)$ は HMM パラメータ集合である。

式 (16) を Baum-Welch アルゴリズムに組み込むことにより伸縮係数 ν を推定する。最大化すべき目的関数を次のように定める。

$$\Phi(\Theta', \Theta) = \sum_{j=1}^J \sum_{t=1}^T P(O, q_t = j|\Theta') \log b_j(\tilde{c}_t) \quad (32)$$

$P(O, q_t = j|\Theta)$ はモデル Θ が時刻 t で状態が j であり O を生成する同時確率、 T は時間、 J は状態数である。連続ガウス分布を仮定した確率密度関数 $b_j(\tilde{c}_t)$ は次式で与えられる。

$$b_j(\tilde{c}_t) = \frac{1}{\sqrt{(2\pi)^M |\Sigma_j|}} \exp \left\{ -\frac{1}{2} (\tilde{c}_t - \mu_j)^t \Sigma_j^{-1} (\tilde{c}_t - \mu_j) \right\} \quad (33)$$

Σ_j は対角共分散行列、 σ_{mj}^2 はその対角成分 (m は次元、 j は状態) である。 μ_j は状態 j の平均ベクトル、 M は MFCC パラメータの次元数である。目的関数の ν についての微分の解が最適な ν であるので、微分方程式は次式ようになる。

$$\frac{\partial \Phi(\Theta', \Theta)}{\partial \nu} = \sum_{j=1}^J \sum_{t=1}^T \frac{P(O, p_t = j|\Theta') \frac{\partial b_j(\tilde{c}_t)}{\partial \nu}}{\partial b_j(\tilde{c}_t)} = 0 \quad (34)$$

これにより式 (16) から式 (34) は次のように変形される。

$$\sum_{j=1}^J \sum_{t=1}^T P(O, q_t = j|\Theta') \left[\sum_{m=1}^M \frac{1}{\sigma_{mj}^2} (\tilde{c}_{mt} - \mu_{mt}) \cdot (-c_{mt}) \right] = 0 \quad (35)$$

この式から伸縮係数 ν^* を抜き出すと次式が得られる。

$$\nu^* = \frac{C \log \left(\exp \left(C^{-1} \sum_{j=1}^J \sum_{t=1}^T P(O, q_t = j|\Theta') \right) \right)}{\sum_{j=1}^J \sum_{t=1}^T P(O, q_t = j|\Theta')}$$

$$\left[\sum_{m=1}^M \frac{1}{\sigma_{mj}^2} \mu_{mt} c_{mt} \right] \Bigg) \Bigg] \quad (36)$$

c_{mt} は時刻 t における MFCC パラメータ第 m 次元の成分である。アルゴリズムより求められる占有度数 $\gamma_t(j)$ は次式

$$\gamma_t(j) = \frac{P(O, q_t = j|\Theta)}{\sum_{j=1}^J P(O, q_t = j|\Theta)} \quad (37)$$

式 (37) を式 (36) へ代入すると

$$\nu^* = \frac{C \log \left(\exp \left(C^{-1} \sum_{j=1}^J \sum_{t=1}^T \gamma_t(j) \left[\sum_{m=1}^M \frac{1}{\sigma_{mj}^2} \mu_{mt} c_{mt} \right] \right) \right)}{\sum_{j=1}^J \sum_{t=1}^T \gamma_t(j) \cdot \left[\sum_{m=1}^M \frac{1}{\sigma_{mj}^2} c_{mt} \right]} \quad (38)$$

式 (38) を HMM の学習と認識のアルゴリズムに組み込む。 Δ ケプストラムについても、 Δ^2 ケプストラムに関しても同様に求めることができる。

4. 実験

4.1 比較実験

本節では本手法の定式化の妥当性を検証するため、従来手法との比較を行った。

実験条件は以下の通りである。

単語データ	: ATR データベース A セット
学習話者	: 男性 9 話者 (mau,mht, mnm,msh,mmy,mtk, mtm,mtt,mxm)
	: 女性 4 話者 (ffs,fms, fkn,fyn)
学習データ数	: 奇数番目 2620 単語
評価話者	: 男性 1 話者 (mms)
	: 女性 1 話者 (faf)
評価データ数	: 偶数番目 655 単語
サンプリング周波数	: 12 KHz
ハミング窓幅	: 20ms
フレーム間隔	: 10ms
特徴量	: MFCC13 次元
音響モデル	: 状態数 3, 混合数 3
共分散行列	: 対角共分散
初期モデル	: 男性 4 話者、女性 4 話者
初期モデル数	: 話者を組み替えた 3 種
辞書数	: 2620 単語

適応単語は音素バランス単語の偶数番目より認識で用いていない単語から選択したものを使用した。

図 1 では認識実験結果を適応単語数に分て示すと共に、提案手法と VTLN-R [1] による比較の結果を

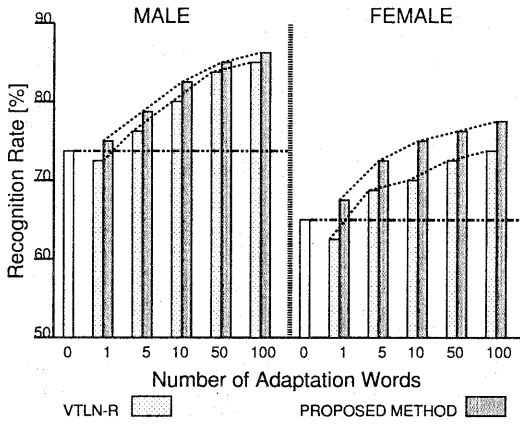


図1 VTLN-Rと提案手法の比較認識性能

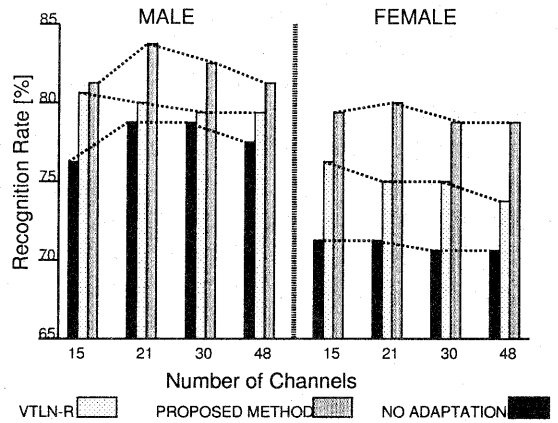


図2 チャンネル数に対する認識率の変化

示す。

VTLN-Rは全域通過フィルタを周波数伸縮関数として用いる手法であり、高速かつ高い認識効果を与える声道長正規化手法である。図1から分かるように提案手法は適応単語数に依らず、VTLN-Rよりも高い認識率を与えている。

4.2 チャンネル数の検討

本節では声道長正規化の効果を向上させるための予備実験として、伸縮係数 ν 推定の際に用いるMFCCのフィルタバンクのチャンネル数に関する検討を行った。実験条件は以下の通りである。

単語データ	: ATR データベース A セット
学習話者	: 男性 2 話者 (mau,mht) : 女性 2 話者 (ffs,ffms)
学習データ数	: 奇数番目 2620 単語
評価話者	: 男性 1 話者 (mms) : 女性 1 話者 (faf)
評価データ数	: 偶数番目 655 単語
サンプリング周波数	: 12 KHz
ハミング窓幅	: 20ms
フレーム間隔	: 10ms
特徴量	: MFCC13 次元
音響モデル	: 状態数 3, 混合数 3
共分散行列	: 対角共分散
初期モデル	: 男性 2 話者、女性 2 話者
辞書数	: 2620 単語

適応単語は音素バランス単語の偶数番目より認識で用いていない単語から選択した5単語を使用した。MFCC作成の際のチャンネル数を15、21、30、48と変更し実験を行った。

図2より、VTLN-Rを用いた場合はチャンネル数が少ない時に男性、女性両方のデータにおいて最も高い認識性能を与えている。VTLN-Rに対して本手法

を用いた場合、男性、女性共にチャンネル数21の時に最も良い認識性能が得られている。本手法は最大で約2ポイントの認識性能の差が現れた。これは、本手法がチャンネル数の違いによる影響が大きく、適切なチャンネル数を与えることでより高い適応効果が見込めることを示している。VTLN-Rとの比較から周波数領域での伸縮の違いが、チャンネル数を変えた際の適応効果と関係があることが確認された。

4.3 伸縮係数の推定精度

本節では伸縮係数を式(38)の推定式を用いずに直接操作することで、どのように認識率が変化するかを調べた。図3は男性話者データ mau の偶数番目データ 655 単語を用いた場合の認識率の変化の様子を調べたものである。これより、認識率が最大となる伸縮値は0.92付近にあると推定される。

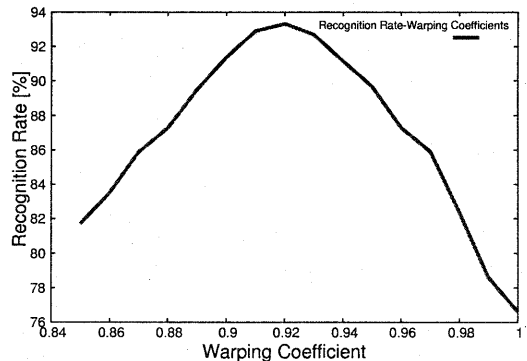


図3 伸縮係数に対する認識率の変化

次に提案手法を用いた場合、どのように伸縮係数が変化するかを示す。実験に用いたHMMモデルは

先の実験で用いた男女2話者計4話者のデータで作成した多数話者モデルを用いた。

実際に適応単語数が増えることで、どのように伸縮係数値が増え、いくかを調べたのが図4である。この図4から適応単語数が10単語を超えたあたりから急速に収束が始まり、100単語程度で伸縮係数値0.92付近に収束することが分かる。この結果は認識率最大を与える伸縮係数 ν と提案手法による尤度最大基準を求めたそれがほぼ一致していることを示している。

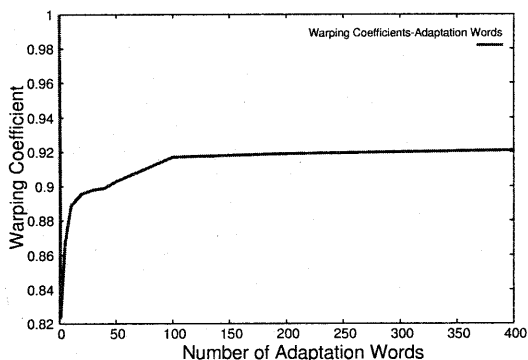


図4 適応単語数に対する伸縮係数の変化

5. むすび

本研究では最尤推定により求まるただ一つの声道長伸縮パラメータにより構成される線形変換によって話者適応を行う手法を提案した。本手法は少量の発話しか得られない環境において特に有効な手法であることを確認した。また、伸縮係数値を任意に変化させることにより本手法によって学習が認識率が最大となる伸縮係数値付近に収束することを確認した。今後は認識実験で用いる話者数を増やして評価を行う。さらに、雑音、伝達特性などの環境要因も含んだアルゴリズム構築を行い、実環境における有効性を評価する。

文 献

- [1] 江森正, 篠田浩一, “音声認識のための高速最尤推定を用いた声道長正規化,” 電子情報通信学会論文誌 DII, J83-DII, No.11, pp.2108-2117, November 2000.
- [2] 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.
- [3] E. Eide and H. Gish, “A parametric approach to vocal tract length normalization,” ICASSP96, Vol.1, pp.346-348, 1996.

- [4] B. Atal, “Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification,” J. Acoust. Soc. Am, Vol.55, pp.1304-1312, 1974.
- [5] L. Lee and R. C. Rose, “Speaker normalization using efficient frequency warping procedure,” ICASSP96, Vol.1, pp.353-356, 1996.
- [6] L. Welling and S. Kanthak and H. Key, “Improved methods for vocal tract normalization,” ICASSP99, pp.1436, 1999.
- [7] P. Zhan and M. Westohal, “Speaker normalization based on frequency warping,” ICASSP97, pp.1039-1042, 1997.
- [8] T. Anastasakos, J. McDonough, R. Schwartz and J. Makhoul, “A compact model for speaker-adaptive training,” ICSLP96, Vol.2, 1996.
- [9] H. Wakita, “Estimation of vocal tract shapes from acoustical analysis of the speech wave,” IEEE Acoust. Speech, Signal Processing, ASSP27, pp.281, 1979.
- [10] 服部浩明, 嵯峨山茂樹, “移動ベクトル平滑化話者適応の原理,” 信学技報 SP92, No.15, 1992.
- [11] 大倉計美, 杉山雅英, 嵯峨山茂樹, “混合連続分布 HMM を用いた移動ベクトル平滑化話者適応方式,” 信学技報 SP92, No.16, 1992.
- [12] Yunxin Zhao, “An EM Algorithm for Linear Distortion Channel Estimation Based on Observations from a Mixture of Gaussian Source,” IEEE Transactions on Speech and Audio Processing, Vol.7, No.4, July 1999.