

論文 / 著書情報  
Article / Book Information

|                   |                                             |
|-------------------|---------------------------------------------|
| 論題(和文)            | 識別誤り最小化による話者適応化法の検討                         |
| Title(English)    |                                             |
| 著者(和文)            | 松井 知子, 古井 貞熙                                |
| Authors(English)  | SADAOKI FURUI                               |
| 出典(和文)            | 1995年日本音響学会春季講演論文集, , No. 3-5-10, pp. 95-96 |
| Citation(English) | , , No. 3-5-10, pp. 95-96                   |
| 発行日 / Pub. date   | 1995, 3                                     |

◎ 松井知子 古井貞照

(NTT ヒューマンインタフェース研究所)

## 1 まえがき

不特定話者用の音韻 HMM による認識には限界があるため、少量の音声を用いて話者適応化を行う技術が盛んに研究されている。その代表的な方法としては、適応化音声に対する音韻 HMM の尤度が最大となるように、不特定話者用音韻 HMM のパラメータを適応化する方法 (maximum likelihood method: ML 法) がある。しかし、この ML 法では認識誤り率に関して、必ずしも最も良い性能が得られるとは限らない。そのため、識別誤りが最小となるように、音韻 HMM のパラメータを推定する方法 (minimum classification error method: MCE 法) が検討されている [1]。

一方、話者認識において、各話者の学習音声から各話者の HMM を作成する時に、まず ML 推定法を用いて HMM のパラメータを推定し、次に MCE 法を適用すると、認識精度の高い話者 HMM が作成できることが示されている [2]。

そこで本論文では、不特定話者音韻 HMM の話者適応化において、ML 法と MCE 法を組み合わせて用いる方法について検討する。

## 2 ML+MCE 話者適応化

本方法は、ML および MCE 法の二つの手続きからなる。

まず ML 法において、尤度関数が最大となるように、不特定話者音韻 HMM のパラメータを Baum-Welch アルゴリズムで適応化する。

次に、MCE 法 [2][3] においては、識別関数  $g_k(\cdot)$  と、識別誤り関数  $d_k(\cdot)$  をそれぞれ式 (1)、式 (2) で定義する。

$$g_k(X, A) = \log[\mathcal{L}_k(X, A)] \quad (1)$$

$$d_k(X, A) = -g_k(X, A) + G_k(X, A) \quad (2)$$

$$G_k(X, A) = \log\left\{\frac{1}{K-1} \sum_{j \neq k} \exp[\eta g_j(X, A)]\right\}^{\frac{1}{K}}$$

ここで  $X$  は観測ベクトル系列、 $A$  は HMM パラメータ集合、 $\mathcal{L}_k(\cdot)$  は、Viterbi アルゴリズムによる尤度関数である。そして、識別誤り関数  $d_k(\cdot)$  が最小となるように、segmental GPD (segmental generalized probabilistic descent) アルゴリズム [4] を用いて  $A$  を推定する。実際には、識別誤り関数  $d_k(\cdot)$  の代わりに、微分可能な損失関数  $l(d_k)$  (式 (3)) を定義し、この関数が減少するように式 (4) で  $A$  を逐次的に更新していく。

$$l(d_k) = \frac{1}{1 + \exp(-\phi(d_k + \psi))} \quad (3)$$

$$A_{t+1} = A_t - \epsilon_t \nabla l(d_k) \quad (4)$$

なお、 $\epsilon_t$  は更新量を調節する係数で、実験的に設定する。(4.3 節では、 $\epsilon_t$  の設定法について検討する。)

## 3 実験

実験で使った音韻 HMM は、混合分布数 256 の半連続型 HMM である。音韻 HMM は音韻環境独立の 43 種類である。不特定話者用音韻 HMM の作成には、ATR の日本語音声データベース B セットの男 5 名、及び日本音響学会研究用データベースの男 30 名の計 7,016 文章を用い、Baum-Welch アルゴリズムによって HMM パラメータの推定を行った。また、話者適応化には、ATR の日本語音声データベース C セットの男 2 名の文章データを用い、連続音声の中の音韻認識率により評価した。音韻認識実験では、発声内容の書き下しを与えて音声区間に対して Viterbi アライメントを取り、それを正解の音韻区間と仮定し、その音韻区間で全ての音韻 HMM のうち、最大尤度を示すものを認識結果とした [5]。音韻認識実験は、話者適応化に用いたものと同じ文章セットを用いた場合 (closed)、それとは異なる 100 文章 (open) を用いた場合について行った。特徴パラメータとして、標本化周波数 12kHz、フレーム長 32ms、フレーム周期 8ms、LPC 分析次数 16 でケプストラムを抽出した。

なお、ML 法による話者適応化では、学習の繰り返し回数は 5 回、また各混合分布の平均値だけを推定した。また、MCE 法では、繰り返し回数は 10 回、各混合分布の平均値と重み係数を推定した。両方法ともに、各繰り返しにおいて全学習データを適用した後に、バッチ方式で HMM パラメータを更新した。

## 4 結果

## 4.1 従来法との比較

図 1 に、不特定話者用音韻 HMM をそのまま用いた場合 (SI)、ML 法だけを用いて話者適応化した場合 (ML)、MCE 法だけを用いた場合 (MCE)、本方法 (ML+MCE) の音韻認識率を比較した結果を示す。なお、話者適応化には各話者 50 文章を用いた。

認識率は本方法、MCE、ML、SI の順であった。これらの結果から、本方法が有効であることがわかる。また、MCE 法は open、closed 両データに対して効果があるが、両データに対する効果の差が比較的大きい。

## 4.2 学習データ量との関係

図 2 に、ML 法、本方法による話者適応化において、50 文章の学習データを用いた場合と、10 文章を用いた場合の音韻認識率を示す。

\*A study on speaker adaptation based on minimum classification error criterion.  
By Tomoko Matsui and Sadaoki Furui (NTT Human Interface Laboratories)

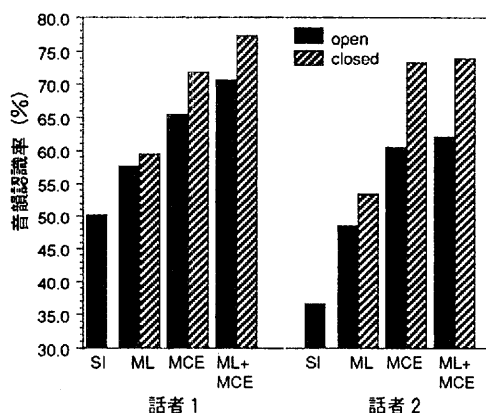


図 1: 従来法との比較

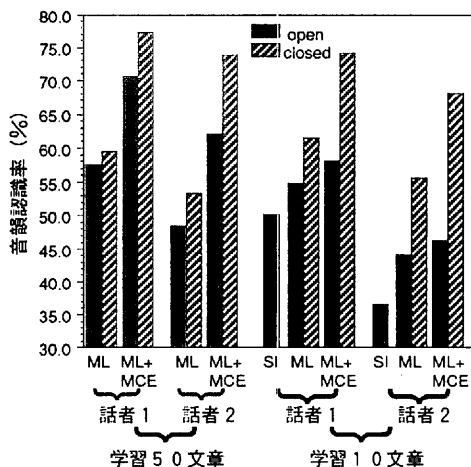


図 2: 異なる学習データ量による音韻認識率 (%)

特に open データに対して、学習データが 10 文章になると一般性が低下するため、MCE 法を適用する効果が少なくなる。

### 4.3 $\epsilon_t$ の設定法

4.1 節、4.2 節の実験では、式(4)の  $\epsilon_t$  の値を適当な値に固定していた。ここでは、学習を繰り返す度に  $\epsilon_t$  の値を適当に振り、学習データについて音韻認識実験を行って、最も高い認識率が得られた時の  $\epsilon_t$  の値に適応的に設定する。図 3 に  $\epsilon_t$  を固定した場合、適応的に設定した場合の音韻認識率を示す。なお、話者適応化には各話者 10 文章を用いた。

この図から、 $\epsilon_t$  を適応的に設定した方が高い認識率が得られることがわかる。

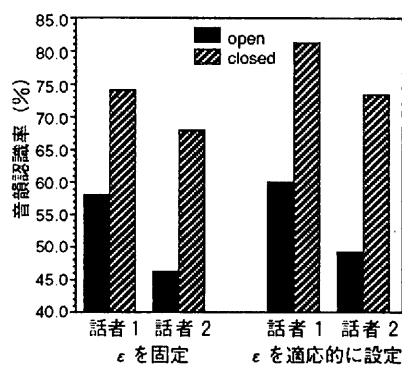


図 3: 異なる  $\epsilon_t$  の設定法による音韻認識率 (%)

## 5 まとめ

本論文では、不特定話者音韻 HMM の話者適応化において、ML 法と MCE 法を組み合わせる方法について検討し、それぞれ単独に用いる場合と比べて、高い音韻認識率が得られることを示した。話者適応化に 10 文章を用いた時、話者 2 名の平均で open データに対し、不特定話者音韻 HMM をそのまま用いた場合には音韻認識率 43.3 %、ML 法だけを用いた場合には 49.5 %であったが、本方法を用いることにより、音韻認識率 54.6 %が得られた。

今後は、話者数を増やして本方法の効果を確かめるとともに、学習データが少ない場合にも頑健に話者適応化できる方法について検討していく。

## 謝辞

熱心に討論して頂いた、ヒューマンインタフェース研究所古井特別研究室の方々に感謝致します。

## 参考文献

- [1] C.-H. Lin, P.-C. Chang, and C.-H. Wu, "An initial study on speaker adaptation for mandarin syllable recognition with minimum error discriminative training," *Proc. ICSLP, Yokohama*, pp. 1-307-310, 1994.
- [2] C.-S. Liu, C.-H. Lee, and W. Chou, "A study on minimum error discriminative training for speaker recognition," *J. Acoust. Soc. Amer.*, 97, 1, 1995.
- [3] B.-H. Juang and S. Katagiri, "Discriminative training," *J. Acoust. Soc. Jpn.*, 13, 6, pp. 333-339, 1992.
- [4] W. Chou, B.-H. Juang, and C.-H. Lee, "Segmental GPD training of HMM based speech recognizer," *Proc. ICASSP, San Francisco*, pp. 1-473-476, 1992.
- [5] 松岡達雄、P.-E. Gros、南泰浩、古井貞熙、"電話音声認識のための音響モデルの回線特性への適応化"、日本音響学会秋季研究発表会、pp. 1-103-104, 1994.