論文 / 著書情報
Article / Book Information

論題(和文)	話者適応化における学習語彙依存性の改善
Title(English)	
著者(和文)	篠田浩一, 渡辺隆夫
Authors(English)	Koichi Shinoda
出典(和文)	日本音響学会平成4年度秋季研究発表会講演論文集, Vol. 2-5-7, No. , pp.
Citation(English)	, Vol. 2-5-7, No. , pp.
発行日 / Pub. date	1992,

2—5—7　話者適応化における学習語彙依存性の改善 *

⊙ 篠田浩一　渡辺隆夫　(日本電気(株) C & C 情報研究所)

1　はじめに

不特定話者 HMM を初期モデルとした教師あり話者適応化では、適応化に用いる学習データが少量の場合、しばしば認識性能が劣化する。この原因の一つとして、少量の学習データ中の語彙に依存したパラメータが学習されてしまうことが挙げられる。これを、本稿では語彙依存性と呼ぶ。筆者らは先に、スペクトル内挿写像を用いた話者適応化法を提案し [1]、混合分布連続 HMM を用いた不特定話者認識システムに適用した [2]。本文では、上述の語彙依存性を改善する手法を提案するとともに、本手法をスペクトル内挿話者適応化と組み合わせ評価実験を行なった結果を報告する。

2　スペクトル内挿話者適応化法 [1,2]

サブワードを認識単位とする混合分布連続 HMM を用いた認識システムを考える。本適応化では、各分布の平均ベクトル μ_j の移動量を求める。この移動量を適応化ベクトルと呼び、Δ_j と書く。ここに、j は分布を表す添字である。

まず、学習データを用い、各分布の適応化ベクトル Δ_j を求める。他のパラメータを固定して平均ベクトルの Viterbi 学習を行う。Viterbi Alignment の過程では、同一状態内の各分布のうち出現確率の最も大きい分布が選ばれる。

学習データが極めて少量の場合、Viterbi 学習で学習されない、すなわち、対応する学習データのない、分布が存在する。これらの分布の適応化ベクトルを以下のスペクトル内挿により求める。

$$\Delta_i^B = \sum_j w_{ij} \Delta_j^A \qquad (1)$$

$$w_{ij} = \frac{\| \mu_i^B - \mu_j^A \|^{-m}}{\sum_{j'} \| \mu_i^B - \mu_{j'}^A \|^{-m}} \qquad (2)$$

ここで A は学習データありの分布を示す添字、B は学習データなしの分布を示す添字である。

*Normalization of Training Data Dependency in Speaker Adaptation
by K.Shinoda and T.Watanabe (NEC Corporation)

3　語彙依存性の改善

サブワードを認識単位とした HMM において、適応化に用いる学習データが少量の場合、学習用語彙の語彙コンテキストに依存したモデルが作成され、認識性能がかえって劣化する可能性がある。この語彙依存性を多数話者の発声を用いて補正することを考える。まず、多数話者の多数語彙の発声で不特定話者モデル M_{CI} を作成する。このモデルは通常の不特定話者モデルに相当し、学習用語彙に依存しないとみなすことができる。次に、多数話者の学習用語彙の発声を用いて不特定話者モデル M_{CD} を作成する。このとき、学習用語彙以外の条件はモデル M_{CD} 作成時の条件と同一にする。これら 2 つのモデルの違いは語彙コンテキストの違いのみを反映していると考えられ、M_{CD} から M_{CI} への写像を作成し、その写像を用いて教師あり適応化後のモデルを補正する。

Viterbi 学習の後、学習データの存在する分布について以下の処理を行なう (図 1)。

1. 多数話者の学習用語彙の発声データを用いて、学習用語彙の語彙コンテキストに依存した平均ベクトル μ_i^{CD} を学習する。初期モデルは不特定話者モデル M_{CI} を用いる。
2. 平均ベクトル μ_i^{CD} とモデル M_{CI} の平均ベクトル μ_i^{CI} のとの差ベクトル δ_i を求める。
$$\delta_i = \mu_i^{CI} - \mu_i^{CD} \qquad (3)$$
δ_i を分布 i の語彙補正ベクトルと呼ぶ。
3. 語彙補正ベクトルを、適応化ベクトルに加える。
$$\hat{\Delta}_i^A = \Delta_i^A + \delta_i^A \qquad (4)$$

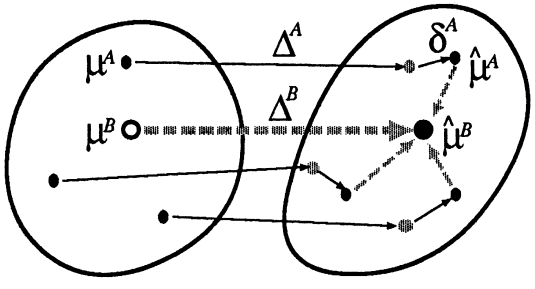

Figure 1: Normalization of Data Dependency

本手法は、事前に学習用語彙を特定する必要はあるものの、スペクトル内挿話者適応化法以外のその他の教師あり適応化法に対しても後置処理として用いることが可能である。

4 実験

4.1 実験条件

評価実験は半音節を認識単位とした混合ガウス分布 HMM[3] を用い、5000 単語を対象とした大語彙離散単語認識 [4] を行なった。混合ガウス分布数は 2 とした。多数話者のデータとして、男性 46 名女性 39 名計 85 名の音素バランスを考慮した 250 単語 1 回発声を用いた。また、評価用データとして、上の 85 名に含まれない話者男性 11 名女性 8 名計 19 名の、学習時とは異なる 250 単語 1 回発声を用いた。

分析条件は、サンプリング周波数 16 kHz、帯域 0.1–7.2 kHz、フレーム間隔 10 ms で、メルケプストラム分析を用いた。特徴ベクトルは正規化パワー差分、メルケプストラム 10 次元、メルケプストラムの変化量 10 次元の計 21 次元である。

半音節の種類は 241 個である。長母音と無音部が 1 状態で、その他の半音節は 4 個の状態をもつ。学習用データのある状態数の全状態数に対する割合は、学習用単語数 10 単語で約 1/3、50 単語で約 2/3 である。

不特定話者モデル M_{CI} は 85 名 250 単語発声のデータを用いて作成した。また、学習用語彙の語彙コンテキストに依存した不特定話者モデル M_{CD} は、M_{CI} を初期モデルとし、同じ 85 名のデータを用い、学習用単語数 10、20、30、40、50、100 の場合について、作成した。単語は M_{CI} 作成に用いた 250 単語中から選択した。

4.2 実験結果

適応化の初期モデルは不特定話者モデル M_{CI} を用いた。全評価話者平均の認識率を図 2 に示す。ここに CI はモデル M_{CI} による認識率、INTP はスペクトル内挿のみの従来の話者適応化法、INTP+NORM は内挿及び語彙補正を行う新適応化法の認識率を表す。また、単語数 30 のときの各評価話者ごとの認識率を図 3 に示す。横軸は従来法の認識率、縦軸は新適応化法の認識率である。従来法に比べ、認識性能が 10 単語から 50 単語にかけ平均で 1.4% 上がった。話者によるばらつきもほとんどなく、語彙補正の効果が確認できた。

5 おわりに

予め用意された多数話者の発声データを利用して、話者適応化の学習用語彙依存性を改善することを試み、良好な結果を得た。これまでに、多数話者の発声データを利用した話者適応化法がいくつか提案されている (e.g.[5, 6])。スペクトル内挿写像と語彙補正を組み合わせた新適応化法は、これらと比較して、周囲雑音・使用マイクなどの周囲環境が多数話者データ収録時と認識時とで異なっている場合に対しても適用が容易であるという利点がある。今後は、周囲環境の違う場合における新適応化法の有効性を調べたい。

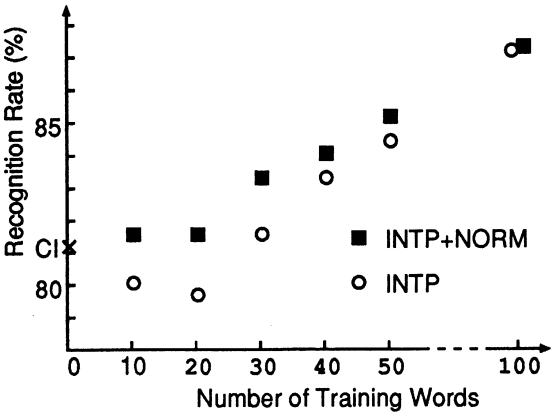

Figure 2: Experimental Result (1)

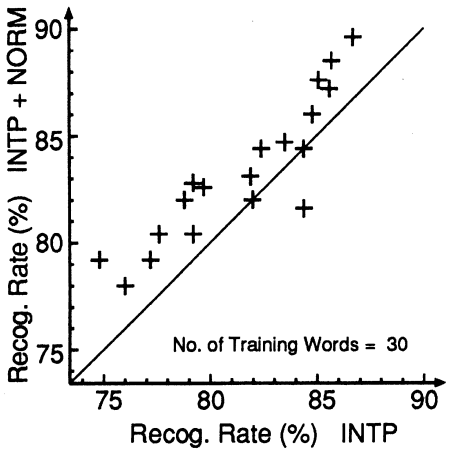

Figure 3: Experimental Result (2)

謝辞
日頃御指導いただく亘理部長、及び、御討論いただくメディア研の諸氏に感謝致します。

参考文献

[1]篠田他：音講論、1-8-12, (1990-9).
[2]篠田他：ICASSP91, S13.7, pp.857-860.
[3]磯谷他：音講論、3-5-13(1991-3).
[4]古賀他：音講論、2-P-5(1989-10).
[5]C.H.Lee 他：ICASSP90, S3.4, pp.145-148.
[6]松岡、鹿野：音講論, 1-1-6, (1992-3).