論文 / 著書情報
Article / Book Information

論題(和文)	半音節HMMによる音声認識のための話者適応
Title(English)	
著者(和文)	篠田浩一, 磯健一, 渡辺隆夫
Authors(English)	Koichi Shinoda
出典(和文)	日本音響学会平成2年度秋季研究発表会講演論文集, Vol. 1-8-12, No. , pp.
Citation(English)	, Vol. 1-8-12, No. , pp.
発行日 / Pub. date	1990,

1—8—12　半音節HMMによる音声認識のための話者適応 *

⊙篠田浩一　磯 健一　渡辺隆夫　(日本電気(株) C & C 情報研究所)

1　はじめに

これまでに半音節を認識単位とした連続確率密度HMMを用いることによって、比較的少ない学習データからでも、高精度の特定話者大語彙音声認識システムが構築可能であることが示されている [1, 2]。本稿では、さらに少ない発声でこのシステムを使用者に適応させる新しい話者適応化法を提案する。これまでに話者適応化法としては、教師あり適応化 [3, 4, 5] と教師なし適応化 [6, 7] が知られている。提案する手法は、教師あり適応化と教師なし適応化を組み合わせたものであり、適応化用発声には含まれていない半音節の適応化も可能にする。評価実験として大語彙離散単語認識実験を行い、方式の有効性を確認した。

2　話者適応化法

単一ガウス分布を出力分布とした連続確率密度HMMには3種類のパラメータがある。出力分布の平均ベクトル、出力分布の分散、遷移確率である。提案した話者適応化法では、そのうち平均ベクトルのみを適応化し、標準話者のHMMを未知話者に適応させる。

適応化データが少ない場合、全半音節のうち適応化データに存在しない半音節がある。適応化データに存在する半音節のHMMの状態の集合を A、適応化データに存在しない半音節のHMMの状態の集合を B とする。

集合 A に属する状態については、発声内容が既知の適応化データを用い、平均ベクトルを適応化する (教師あり適応)。適応化においては、出力分布の分散および遷移確率を固定し、HMM学習を行なう。 HMM学習にはBaum-Welchアルゴリズム、もしくはViterbiアルゴリズムを用いる。

また、集合 B に属する状態に対しては次のような教師なし手法を用いる。まず、集合 A に属する状態 $j \in A$ の適応化後の平均ベクトル $\hat{\mu}_j^A$ と適応化前の平均ベクトル μ_j^A の差のベクトル (適応化ベクトル)Δ_j^A を求める。

$$\Delta_j^A = \hat{\mu}_j^A - \mu_j^A. \tag{1}$$

次に、(2) 式にしたがい集合 B に属する状態の適応化後の平均ベクトル $\hat{\mu}_i^B$ を求める。

$$\hat{\mu}_i^B = \mu_i^B + \sum_j w_{ij} \Delta_j^A \tag{2}$$

(2) 式の w_{ij} は次の (3)、(4) 式で求める。

$$d_{ij} = \| \mu_i^B - \mu_j^A \|, \tag{3}$$

$$w_{ij} = \frac{d_{ij}^{-m}}{\sum_{j'} d_{ij'}^{-m}} \tag{4}$$

ここで、d_{ij} は平均ベクトル μ_i^B、μ_j^A 間の距離である。また、w_{ij} は μ_i^B の近傍 K 個の μ_j^A についてのみ計算する。m は平均ベクトル間の距離と重みとの関係を定める定数である。m が0に近づくと、すべての w_{ij} は一定値に近づく。一方 m が大きくなっていくと、d_{ij} が小さい状態 j の寄与が大きくなっていく。

3　実験

3.1　音声資料

話者は男性4名を用意した。1名 (A) を標準話者とし、その他3名 (B、C、D) を未知話者とした。発声データは、学習および話者適応化用に250語を用意し、評価用にはそれとは異なる250単語を用意した。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。評価実験は5000単語を対象とした大語彙離散単語認識実験を行なった。なお、この実験では効率化を計るため以下のような類似単語認識実験を行なった。すなわち、予め定められた音素間距離を用いて、各々の単語に対し全認識対象単語の中から類似の100単語を選出し、それらを認識対象として認識実験を行なう [8]。

半音節の種類は241個であり、そのうち長母音と無音部が1状態で、その他の半音節は4個の状

*Speaker Adaptation for Demi-Syllable Based Speech Recognition Using HMM.
by K.Shinoda, K.Iso, and T.Watanabe (NEC Corporation)

態をもつ。集合 A に属する状態数の全状態数に対する割合は、適応用データの単語数 10 単語で約 1/3、50 単語で約 2/3 である。

3.2 話者適応化

提案した話者適応化法の効果を調べる認識実験を行なった。標準話者 A、未知話者 B、適応化用単語数 N_w が 10 及び 50 のときの実験結果を Table 1 に示す。比較のため、分散及び遷移確率の学習も行なう通常の Baum-Welch アルゴリズムによる学習を行なった場合 (B-W)、教師あり話者適応化のみを行なった場合 (sup.ad.) の認識結果も併せて示した。教師あり適応化においては、Viterbi、Baum-Welch のアルゴリズムの違い及び学習の繰り返しの回数による認識率の違いがほとんどない。ここでは、教師あり適応化として、分散、遷移確率を固定した Viterbi アルゴリズムによる学習の繰り返し回数 1 回の学習を行なうこととした。また、教師なし適応化においては、計算する近傍数 K は 10 に、m は 1.0 に固定した。これらの K、m の値は予備実験により求めた最適値であるが、これらの値により認識率はほとんど変動しなかった。

実験結果をみると、Baum-Welch アルゴリズムに比べ、教師あり話者適応化を行なった方が認識率が良く、教師あり話者適応化のみよりも教師なし話者適応化を組み合わせた方が認識率が良いことがわかる。教師なし話者適応化の効果は適応化用単語数が小さい時に特に大きい。

次に上と同一の未知話者 (B) について、適応化用単語数を変化させたときの本手法による話者適応化後の認識率の変化を Figure 1 に示す。1 単語で認識率の改善がみられ、250 単語を用いると特定話者の認識率 (後述) とほぼ等しくなる。

次に 3 名の話者について、話者適応化 (sp.adp.) を行なったときの認識率を Table 2 に示す。標準話者は上と同じ話者 (A) である。単語数 N_w は 50 である。参考のため、適応前 (crs.sp.) の認識率と特定話者学習 (sp.dep.) を行なったときの認識率を合わせて示す。ここで、特定話者学習は Baum-Welch アルゴリズムでの 250 単語学習である。

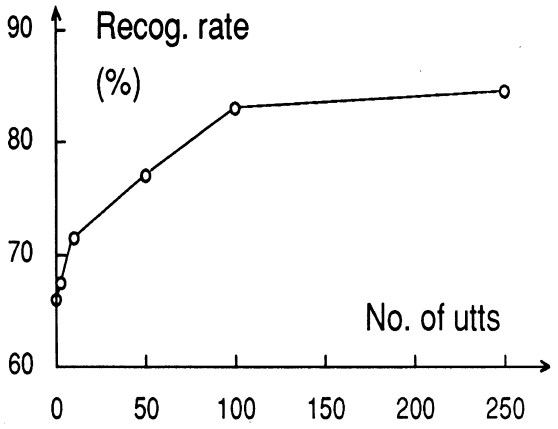

Figure 1: Recognition results vs. No. of utterances used for adaptation

4 終りに

連続 HMM に適用する新しい話者適応法を提案し、大語彙認識実験により有効性を確認した。本手法は、適応化データが少ない場合、特に効果がある。また、標準話者の発声を必要としないので、適応化用単語の変更が容易であるという利点がある。この手法は、話者適応化のみでなく、他の適応化、例えば、雑音下の発声を適応化するノイズ適応化などにも適用することが容易に可能である。

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]吉田他：音講論, 2-P-24(1988-10).
[2]渡辺他：信学論、J72DII、No. 8(1989)
[3]K.Shikano et al.: ICASSP-86, S49.5.
[4]R.M.Schwartz et al.: ICASSP-87, S15.3.
[5]平田, 中川：信学技報, SP90-16.
[6]古井：信学技報, SP88-21.
[7]松本他：信学技報, SP88-122.
[8]古賀他：音講論, 2-P-5(1989-10).

Table 1: Comparison of adaptation methods

	$N_w = 10$	$N_w = 50$
B-W	58.4 %	66.8 %
sup. ad.	62.4 %	73.6 %
this method	71.6 %	77.2 %

Table 2: Recognition results for 3 speakers

	B	C	D	Ave.
crs.sp.	66.0 %	63.2 %	67.2 %	65.5 %
sp.adp.	77.2 %	75.2 %	81.2 %	77.9 %
sp.dep.	84.4 %	86.8 %	92.8 %	88.0 %