論文 / 著書情報
Article / Book Information

論題(和文)	音声認識のためのタスク適応化
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年度春季研究発表会講演論文集, Vol. 1-P-15, No. , pp.
Citation(English)	, Vol. 1-P-15, No. , pp.
発行日 / Pub. date	1992,

1—P—15

音声認識のためのタスク適応化 *

⊙ 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

不特定話者音声認識においては、学習時のタスクと認識時のタスクが異なる場合、認識性能が劣化することが報告されている。この問題を解決するために、少数の話者の新しいタスクの発声を有効に用いて新しいタスク向けの不特定話者認識システムを構築する試みが、いくつか行なわれてきた[1, 2]。[1] では、少数話者の大量発声を用いて不特定話者のHMMを学習する手法について検討している。[2] は、連続HMMにおいて、予め対応づけられた特定話者モデルと不特定話者モデルを用いて、新しいタスクの特定話者モデルを不特定話者に適応化させる手法である。

本報告では、不特定話者認識システムにおいて、複数の参照話者の基準のタスクと新しいタスクの発声を用いて、システムを新しいタスクに適応化するタスク適応化の手法を提案する。提案手法は、以前報告した話者適応化の手法[3]を応用したものである。

2　タスク適応化法

本手法は、半音節を認識単位とした連続HMM認識システムに適用するタスク適応化の手法である。本タスク適応化の概念図をFigure 1 に示す。

まず、基準となるタスクの多数話者の発声を用いて、不特定話者HMM M_0 を作成する。基準のタスクは、すべての認識単位がなるべく均等に存在するように設計される。

次に、基準のタスク、新しいタスクそれぞれについて、複数の参照話者の特定話者HMM $M_A(p)$、$M_B(p)$、$p = 1, \ldots, P$ を作成する。ここで、P は参照話者数、A、B はそれぞれ、基準タスク、新タスクを表す添字である。作成は以下の手順で行なう。まず、参照話者の発声した、基準タスク及び新しいタスクの発声データを用意する。そして、M_0 を用いて、これらの発声データのセグメンテーションを行ない、データをHMMの各状態に対応づける。出力分布が混合分布の場合には、さらに状態内における分布まで同定する。次に、各分布に対応するデータの特徴ベクトルを平均し、その分布の平均ベクトルとする。分散、遷移確率などのその他のパラメータは M_0 のものを用いる。すべての参照話者の発声データに対し、同じ不特定話者HMM M_0 を用いてセグメンテーションを行なっているのは、各 $M_A(p)$、$M_B(p)$ の各分布の表す音響的特徴と、不特定話者モデルの対応する分布の表す特徴とを対応させるためである。

次に、$M_A(p)$、$M_B(p)$ を用いて、基準タスクから新しいタスクへの各分布の平均ベクトルの写像を作成する。今回は、対応する各分布の平均ベクトルを1対1に対応させることにより、写像を作成する。この写像を用いて、M_0 の各分布のHMMの平均ベクトル μ を変換し、新タスクの発声に適応化した平均ベクトル $\hat{\mu}$ を推定する。この手法をスペクトル内挿タスク適応化と呼ぶ。

$$\hat{\mu} = \mu + \sum_{p=1}^{P} w(p)(\mu_B(p) - \mu_A(p)), \quad (1)$$

$$w(p) = \frac{1/d(p)^{-m}}{\sum_{p'=1}^{P} 1/d(p')^{-m}}. \quad (2)$$

ここで、$\mu_A(p), \mu_B(p)$ はそれぞれ、$M_A(p)$、

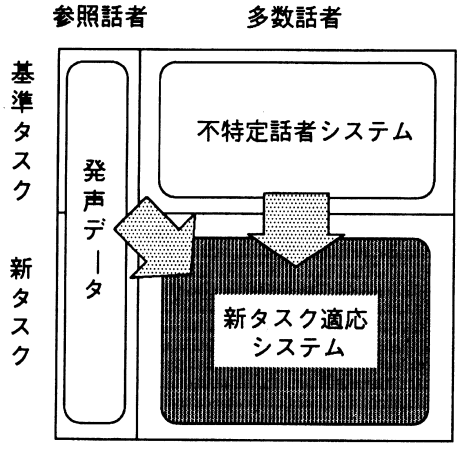

Figure 1: Task Adaptation

*Task Adaptation in Speech Recognition
by K.Shinoda and T.Watanabe (NEC Corporation)

$M_B(p)$ における当該分布の平均ベクトル、$d(p)$ は、μ と $\mu_A(p)$ の分散重みづけ距離である。m は適当な実数である。スペクトル内挿の概念図を Figure 2 に示す。

参照話者数が少ない場合、あるいは、参照話者1名あたりの発声データが少ない場合には、特定話者 HMM の学習が十分に行なわれないために、写像の精度が劣化する可能性がある。そこで、下式に示すように、不特定話者 HMM M_0 とタスク適応化後の HMM を混合し、新しい HMM を作成する。

$$\hat{\mu}' = k\hat{\mu} + (1-k)\mu. \qquad (3)$$

ここで、$0 < k < 1$ である。

3 実験

3.1 音声資料

基準のタスクは音素のバランスを考慮した250単語を用い、その多数話者データベースとして、男性46名の発声を用意した。新しいタスクは、連続数字とした。参照話者として3名の男性話者を用意し、各々の話者について、基準タスクの発声と新タスクの発声を用意した。新タスクの適応化用発声は1～5桁数字200発声である。また、別に評価用として、電子協連続数字データから、話者番号 No.1 ～ No.28 の男性話者の1桁及び4桁数字合わせて45文の4回発声を用意した。

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。

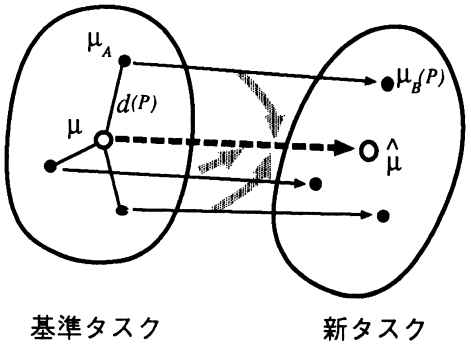

Figure 2: Spectral Mapping

3.2 実験結果

タスク適応化の効果を調べる認識実験を行なった。適応化及び認識実験には、半音節を認識単位として用いた不特定話者連続 HMM 音声認識システム [4] を用いた。出力分布は、2- ガウスの混合分布とした。不特定話者 HMM は男性46名の基準タスクの発声で学習した。式 (1) における m は1.0に固定した。(3) 式における k の値を4通りに変えて行なった認識実験の結果を Table 1 に示す。ここで、Standard SI は、不特定話者 HMM で認識した結果、Task Dependent SI は、参照話者3名の新しいタスクの発声を用いて、全パラメータを Baum-Welch 学習した HMM による認識結果である。

不特定話者の結果 84.21% に対し、タスク適応化の結果は、$k = 0.5$ のとき、86.04% と、認識率の改善が見られた。

4 終りに

タスク適応化手法を提案し、認識実験により有望であるとの見通しを得た。今後は、さらに、写像の作成方法に工夫を加えるとともに、平均ベクトル以外のパラメータの適応化も検討したい。

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]F.Kubala *et al.*: ICASSP-91, S13.1.
[2]D.J.B Pearce & L.C.Wood: ICASSP-91, S13.2.
[3]篠田他：音講論, 1-8-12(1990-9).
[4]磯谷他：音講論, 3-5-13(1990-3).

Table 1: Task Adaptation

Standard SI	84.21 %
Task Dependent SI	73.77 %
Task Adaptation(k = 0.25)	85.56 %
Task Adaptation(k = 0.50)	86.04 %
Task Adaptation(k = 0.75)	85.40 %
Task Adaptation(k = 1.00)	83.99 %