論文 / 著書情報
Article / Book Information

論題(和文)	テキスト指定形話者認識のための話者適応化法
Title(English)	
著者(和文)	松井 知子, 古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会1994年春季講演論文集, , , pp. 97-98
Citation(English)	, , , pp. 97-98
発行日 / Pub. date	1994, 3

3—7—5

テキスト指定形話者認識のための話者適応化法 *

⊙ 松井知子　　古井貞熙
(NTT ヒューマンインタフェース研究所)

1 まえがき

本論文では、テキスト指定形話者認識のための半連続型 HMM に基づく話者適応化法について検討する。テキスト指定形話者認識では、本人がシステム側から指定されたテキストを正しく発声したかどうかを判定するために、各話者ごとに音韻モデルを生成する。その際に、この音韻モデルにいかに正確に音韻性・話者性を持たせるかが重要な課題である。

これまで、連続型の不特定話者音韻 HMM を各話者の音声に適応化する方法について検討してきたが、話者性を十分に持たせることができなかった。このために、音韻独立話者 HMM から得られる結果を用いて、話者性を補う方法を提案してきた [1]。しかし、この方法には話者、テキスト照合を別々に行わなければならないという問題があった。また、不特定話者音韻モデルとして連続型 HMM を用いていたが、半連続型 HMM を用いた方が次の二点で優れていると思われる。第一に、半連続型 HMM では、全ての音韻モデルで平均・分散値が結びの状態になっているため、学習データ量が少ない時でもうまく HMM パラメータを推定できる [2]。第二に、音韻共通の混合分布に、テキスト独立形の話者認識 [3] で有効に使われる、全ての音韻に共通の話者性を反映させることができる。

ここで提案する方法 (phoneme-dependent/independent iterative training: PDI) では、不特定話者音韻モデルとして半連続型 HMM を用い、音韻依存／独立反復学習を通して、不特定話者音韻 HMM を各話者の音声に適応化する。その結果、学習データに含まれている数の少ない音韻のモデルについても、その話者の音声に適応化できる。

2 テキスト指定形話者認識法

テキスト指定形話者認識法 [1] では、まず各話者の学習データを用いて、各話者の音韻 HMM を生成する。認識時には、システム側から指定したテキストに従って、各話者ごとにその話者の音韻 HMM を連結し、そのテキストのモデルを生成する。次いで、入力音声をそのテキストのモデルに与えた時の尤度を計算し、話者およびテキストの判定を行なう。

テキストに応じて連結した HMM の尤度の値は、テキストや発声時期の違いによって大きく変動する。このことは、話者・テキスト照合において、全ての音声に共通な閾値を安定して設定することを難しくする。ここでは、尤度の値のばらつきを事後確率に基づく尤度正規化法 [4] によって正規化する。

3 話者適応化法 PDI

話者認識では、学習データを大量に必要とする方法は現実的ではない。各話者ごとに、少量の学習データから音韻モデルを生成しようとすると、出現頻度の低い音韻が生じ、そのような音韻のモデルについてはうまく話者の音声に適応化させることは難しい。出現頻度の高い音韻のモデルから、低い音韻のモデルを効果的に推定する方法は今のところない。推定誤りによる性能劣化を避けるために、ここでは、全音韻について平均、分散値に加え、重みも結びの状態にした学習を行うことによって、推定すべきパラメータ数を減らす。これにより、出現頻度の低い音韻のモデルについても、推定誤りを小さくすることを考える。学習データの分布だけを用いるテキスト独立形話者認識 [3] がうまくできることから、学習データの分布をその話者の全音韻の分布であると仮定し、それをより正確に表現することが有効であると考える。

PDI では、図1で示すように、音韻依存／独立学習を一系列で繰り返し行う。学習データが少ない時にモデルパラメータを頑健に推定する方法として、削除補間法 [5] がよく知られている。この削除補間法の場合、音韻依存、独立学習は別々に行われることとなり、両者を補間することは難しいと思われる。PDI では、音韻依存学習は従来の音韻モデルの話者適応化に相当し、音韻独立学習は半連続 HMM の混合分布の重みも全ての音韻で結びにした学習で、全ての音韻に共通な話者性を反映させる。その効果として、出現頻度の少ない音韻のモデルについても、その話者の音声に適応化できる。

音韻依存学習では、まず学習テキストに従い、不特定話者音韻 HMM を連結し、学習音声を与えて、全音韻で結びの平均値、および各音韻ごとの重みを推定する。この学習では、学習データに含まれない音韻のモデルについては、積極的には適応化されない。

音韻独立学習では、各話者の半連続型 HMM の混合分布を用いて、1状態の HMM を生成する。この1状態 HMM は、全音韻について重みも結びの状態にしたモデルであると言える。ここでは、重みの初期値は各分布に共通な値に設定した。それに学習音声を与え、平均値と重みを推定する。そして、各話者の混合分布の平均値だけをその1状態 HMM の平均値で置き換える。この学習では、出現頻度の少ない音韻のモデルについても、その話者の特徴量分布に適応化される。

4 実験

使用したデータは、男10名、女5名が約10カ月に渡る5時期に発声した文章データである。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。学習には、ある1時期に発声した10文章（1文章長は平均約4秒）を用い、テストでは、それとは異なる4時期に発声した各5文章を1文章づつ判定に用いた。音韻数は 41 である。半連続型 HMM として、3状態、256混合分布の HMM を用いた。また、比較のために、3状態、4混合分布の連続型 HMM についても実験を行った。いずれの場合においても、HMM パラメータは Baum-Welch アルゴリズムを用いて推定した。

*Speaker adaptation method for text-prompted speaker recognition.
By Tomoko Matsui and Sadaoki Furui (NTT Human Interface Laboratories)

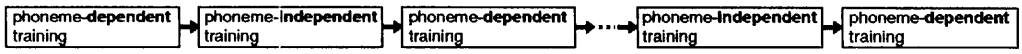

図 1: PDI のブロック図

表 1: 話者・テキスト照合平均誤り率 (%). [ ]: 尤度正規化法を用いた場合

時期	連続型	半連続型	*PDI*
T1	2.1 [2.2]	1.1 [1.6]	0.8 [0.9]
T2	2.5 [1.8]	3.6 [1.0]	3.0 [0.3]
T3	4.7 [2.3]	5.0 [1.7]	4.4 [1.0]
T4	1.5 [4.6]	1.5 [1.6]	1.1 [0.9]
Average	2.7 [2.7]	2.8 [1.3]	2.3 [0.7]

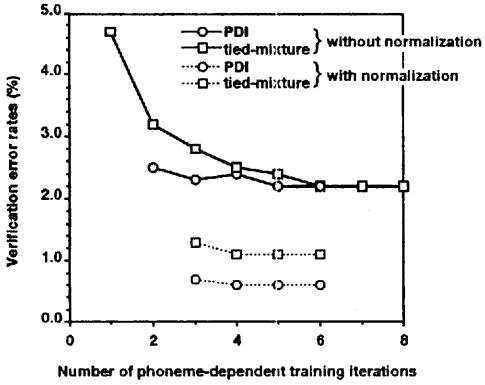

図 2: PDI vs. tied-mixture.

本実験では、本人が異なるテキストを発声した音声も、棄却すべき音声として含めて認識対象とした場合の話者・テキスト照合率で評価した。尤度の値に閾値を設けて、その閾値よりも尤度の大きい入力音声は、本人が正しいテキストを発声した音声として受理し、それ以外は棄却する。閾値は、各話者ごとに、2 種類の誤り（詐称者受理率、本人棄却率）が等しくなる値に事後的に設定した。

5 結果

表 1 に話者・テキスト照合誤り率を示す。“連続型”、“半連続型”はそれぞれ連続型、半連続型 HMM を用い、従来の音韻依存学習によって各話者の音韻モデルを生成する方法のことである。[ ] は、尤度正規化法を用いた場合の照合誤り率である。全ての方法において、音韻依存学習は 3 回繰り返した。（PDI では、音韻依存、独立学習は交互に行われるので、全学習回数は (音韻依存学習回数 × 2 − 1) 回である。）これらの結果は、PDI が連続型、半連続型 HMM による従来の方法に比べて有効であること、特に尤度正規化法を用いた場合に優れていることを示している。

図 2 は、音韻依存学習の適用回数を変えた場合について、PDI と半連続型の性能を比較した結果を示している。この図から、特に尤度正規化法を用いた場合に PDI の方が有効であることがわかる。

Viterbi アライメントを取って、各音韻モデルごとにその平均尤度を調べた結果、PDI を用いた場合、従来の方法に比べて、出現頻度の低い音韻の尤度が大きくなることが確認できた。このことは、出現頻度の低い音韻のモデルについても、適応化がうまくできていることを表している。

尤度正規化法では、申告話者のモデルの尤度と他の話者のモデルの尤度との相対値が用いられる。従来の方法では、上述のように、出現頻度の低い音韻のモデルの尤度は小さく、また、各音韻の出現頻度は話者によって異なるので、その相対値はテストデータによってばらつく。PDI では、出現頻度の低い音韻のモデルの尤度もある程度大きいので、その相対値はばらつかず、安定して閾値が設定できるため、尤度正規化法による効果が大きいと思われる。

6 まとめ

本論文では、テキスト指定形話者認識のための半連続型 HMM に基づく話者適応化法 PDI を提案した。PDI と従来の話者適応化法の性能を比較し、PDI の方が効果的であることを示した。PDI では、学習データに含まれている数の少ない音韻のモデルについても、その話者の音声に適応化できる。また、話者、テキスト照合を同時に行うことができる。話者 15 名、約 10 カ月に渡る 5 時期のデータを用いて評価した結果、99.3% の話者・テキスト照合率が得られた。

今後は、出現頻度の高い音韻のモデルから、低い音韻のモデルを推定するために、各音韻の特徴量空間の幾何学的な構造について調べていきたい。

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] 松井知子、古井貞熙、“連結音韻モデルによる話者認識、”信学技報、*SP92-128 (1993)*

[2] X. D. Huang, “Phoneme Classification Using Semi-continuous Hidden Markov Models,” *IEEE Trans. ASSP. Vol. 40, No. 5, pp. 1062-1067 (1992)*

[3] 松井知子、古井貞熙、“VQ、離散／連続 HMM によるテキスト独立形話者認識法の比較検討、”信学技報、*SP91-89 (1991)*

[4] 松井知子、古井貞熙、“テキスト指定形話者認識のための事後確率に基づく尤度正規化法、”日本音響学会秋季研究発表会、*1-7-20*、山梨 *(1993)*

[5] F. Jelinek and R.L. Mercer, “Interpolated estimation of Markov source parameters from sparse data,” *in Pattern Recognition in Practice, E.S. Gelesma and L.N. Kanal, Eds. Amsterdam, The Netherlands: North-Holland, pp. 381-397 (1980)*