

論文 / 著書情報
Article / Book Information

論題(和文)	連結音韻HMMによる話者認識
Title(English)	
著者(和文)	松井 知子, 古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	1992年電子情報通信学会秋季大会, , , pp. 1-265-266,
Citation(English)	, , , pp. 1-265-266,
発行日 / Pub. date	1992, 9
URL	http://www.ieice.org/jpn/books/t_g.html
権利情報 / Copyright	本著作物の著作権は電子情報通信学会に帰属します。 Copyright (c) 1992 Institute of Electronics, Information and Communication Engineers.

SA-6-2

連結音韻 HMM による話者認識

Speaker Recognition Using Concatenated Phoneme HMMs

松井知子

Tomoko MATSUI

古井貞熙

Sadaaki FURUI

NTT ヒューマンインタフェース研究所

NTT Human Interface Laboratories

1 まえがき

テキスト依存形話者認識では、あらかじめ認識に用いる言葉(キーワード)を決めておくため、それに固有な音韻情報も利用することができ、一般的にテキスト独立形話者認識に比べて性能が高い。しかし、実際にカードや鍵の代わりに、音声を用いて自動的に話者を認識することを考えると、キーワードが固定されている場合、テープレコード等によって話者の声を録音し、認識装置の前で再生すれば、それを受理してしまう危険性がある。従って、信頼性の高い話者認識を実現するためには、キーワードを可変に設定することができ、本人がそのキーワードを正しく発声した時にだけ受理するような機構が必要である。

従来、10数字や複数の単語をキーワードとし、認識のためにそれらのキーワードを新たな順序で並び変えた系列を指定して、それを話者に発声させるという方法が試みられてきた[1][2]。しかし、電子化した記憶装置を用いれば、キーワードの任意の系列を容易に再生できるようになると思われ、そのような方法も安全性に欠ける。

本論文では、認識のために任意の文章(キー文)を指定することが可能な話者認識法を提案する。この方法では、各話者ごとに音韻クラスHMM(hidden Markov Model)を生成し、キー文の内容に合わせて、それら音韻クラスHMMを連結して、認識用の音声その連結HMMに与えた時の尤度によって、話者、および発声内容を判定する。この方法を使えば、極めて高い精度で話者認識ができ、しかも、本人の音声でも、録音音声の再生により異なる内容である場合には、高い割合でそれを棄却することができる。

2 話者認識法

ここで検討する方法では、各話者の音声を話者認識システムに登録する時に、各話者ごとに音韻クラスHMMを生成する。図1に各話者の音韻クラスHMMの学習手続きを示す。話者認識では、登録時に各話者に膨大な量の発声を要求するような方法は現実的ではない。そのために、限られた量の学習データから、如何にして各話者の音声に適合した音韻クラスHMMを作成するかが問題となる。ここではまず、各話者ごとにその話者の全ての学習データを用いて、テキストに独立な話者モデル(1状態混合ガウス分布

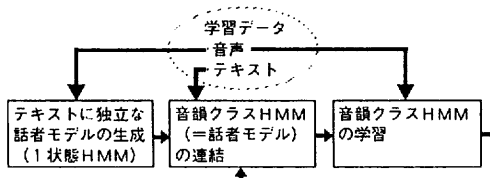


図1: 各話者の音韻クラスHMMの学習手続き

のHMM)を作成する。次にこの話者モデルを、その話者の各音韻クラスHMMに共通な初期モデルとし、学習データを再度用いて、そのテキストに合わせて音韻クラスHMMを連結する。連結学習[3][4]で各ガウス分布の重みのみを再推定することによって、各話者の音韻クラスHMMを学習する。ただし、音韻クラスの数と、その種類についてはあらかじめ決めておく。

図2に認識の手続きを示す。認識では、話者認識システムが指定したキー文(キー文は登録時には発声しなかった文章でも可)のテキストに合わせて、音韻クラスHMMを連結し、その連結HMMに対する入力音声の尤度を計算する。この尤度の値に基づいて、話者の判定を行う。



図2: 認識の手続き

3 評価実験

使用したデータベースは、男23名、女13名が約6カ月に渡る3時期(時期A、時期B、時期C)に発声した文章データで構成されている。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、時期Aに発声した10文章(1文章長は平均4秒)を用い、テストでは、時期B、時期Cに発声した5文章を1文章ずつ用いた。各音韻クラスHMMは、1状態、64混合分布の無相関連結HMMで表現した。

ここで提案した方法を、音韻クラス数を変えて、次の三つの点から評価した。

第一は、話者照合率である。尤度の値に閾値を設けて、その閾値よりも大きい入力音声は本人の声として受理し、それ以外は棄却した。この閾値は、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。評価者は女5名、詐称者は評価者5名を含む女13名である。

第二は、話者識別率である。評価は、男10名、女5名を上記のデータベースから無作為に選んで行った。認識の判定としては、計15名の話者の中から、最も類似度(尤度)の大きい話者を選択した。

以上の実験は、各話者が正しい文章を発声した場合について行った。

第三は、本人がキー文とは異なる文章を発声した時の棄却率である。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、本人がキー文を発声した音声として受理し、それ以外は棄却する。この閾値は、各話者ごとに、本人がキー文を発声した音声棄却されることの

ないように事後的に設定した。評価は、話者識別実験と同じ15名の音声を用いて行った。

4 実験結果

4.1 音韻クラス分類の効果

表1に話者照合の平均誤り率を示す。音韻クラスHMMを利用することによって、音韻クラスに分けない場合（音韻クラスに共通のHMM）に比べ、平均誤り率はほぼ2/5になる。

表1: 話者照合誤り率 (%)

音韻クラス数	1	25
時期B	5.3	2.0
時期C	7.8	3.3
平均	6.6	2.7

話者識別誤り率は、音韻クラス数を変えても変化なく、時期Bのテストデータに対する平均誤り率は1.3%、時期Cに対しては8.0%、平均4.7%であった。

4.2 話者照合における尤度正規化の効果

話者照合において、尤度の値のばらつきを効果的に正規化する方法について検討した。従来、Higgins[5]らにより、認識すべき音声を本人以外の話者のモデルに与えて、その音声と本人以外の話者のモデルとの類似度（本論文の場合はHMMの尤度）を計算し、その類似度を使って、本人のモデルに対する類似度の値を正規化する方法が試みられてきた。

ここでは、本人を含め登録話者全てについて、音韻クラス数が1の、テキストに独立な話者モデル（1状態、64混合分布HMM）に対する尤度を計算し、上位n名の平均尤度を、キー文のテキストに応じて本人の音韻クラスHMMを連結したモデルに対する尤度から差し引くことによって、正規化を行った。表2に、音韻クラス数が1の場合、尤度の正規化を施した時の話者照合平均誤り率を示す。この実験では、評価に男23名、女13名の計36名の話者のデータを用いた。尤度の値に適切な正規化を施せば、話者照合誤り率がほぼ一桁小さくなる。

表2: 話者照合における尤度の正規化の効果 (%)

正規化	なし	あり					
上位n位	-	1	2	3	5	10	36
時期B	2.8	0.3	0.1	0.2	0.5	0.8	3.1
時期C	3.4	0.3	0.2	0.0	0.2	0.4	2.5
平均	3.1	0.3	0.2	0.1	0.4	0.6	2.8

4.3 異文章の棄却性能（尤度正規化なし）

表3に、連結HMMに対する入力音声の尤度を直接用いた場合に、本人がキー文とは異なる文章を発声した音声を棄却できる割合（棄却率）を示す。音韻クラス数を25に増やしても、本人がキー文とは異なる文章を発声した音声は、半分程度しか棄却することができない。これは、尤度の値が、認識すべき音声のテキストや収録時期の違いに

よって大きく変動することに原因がある。全ての音声に共通な閾値をうまく設定することは難しい。

表3: 異なる文章の棄却率 (%)

音韻クラス数	6	15	25
時期B	41.3	63.7	69.0
時期C	37.7	54.0	60.7
時期B+時期C	24.3	39.7	48.3

そこで次に、尤度の値のばらつきを効果的に正規化する方法について検討した。

4.4 異文章の棄却における尤度正規化の効果

正規化は、音韻クラス数が1の、テキストに独立な話者モデル（1状態、64混合分布HMM）に対する尤度を差し引くことによって行った。表4に、音韻クラス数が25の場合、尤度の正規化をした時、本人がキー文とは異なる文章を発声した音声の棄却率を示す。尤度の正規化を施した場合、棄却誤り率はほぼ半分になる。

表4: 異なる文章の棄却における尤度正規化の効果 (%)

正規化	なし	あり	差
時期B	69.0	82.0	+13.0
時期C	60.7	81.3	+20.6
時期B+時期C	48.3	80.7	+32.4

5 むすび

認識のために任意の文章を指定することが可能な話者認識法について検討した。尤度の値を正規化することにより、キー文とは異なる文章の棄却率は約80%、話者照合率は99.9%を得た。今後は、安定した個人性情報を豊富に含む音韻クラスHMMを重視した方法など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] A. E. Rosenberg and F. K. Soong, "Recent Research in Automatic Speaker Recognition," in *Advances in Speech Signal Processing*, S. Furui and M. M. Sondhi (Ed.), Marcel Dekker, New York, pp. 701-738 (1992)
- [2] J. M. Naik, "Speaker Verification: A Tutorial," *IEEE Communications Magazine*, 28, 1, pp. 42-48 (1990)
- [3] 丸山活輝、花沢利行、川端豪、鹿野清宏、"HMM音韻連結学習を用いた英単語音声の認識," 信学技報、SP88-119 (1988)
- [4] 南泰浩、松岡達雄、鹿野清宏、"不特定話者連続音声データベースによる連結学習HMMの評価," 信学技報、SP91-113 (1992)
- [5] A. Higgins, L. Bahl, and J. Porter, "Speaker verification using randomized phrase prompting," *Digital Signal Processing*, 1, pp. 89-106 (1991)