

論文 / 著書情報
Article / Book Information

論題(和文)	合成HMMによる環境適応型話者認識法
Title(English)	
著者(和文)	菅野 智仁, 松井 知子, 古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	1995年日本音響学会春季講演論文集, , No. 2-5-13, pp. 55-56
Citation(English)	, , No. 2-5-13, pp. 55-56
発行日 / Pub. date	1995, 3

△ 菅野智仁[†] ○ 松井知子[†] 古井貞熙^{††}

([†] 東京工業大学 ^{††} NTT ヒューマンインタフェース研究所)

1 まえがき

実環境下で話者認識を行うためには、話者ごとに環境に適応したモデルを生成する必要がある。これまで、話者モデルと雑音モデルを近似的に合成したモデルを用いると、雑音下において、高い話者認識性能が得られることが報告されている [1]。一方、音声認識において、音韻 HMM と雑音 HMM を NOVO (NOise&VOice) 変換によって合成した HMM は、耐雑音性に優れていることが報告されている [2]。そこで本論文では、テキスト独立型話者認識において、話者 HMM と雑音 HMM を NOVO 変換によって合成した HMM を用い、その有効性について検討する。

さらに、従来の NOVO 変換では、HMM を合成する時に SN 比が既知である必要があった。本論文では SN 比が未知の場合にも有効な方法、SN 比選択型適応化法を提案する。

2 テキスト独立型話者認識

ここでは、話者認識法としてテキスト独立型話者認識を用い、各話者ごとにその学習データから話者 HMM を生成する。認識時には、任意のテキストの発声に対する話者 HMM の尤度を求め、話者の判定に用いる。なお、HMM の尤度は、入力テキスト、発声時期、雑音などによって変動するため、値のばらつきを事後確率に基づく尤度正規化法 [3] によって正規化する。

3 NOVO 変換による HMM 合成

NOVO 変換 [2] によって、雑音を含まない音声データから生成した話者 HMM と、雑音データから生成した雑音 HMM の出力確率分布を重畳した出力確率分布を有する合成 HMM を生成する。実際には、両分布をケプストラム領域からスペクトル領域に変換し、スペクトル領域で両分布を足し合わせ、ケプストラム領域に再変換して求める。この NOVO 変換では、両分布を SN 比に応じた重みで足し合わせるため、SN 比が既知である必要がある。

4 SN 比選択型適応化法

SN 比選択型適応化法 (選択法) は、SN 比が未知の場合に用いる方法である。まず各話者ごとに、複数の SN 比で話者 HMM と雑音 HMM を NOVO 変換した合成 HMM を用意する。認識時には、各話者について、入力音声に対するこれら複数の合成 HMM の尤度を計算し、その最大値を各話者の尤度とする (図 1)。

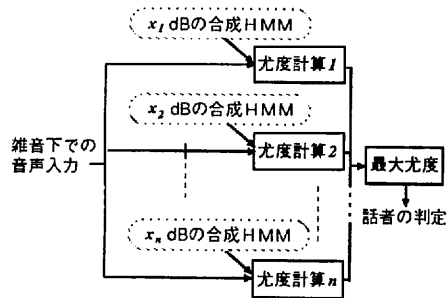


図 1: 各話者 HMM に対する SN 比選択型適応化法の手続き (x_i : 話者 HMM + 雑音 HMM の SN 比)

表 1: 雑音とその特徴

雑音の種類	特徴
エレベータ	足音や話し声
人混み	足音や話し声 (全体に定常的な白色系の雑音)
計算機室	定常的な白色系の雑音
公衆電話	時間的変化の激しい自動車の排気音
走行車内	定常的な白色系の雑音や自動車の排気音
展示会場	スピーカの声や人のざわめき

5 実験

5.1 条件

使用した音声データは、男 10 名が約 6 カ月に渡る 3 時期に、静かな環境で発声した文章データである。文章テキストは、ATR 連続音声テキスト 503 文 (平均 1 文章長は 4.2 秒) から抜粋した。長い無音区間は自動的に取り除いて用いた。話者 HMM の学習には、ある 1 時期に発声した 10 文章を用いた。雑音 HMM の学習には、電子協の雑音データベースから、表 1 の雑音データをそれぞれ 2 分程度切り出して使用した。テストでは、学習とは異なる 2 時期に発声した 5 文章を 1 文章ずつ用い、その音声データに、雑音データを計算機上で SN 比 0dB から 18dB まで 6dB ごとに重畳した雑音重畳音声を使用した。なお、テキストは学習用とテスト用で異なり、テストデータを作成する時に用いた雑音データは、雑音 HMM の学習に用いた区間は異なる区間から切り出した。特徴パラメータとして、標準化周波数 12kHz、フレーム長 32ms、フレーム周期 8ms、LPC 分析次数 16 のケプストラムを抽出した。話者 HMM としては状態数 1、混合ガウス分布数 64 の HMM を、雑音 HMM としては雑音の種類に拘らず、状態数 1、単一ガウス分布の HMM を用いた。

*Speaker recognition using HMM composition in noisy environments.

By Tomohito Kauno[†], Tomoko Matsui[†], and Sadaaki Furui^{††} ([†]Tokyo Inst. of Technology, ^{††}NTT Human Interface Labs.)

表 2: 種々の SN 比のテストデータに対する話者識別誤り率 (%)

雑音の種類	0dB			6dB			12dB			18dB		
	clean	合成 HMM	選択法	clean	合成 HMM	選択法	clean	合成 HMM	選択法	clean	合成 HMM	選択法
エレベータ	74.0	47.0	48.0	49.0	11.0	8.0	25.0	5.0	5.0	7.0	4.0	4.0
人混み	44.0	15.0	11.0	8.0	5.0	4.0	6.0	3.0	4.0	3.0	3.0	3.0
計算機室	82.0	25.0	11.0	55.0	11.0	5.0	20.0	7.0	4.0	6.0	3.0	3.0
公衆電話	27.0	16.0	15.0	11.0	8.0	6.0	6.0	5.0	5.0	3.0	3.0	3.0
走行車内	15.0	3.0	4.0	7.0	3.0	3.0	4.0	3.0	3.0	3.0	3.0	3.0
展示会場	50.0	12.0	9.0	20.0	5.0	4.0	7.0	2.0	2.0	4.0	2.0	2.0
平均	48.7	19.7	16.3	25.0	7.2	5.0	11.3	4.2	3.8	4.3	3.0	3.0

表 3: 種々の SN 比のテストデータに対する話者照合誤り率 (%)

雑音の種類	0dB			6dB			12dB			18dB		
	clean	合成 HMM	選択法	clean	合成 HMM	選択法	clean	合成 HMM	選択法	clean	合成 HMM	選択法
エレベータ	36.6	26.9	26.6	15.9	4.6	4.1	4.5	2.0	1.7	1.1	0.7	0.7
人混み	14.3	3.3	2.4	1.9	1.3	1.0	0.5	0.7	0.5	0.1	0.4	0.4
計算機室	22.5	4.8	5.3	9.2	2.5	1.8	2.2	1.3	1.1	1.2	0.8	0.8
公衆電話	10.4	3.2	3.5	1.7	0.8	0.7	0.6	0.2	0.1	0.1	0.1	0.1
走行車内	1.1	0.9	0.6	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1
展示会場	8.8	3.5	3.5	1.8	1.4	1.2	0.7	0.9	0.5	0.1	0.5	0.5
平均	15.6	7.1	6.9	5.1	1.8	1.5	1.4	0.9	0.7	0.4	0.4	0.4

実験は、

1. 雑音を含まないデータで学習した話者 HMM をそのまま用いた場合 (clean)
2. SN 比が既知として NOVO 変換による合成 HMM を用いた場合 (合成 HMM)
3. SN 比が未知として選択法を用いた場合 (選択法)

の三つの場合について行い、話者識別・照合誤り率で評価した。話者識別実験は 10 人から 1 人を選択する方法で行い、話者照合実験では 10 人のうち、本人以外を詐称者として用いた。なお、照合実験では、しきい値を本人棄却率と他人受理率が一致する値に事後的に設定した。また選択法では、SN 比 0dB から 18dB まで 3dB ごとに合成 HMM を用意した。

5.2 結果

識別誤り率を表 2、照合誤り率を表 3 に示す。テストデータとして、雑音を含まないデータを用いた場合、識別誤り率は 4.0%、照合誤り率は 0.2% であった。

これらの結果から、雑音の種類に拘らず、合成 HMM を用いると誤り率が小さくなるのがわかる。また、“3. 選択法”は“2. 合成 HMM”と比べて、どの SN 比のテストデータに対しても同等以上の性能を示した。特に SN 比が悪いほど、“3. 選択法”を用いる効果が高い。“3. 選択法”を用いれば、SN 比が未知であっても高い認識性能を得ることができる。

6 むすび

HMM 合成を行なうことにより、雑音環境下でも高い話者認識性能が得られることを示した。さらに SN 比未知の場合でも、SN 比選択型適応化法を用いることにより、十分高い認識率が得られることを示した。

今後は、実際に雑音下で収録した音声データを用いて、本方法の効果を調べるとともに、話者照合のしきい値を事前に設定する方法についても検討して行きたい。

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] R.C. Rose, E.M. Hofstetter, and D.A. Reynolds, "Integrated Models of Signal and Background with Application to Speaker Identification in Noise," IEEE Trans. Speech and Audio Processing, Vol. 2, No. 2, pp.245-257, 1994.
- [2] F. Martin, K. Shikano, Y. Minami, and Y. Okabe, "Recognition of Noisy Speech by Using the Composition of Hidden Markov Models", 信学会音声研究会資料, SP92-96, 1992.
- [3] 松井知子、古井貞照、 “テキスト指定型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、1993.