話者照合システム

6C－8

広野 亮　斎藤 博昭　中西 正和
慶應義塾大学大学院 理工学研究科 計算機科学専攻

1. はじめに

音声によって誰であるかを自動的に判定する話者認識に対する有用性が高まっている。これが実現できれば、車や建物の入出チェックにおける「音声キー」の実現が可能となるほか、音声メールやコンピュータのリモートアクセスなどにおいて、音声による本人確認ができるようになり、コンピュータネットワークが急速に社会に普及しつつある現在、音声認識の分野の中でも一層注目を集めている。

本研究では話者認識の中でも音声が特定の個人のものであるか判定する話者照合システムについて次に挙げる事柄に焦点を当てて研究を進めていく。

- 音声の個人特徴量の自動抽出
- 優れた確率モデル（HMM）の構築
- 単語特有の個性を利用したテキスト指定

2. 音声分析と個人特徴量の抽出

サンプリング

- DAT-LINK による A/D 変換 [7]。
- サンプリングレート 24000Hz。
- 量子化ビット数 16bit。

特徴量抽出

- FFT ケプストラム 16 次元。
- ハミング窓かけ。
- 平均スペクトルパワーによる無音区間の検出とその区間のスキップ
- フレーム長 25.6msec。
- フレーム周期 12.8msec。

無音区間の検出、除去については次の方法を用いた。各分析フレームの平均スペクトルパワーの履歴を 10 フレームまで記録することによって、

- 初めて閾値を下回る区間が 10 フレーム続いた
  → 10 フレーム前から最初の無音区間が開始。
- 無音区間検出後、初めて閾値を上回る区間が 10 フレーム続いた
  → 10 フレーム前から最初の有音区間が開始。
- 有音区間検出後、初めて閾値を下回る区間が 10 フレーム続いた
  → 10 フレーム前から 2 度目の無音区間が開始。

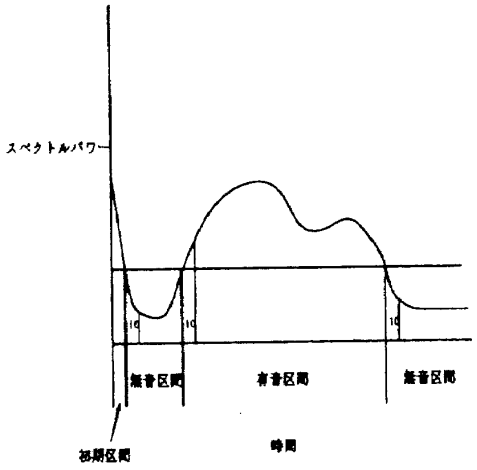

図 1: 無音区間除去

この方法には、録音開始時に生じるノイズ、誤った有音区間の検出を防ぐことを狙いとしたものである。

3. HMM 法

音声認識で用いられる HMM(Hidden Markov Model:隠れマルコフモデル）において特徴量ベクトルの出力確率が離散分布である場合にはシンボルは有限集合で出現確率は離散的なので、テーブル化して表現できる。しかし連続分布モデルを用いる場合には出現確率はガウス分布のような連続分布で与えられるので、出力シンボルが観測される度に出力確率を求めなければならない。離散分布の場合には音声の特徴パラメータは通常ベクトル量子化などの手段によって有限個のシンボルに変換される。しかし、ベクトル量子化の際に生じる量子化誤差が避けられず、話者認識等では問題になるため、出力確率分布は連続分布として扱われることが多い [3]。本研究では特徴量ベクトルが無相関正規分布に従うとした。学習データによる HMM のパラメータ再推定法は Baum-

---

Speaker Verification

Ryo HIRONO Hiroaki SAITO Masakazu NAKANISHI
Department of Computer Science, Faculty of Science and Technology, Keio University 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223, Japan

Welch(Forward-Backward) アルゴリズム [2] を用いた。

4. 話者照合における各手法

話者照合の方法は照合に用いる、即ち発声する言葉の種類によって次の方法に分けられる [1]。

テキスト依存型 照合に用いる言葉をあらかじめ決めておく方法。

テキスト独立型 どんな言葉を発声してもよい。

テキスト指定型 任意のテキストを照合のたびにシステムが指定する手法。

本システムでは登録者にはあらかじめ複数の単語を登録してもらい、照合時にはシステムがランダムにテキストを選択し、そのテキストをユーザーに発声してもらう。登録したテキスト全てが何であるか他人に洩れることがなければ、音声キーを盗むことは極めて困難になり、かつテキスト独自の個人特徴を照合に利用できる。

また本システムでは照合が困難な場合 (データのモデルからの出現確率が閾値に近く、誤った受理、棄却の恐れがある場合) に照合者のもっとも個性が出ているテキストをモデルから選択し、再照合を行なう。全 HMM の各パラメータの平均値を求め、1 つの HMM のパラメータとの自乗誤差をそのテキストのスコアとして計算し、最もスコアの高かったものを登録者の個性が一番表れているテキストとして選択する。

$$
\begin{aligned}
S_\pi &= \sum_i^N (\pi_i - \mu_{\pi_i})^2 \\
S_a &= \sum_i^N \sum_j^N (a_{ij} - \mu_{a_{ij}})^2 \\
S_m &= \sum_i^N \sum_j^M (m_{ij} - \mu_{m_{ij}})^2 \\
S_v &= \sum_i^N \sum_j^M (v_{ij} - \mu_{v_{ij}})^2
\end{aligned}
\qquad (1)
$$

式 (1) で $\pi_i, a_{ij}, m_{ij}, v_{ij}$ はそれぞれ一つの HMM の各状態、各次元の初期状態確率、状態遷移確率、無相関正規分布に従うと仮定した特徴量ベクトルの平均値、分散値であり、$\mu_{[]}$ はそれらの全 HMM の平均値である。また N は状態数、M は特徴量の次元数である。そして最終的に式 (2) をその HMM のスコアとした。

$$
S = S_\pi \times S_a \times S_m \times S_v \qquad (2)
$$

5. 実験方法

照合に使用した各単語 HMM モデルについては、登録話者 3 人、登録単語数 6 とし、10 個の学習用音声データを用い、10 回の再推定学習を行なう。4. 節で述べた再照合を行なった場合とランダムテキスト指定のみの場合の 2 つの方法について照合実験を行ない、各モデルに事後的に閾値を設定し、照合では登録者本人、詐称者 3 人による照合実験を行なう。

6. 評価方法

各方法での照合結果の正解確率、学習、照合時に生成した 2 次的なデータ、各学習済モデルについて ROC 曲線 [1] を比較検討することによって、本システムの性能を評価、考察する。

7. 結果、考察

現在、HMM による学習を終え、認識実験を繰り返しているが、信頼できる認識結果が得られていない。無音区間の除去、特徴量抽出は成功してると思われるが、HMM の認識精度に問題が生じている。また、テキスト固有のモデルのスコア比較による再照合の効果もはっきりしていない。特に問題があると思われるのが高サンプリングレートの設定による冗長なデータによる弊害、話者発声データの長期的変動の影響に認識システムが対応できていないと思われる。今後実験を重ね、種々の問題を統合して解決する必要がある。

参考文献

[1] 古井 貞熙「ディジタル音声処理」東海大学出版, 1985.

[2] 中川 聖一「確率モデルによる音声認識」電子情報通信学会, 1988.

[3] 野々村 行「音声認識システムの構築」慶應義塾理工学部学士論文,1993.

[4] 松井 知子「HMM による話者認識」信学技報,SP95-111,pp17-24, 電子情報通信学会,1996.

[5] 尾島 敬司「韻律情報を用いた音声認識」慶應義塾理工学部学士論文,1995.

[6] T.Matsui,K.Kanno and S.Furui 「Speaker recognition using HMM composition in noisy environments」,Proc.Eurospeech,pp.I-621-624,1995.

[7] DAT-Link http://www.tc.com/