

## 音源に関する特徴量を利用した話者認識

4D-4

松村 理恵子 岡本敏雄

電気通信大学大学院 情報システム学研究科

## 1 はじめに

話者認識の研究において、音声波に含まれる個人性情報の抽出はもっとも重要な課題の一つである。しかし、個人性情報は音韻性情報と深く関わりあっていて、両者を分離することは非常に困難である。このために、現状では発話内容に依存しない話者認識法はまだ確立されていない。

話者の認識は、発声の内容をあらかじめ決めておくテキスト依存型と任意の内容を発声するテキスト独立型に分けられる。個人性情報の現われ方は、発声するテキストの内容に非常に依存しているため、テキスト独立型の話者認識は依存型に比べてあまり良い認識率は得られない。しかし、法科学の分野などでは必ずしも同じ発声内容の資料を対照にすることは限らないため、テキスト独立型の話者認識を使用しなければならない場合が多い。

## 2 研究の目的

従来の話者認識に関する研究では、声道に関する特徴量であるスペクトル情報が用いられることが多く [2] [4] [5]、それと音源に関する特徴量を組合せた研究は少ない [1]。この要因としては音源に関する特徴量の抽出の難しさ、その適切な利用法の難しさなどがあげられる。しかし、音源に関する情報は話者の個人性に寄与するところが大きい。そこで本研究では、音源に関する特徴量に声道に関する特徴量を組み合わせて、テキスト独立型の話者認識への利用の有効性を示すことを目的とする。

## 3 実験方法

## 3.1 音声資料および分析方法

成人 10 名(女性 4 名)の話者から約 1ヶ月の間隔において母音と 10 秒程度の文章を採取し、音声資料とした。これらの音声資料から声道に関する特徴量とし

## Speaker Recognition Using Pitch Information

Rieko Matsumura, Toshio Okamoto

The Graduate School of Information System,  
University of Electro-Communications

1-5-1 Chofugaoka, Chofu-shi, Tokyo, 182 Japan

て FFT、LPC ケプストラム係数、音源に関する特徴量としてピッチを計算する。分析条件は 16bit、8kHz で A/D 変換した。ケプストラム係数は 1 次の適応プリエンファシスを施した後、フレーム長 20msec、フレーム周期 10msec でフーリエ分析、線形予測分析を行ない、14 次までの係数を使用した。ピッチはケプストラム係数と同じ条件でケプストラム法で抽出した。

## 3.2 母音抽出

母音を判断するにあたって、本研究では判別関数を用いる。発話文の時系列のケプストラム係数データに対して判別関数を適用し、数フレーム以上連続して同じ母音であると判断された部分を母音定常部とみなす。また、判別関数は母音それぞれを一つのカテゴリと考え、単独発声された母音の FFT ケプストラム係数、LPC ケプストラム係数から以下のようにして求める。

$$L_{\eta} = \sum_{\beta} \left( \sum_{\alpha} \omega^{\alpha\beta} \mu_{\alpha\eta} \right) x_{\beta} - \frac{1}{2} \sum_{\beta} \sum_{\alpha} \omega^{\alpha\beta} \mu_{\alpha\eta} \mu_{\beta\eta} \quad (1)$$

ただし、 $\omega^{\alpha\beta}$  は級内分散行列の逆行列の各要素、 $\mu_{\alpha\eta}, \mu_{\beta\eta}$  はケプストラム係数  $x_{\alpha}, x_{\beta}$  についての群  $\eta$  における平均を表す。こうして判別すべき個体について、

$$L_{\eta} + \log_e \pi_{\eta} \quad (2)$$

を全ての群について計算し、その最大値を示す群に所属していると判別する。ここに、 $\pi_{\eta}$  とは、前個体数中に占める群  $\eta$  に属する個体数の割合を意味し、 $\log_e$  は自然数を表す。

本研究では閾値を設定し、閾値以下ならば母音以外であるというカテゴリを定義し、6 カテゴリで母音の判断を行なった。

## 3.3 話者認識

本研究では話者の認識法として話者識別を行なう。登録されている個人の音声データを標準パターン、識別するために入力した音声を入力パターンと呼ぶ。標準パターンのケプストラム係数とピッチを  $c_i, p_i$ 、入力パターンのケプストラム係数とピッチを  $\bar{c}_i, \bar{p}_i$ 、 $\alpha$  を重みとしたとき、標準パターンと入力パターンとの距離を以下のように定義する。

$$d = \alpha d_C(c_1 \cdots c_{14}, \bar{c}_1 \cdots \bar{c}_{14}) + (1 - \alpha) d_P(p, \bar{p}) \quad (3)$$

ここで、 $d_C$  と  $d_P$  はそれぞれ

$$d_C = \sum_{i=1}^k (c_i - \bar{c}_i)^2 \quad (4)$$

$$d_P = (p - \bar{p})^2 \quad (5)$$

とする。 $k$  はケプストラム係数の分析・使用次数である。このようにして求めた距離の値がもっとも小さいものを識別結果とする。

## 4 認識実験結果

話者の識別を行なう際に、標準パターンとして単独発声した母音、入力パターンとして時期差をおいて採取した文章を用いた。

### 4.1 重みの決定

式(3)の $\alpha$ を決定するために $\alpha$ の値を変えて話者の識別実験を行なった。結果を図1に示す。ここで、 $\alpha = 1.00$ のときはケプストラム係数のみ、 $\alpha = 0.00$ のときはピッチのみで識別を行なうことを意味する。

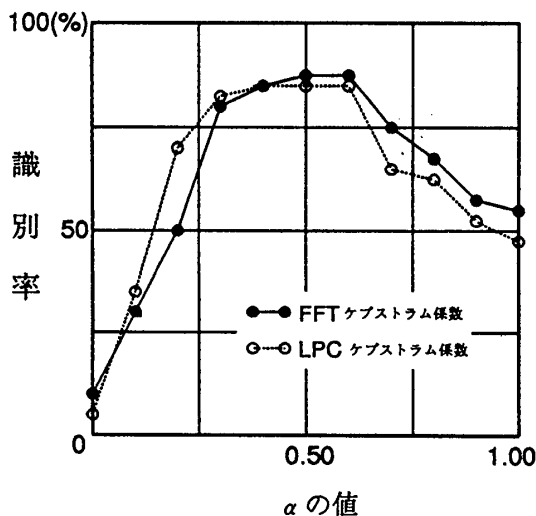


図1: 重み $\alpha$ と識別率の関係

識別率はFFTケプストラム係数の場合は $\alpha$ が0.60のときにもっとも良い値をとっており、LPCケプストラム係数の場合は0.40、0.50、0.60とほとんど変化しない。FFTケプストラム係数とLPCケプストラム係数を比較すると、FFTケプストラム係数を用いて識別を行なったほうが若干良い結果が得られた。

### 4.2 ピッチ計算を利用する話者認識実験

発話文章から母音を抽出する際に、まず最初にピッチを計算することによって文章中の有声部分を正確に抽出することができる。このことは母音の抽出の正当

率を向上させることにつながる。ここでは、最初にピッチを計算するという手順を用いて話者の識別実験を行なった。比較のためにピッチ計算を用いないで母音判断を行なう手順の識別実験結果も併せて表1に示す。ここで $\alpha$ は前実験で結果の良好であった値 $\alpha = 0.60$ を用いる。

表1: ピッチ計算を利用した識別結果

	ピッチ計算を利用した方法	ピッチ計算を利用しない方法
FFTケプストラム	86.4	84.5
LPCケプストラム	84.0	82.3

識別実験を行なった結果、ピッチの計算を利用するとわずかではあるが識別率が向上した。

## 5 むすび

本研究では声道に関する特徴量と音源に関する特徴量を組み合わせて発話内容に依存しない話者の識別を行なった。その結果、声道に関する特徴量であるケプストラム係数のみで話者の識別を行なうよりも、音源に関する特徴量を組み合わせて話者の識別を行なう方法が、より効果的であることがわかった。また、発話文章において最初にピッチを計算することによって文章中の有声部分の抽出、母音の判断をより正確に行なうことができた。このことによって話者の識別率も向上した。今後は発声時間との関係についての実験を行なう予定である。

## 参考文献

- [1] 松井知子・古井貞熙: "音源・声道特徴を用いたテキスト独立型話者認識", 信学論 J75-A,4,pp.703-709,1992
- [2] 中川聖一・坂本光範: "音声認識・話者認識のためのFFTケプストラムおよびLPCケプストラムの検討", 信学論 J66-A,12, pp.1199-1206,1983
- [3] 新美康永: "音声認識", 共立出版,1979
- [4] 長内隆・谷本益巳: "単独発声単語中の母音を用いた話者照合", 音講論,2-3-8(1990-3)
- [5] 長内隆・谷本益巳: "判別関数を用いた話者照合", 音講論, 1-7-5(1992-10)
- [6] 安田三郎・海野道郎: "社会統計学", 丸善,1977
- [7] 松村理恵子・岡本敏雄: "発話内容に依存しない話者認識のための特徴量の比較検証", 信学技報 SP-97, Vol.95, No.431, 1995