

嵯峨山 茂樹 板倉 文忠
(日本電信電話公社 武蔵野電気通信研究所)

1 はじめに

従来の音声の個人性情報の研究では、音声の静的かつ全体的な特徴を用いることが多かった。しかし、部分的あるいは動的な特徴にも個人性は存在すると考えられる。このような観点から、動的特徴に対応するパラメータを導入し、個人性について調べたので報告する。

2 音声パラメータの平均変化係数

音声を一定周期のフレームごとに分析するとパラメータ(例:ケプストラム)の時系列 $\{\theta(t)\}$ が得られる。音声のゆるやかな変化を表現するため、次のような新しいパラメータを導入する。

時刻 t (フレーム番号)を原点にとり直して、 θ の局所的な動きを直線:

$$\hat{\theta}(i) = ai + b \quad (1)$$

で、重みつき最小2乗法により近似する。すなわち、重みを $\{w_i\}$ ($w_i \geq 0$)として、

$$E = \sum_{i=-n}^n w_i (\theta(i) - ai - b)^2 \quad (2)$$

を最小にする a, b を選ぶ(Fig.1)。このときの直線の傾き a と、この時刻における θ の平均変化係数と呼ぶことにする。なお、 b は θ の移動平均になっている。

w_i が対称($w_i = w_{-i}$)ならば a, b は、
$$a = \frac{\sum i w_i \theta(i)}{\sum i^2 w_i}, \quad b = \frac{\sum w_i \theta(i)}{\sum w_i} \quad (3)$$
により簡単に求まる。

$\theta(t)$ は比較的ゆるやかに動くが、種々の擾乱が重畳して、単純な差分ではゆるやかな動きが捉えられない。 a 係数

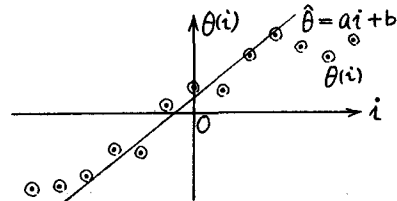


Fig.1 Least squares fitting of a line.

はそのような欠点を補うものである。

3 音声ケプストラムの動的尺度

$\theta(t)$ が何であっても上述の a, b 係数がフレームごとに定義されるが、特にケプストラムの場合は、 a 係数は線形な伝送系の影響を受けないという利点を持つ。(なぜなら、線形な伝送系の影響はケプストラムに和の形で現れるが、式(3)の a では、その成分は消えてしまう。)

また、ケプストラムの各次について求めた a 係数($a_i(t)$ と書く)の2乗和:

$$D(t) = \sum_{i=1}^{\infty} \{a_i(t)\}^2 \quad (4)$$

は、時刻 t におけるlogスペクトルのゆるやかな動きの度合に対応する。この量を動的尺度と呼ぶことにする。

4 特徴点の抽出

Fig.2に、動的尺度のプロット例を示す。ここでは、LPCケプストラムを12次まで用いて $D(t)$ を近似的に求めている。重み $\{w_i\}$ は $w_0 \sim w_4 = 1, 0.8, 0.6, 0.4, 0.2$ である。図の上側はパワパターン、下側は動的尺度パターンである。音韻境界の部分に動的尺度がピークを持つことが観察される。ピークは最も過渡的な点、またピーク間の谷底は最も定常的な点に対

* On Individuality in a Dynamic Measure of Speech, by Shigeki SAGAYAMA and Fumitada ITAKURA, Musashino Electrical Communication Laboratory, NTT.

応すると考えられる。話者識別実験にはこれらの点を特徴点として用いた。

5 ケプストラム平均変化係数の分散分析

9名の男声話者が各3回発声した /namae/ (名前) について、3箇所の過渡部のピーク点における、LPCケプストラムの各次成分の α 係数の個人性について分散分析を行った結果をTable 1に示す。同時に、log area ratio (LAR) について式(3)により求めた α 係数に関する結果も示す。いずれも、過渡部分の変化係数にはかなりの個人性のあることがわかる。

6 動的特徴を用いた話者識別実験

9名の男声話者が3回ずつ、3か月ごとの4時期に発声した /namae/ をサンプルとして、話者識別実験を行った。用いたパラメータは /am,ma,ae/ に対応する動的尺度のピーク点(過渡部)における12次のLPCケプストラムの α 係数および /a,m,a,e/ に対応する動的尺度の谷底点(定常部)における β 係数である。 α のみ、 β のみ、および α 、 β 両方を用いる、3種類について特徴ベクトルを構成し、単純なユークリッド距離最小の原理で、2種類の条件で識別を行った。

実験(I): 第 k 時期の l 番目の各話者の発声を標準パターンとして、その他のすべてのサンプルを識別する。これを $k=1, \dots, 4$; $l=1, \dots, 3$ についてすべて行う。

実験(II): 第 k 時期のサンプルを識別するのに、 k 以外の3時期から各1発声ずつ採りその3つの平均を標準パターンとする。

これらの結果の話者識別率をTable 2に示す。これより、過渡部の動的情報は定常部の静的情報と同程度の個人性情報を持つと考えられる。

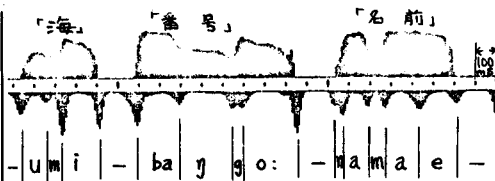


Fig.2 Power pattern and dynamic measure plot.

Table 1 Analysis of variance applied to dynamic feature of speech.

parameter order	cepstrum w=1.,.6,.2			log area ratio w=1.,.8,.6,.4,.2		
	am	ma	ae	am	ma	ae
1	**	**	o	**	**	o
2		**	**	o	**	
3		**	o	**	**	
4	*	**	**	**	**	**
5	**	**	**	*	**	**
6	**	**		**	**	**
7	**	**	o	**	**	**
8	o	**		*	*	**
9	**	**	**	**	**	
10	o	*	**			**
11		**	*	**	**	
12	**	**				**

** : significant at 0.5 %, * : 1.0 %, o : 2.5 %.

Table 2 Results of talker identification experiments.

parameters	experiment I	experiment II
a /am,ma,ae/	81.6 %	92.6 %
b /a,m,a,e/	81.4 %	91.7 %
a & b	82.5 %	93.5 %

7 結論

音声の動的特徴に対応するパラメータを導入し、分散分析や話者認識実験を通じて、動的情報の中に多くの個人性情報があることがわかった。この特徴量は(線形な)伝送系の影響が除去できるなどの利点を持つため、単独では強力でなくとも従来の話者認識パラメータを補強する意味で有用であると期待される。今後は更に動的特徴の扱い方を考えると共に伝送系の影響について調べてゆきたい。

謝辞

日頃御指導戴く野田基礎部長、小池第四研究室長に感謝する。また、この研究に多大な指導・助言を賜、た斎藤収三東京大学教授に深謝する。