

論文 / 著書情報
Article / Book Information

論題(和文)	スペクトルパタンの時間変動特性を考慮した話者認識
Title(English)	
著者(和文)	古井 貞熙, 板倉 文忠
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会1972年講演論文集, , , pp. 229-230
Citation(English)	, , , pp. 229-230
発行日 / Pub. date	1972, 10

2-2-20 スペクトルパタンの時間変動特性を考慮した話者認識*

古井貞熙 板倉文忠 (日本電信電話公社 武蔵野電気通信研究所)

1. まえがき 音声のスペクトルパタンは、種々の時間変動を伴っている。その変動特性は、各話者固有の発声器官の、運動上の特徴および制約等を反映していると考えられる。それらを取扱う一つの方法として、いくつかの単語の有声区間について、部分自己相関関数 (PARCOR) による分析を行ない、パラメータの平均値と、標準偏差、およびパラメータ間相関を用いた話者の認識を試みた。

2. 分析 静かな部屋で録音された音声を、8 KHz, 11 bit でサンプリングおよび量子化し、10 msec 間隔で分析を行なった。分析時間長は 32 msec とし、ハミングウィンドウを乗じてから、PARCOR パラメータを用いて分析した。パラメータとして、ピッチ周期と、1~12 次の PARCOR を用い、13 次元ベクトル $x_t = (x_{1t}, x_{2t}, \dots, x_{13t})$ とする。逐次的に 1~12 次の PARCOR を導出した後の予測誤差の相関関数のピークによって、ピッチ周期の抽出を行なうとともに、ピークの高さに閾値を設けて、有声区間の抽出を行なう⁽¹⁾。単語における全有声区間を用いて、平均値 μ_i 、標準偏差 σ_i 、相関行列 A を求める。

$$A = \begin{bmatrix} \lambda_{ij} \end{bmatrix} \quad : 13 \times 13 \text{ 行列}$$

$$\lambda_{ij} = \frac{\sum_{t=1}^N (x_{it} - \mu_i)(x_{jt} - \mu_j)}{\sigma_i \sigma_j} \quad N: \text{有声区間の数}$$

単語として /bakuon/, /namae/, /bango:/ の 3 種類のサンプルを用いて実験を行なう。

3. 認識方法 3ヶ月おきに 4 時期にわたって録音したサンプルを既知サンプルとし、その最終のサンプルから 3ヶ月後のサンプルを未知サンプルとする。各話者は、各時期に 3 回ずつ発声している。従って、各話者の既知サンプルは、 $3 \times 4 = 12$ サンプルである。未知サンプルは、話者識別の場合、 $3 \times 9 = 27$ サンプル、話者照合の場合、受入れべき本人のサンプルが 3 サンプル、棄却すべき詐称者のサンプルが (3 × 詐称者数 (8 又は 26))

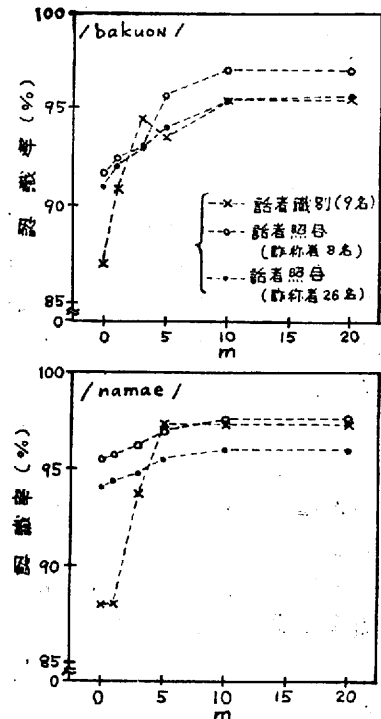


図1. 6. A のうちの認識に用いる成分の数 m を変えたときの認識率

表1. $m = 20$ のときの認識率

単語	話者識別 (9名)	話者照合	
		詐称者 8名	詐称者 26名
/bakuon/	95.4 %	96.9 %	95.6 %
/namae/	97.2 %	97.6 %	95.9 %
/bango:/	96.3 %	96.8 %	95.9 %
平均	96.3 %	97.1 %	95.8 %

*Talker Recognition Using Dynamic Characteristics of the Spectral Pattern.

By Sadaoki Furui and Fumitada Itakura (Musashino E.C.L., N.T.T.)

サンプルである。異なる4時期の未知サンプルについて、同様の実験を行なう。

既知サンプルを用いて、 θ および Λ の各成分について、話者の効果に関する分散分析を行ない、下記の大きい m 成分を選択する。これを μ の成分と組合わせて、特徴ベクトル Z ($13+m$ 次元) とする。未知サンプル Z と、話者 k の標準パターン (既知サンプルの代数平均値) θ との距離 D を次のように定める。

$$D(Z, \theta) = (Z - \theta) V^{-1} (Z - \theta)^t$$

V : 時期的変動に関する共分散行列
(話者間で平均化したもの)

この距離にもとづいて、話者識別 (9名) および話者照合 (9名の各話者に対し、詐称者8名または26名) を行なう。(発声者はいずれも成人男子)

4. 結果 図1に、2種類の単語について、 θ, Λ のうちの認識に用いる成分の数 m を変えたときの、認識結果を示す。異なる4時期の未知サンプルと、9名の話者に関する平均認識率を示す。 θ および Λ のうちの、話者効果の大きい10~20成分を用いることにより、 $m=0$ 、すなわち平均値のみを用いる場合に比較して、誤り率が $2/3 \sim 1/4$ に減少している。

3種類の単語について、 $m=20$ としたときの認識率は表1の通りである。単語による差はほとんどなく、いずれもよい認識率を示している。2種類の単語について、話者識別の Confusion Matrix と、話者照合 (詐称者26名) の話者別誤り率を、表2と図2に示した。

5. むすび 単語を用いて、PARCOR パラメータによる分析を行ない、平均値と、そのまわりでの時間変動特性 (標準偏差とパラメータ間相関) を考慮することにより、よい話者認識率が得られ、この方法の有効性が確認された。話者識別の Confusion Matrix などが単語によって異なることから、2種類以上の単語を組合わせることにより、さらにより認識率を得ることができると思われる。
[謝辞] 日頃御指導頂く、当研究所長 静基徳研究部長、各藤才4研究室長、ならびに研究室の諸氏に感謝する。

表2. Confusion Matrix
($m=20$ のときの話者識別)

/bakuon/

	W	I	F	K	H	S	M	U	T	認識率
W	12									100%
I		12								100
F			12							100
K				9		3				75.0
H					12					100
S						12				100
M							10	1	1	83.3
U								12		100
T									12	100
平均認識率										95.2

/namae/

	W	I	F	K	H	S	M	U	T	認識率
W	12									100%
I		12								100
F			12							100
K				12						100
H					9	3				75.0
S						12				100
M							12			100
U								12		100
T									12	100
平均認識率										97.2

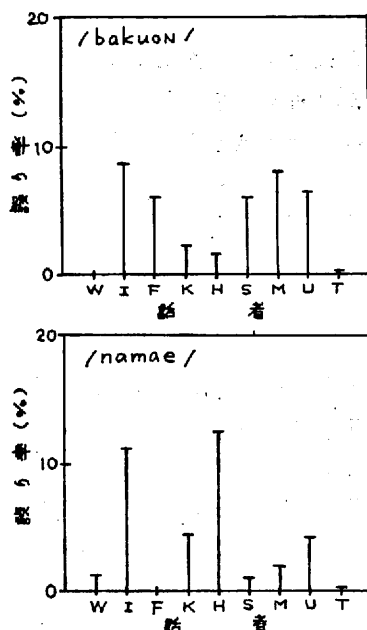


図2. 話者別誤り率
($m=20$ のときの話者照合
— 詐称者 26名 —)

[文献] (1)板倉 各藤: 音響学会講演論文集 2-2-6 (昭44-10)