



# HOKKAIDO UNIVERSITY

Title	音声スペクトルパラメータの時間パターンを用いた話者照合に関する研究
Author(s)	小倉, 久宜; Ogura, Hisayoshi; 廣瀬, 秀治 他
Citation	北海道大學工学部研究報告, 101, 61-69
Issue Date	1980-12-25
Doc URL	<a href="https://hdl.handle.net/2115/41649">https://hdl.handle.net/2115/41649</a>
Type	departmental bulletin paper
File Information	101_61-70.pdf



## 音声スペクトルパラメータの時間パターンを用いた話者 照合に関する研究

小倉 久宜\* 廣瀬 秀治\*\* 柄内 香次\* 永田 邦一\*

(昭和55年6月30日受理)

### Speaker Verification Using Time Patterns of Speech Spectrum Parameters

Hisayoshi OGURA, Hideharu HIROSE, Koji TOCHINAI and Kuniichi NAGATA

(Received June 30, 1980)

#### Abstract

Utterances of the spoken Japanese digit 4 /yoN/ were analyzed by linear prediction, and some statistical features of the time series of predictor coefficients  $\alpha_k$ , PARCOR coefficients  $K_k$  and log area ratios  $L_k$  ( $k; 1 \sim 10$ ) which represent spectrum time patterns of utterances were investigated. And speaker verification experiments have been made using these series.

Changes on spectrum parameter coefficients of utterances between utterances and ones which were uttered 1 year later, were also investigated.

#### 1 ま え が き

話者照合は登録話者の真の音声であるか、詐称者の音声であるかを話者の登録音声と照合し、判定するもので、バンキングサービス、クレジットサービス等本人を確認する必要がある場合の一手段として用いることが可能である。音声スペクトルの時間パターンのマッチングにより話者認識を行う試みは既に多数報告されている。<sup>(1)~(5)</sup> 本研究は音声スペクトルの時間パターンを用いた話者照合に関するもので、特定単語4 /yoN/について同一話者が複数回発声した場合の音声スペクトルパラメータ係数の時間パターンの分布について行った調査結果について述べ、この結果よりほぼ同一時期に発声された音声を用いた話者照合を行う場合には各時刻における話者の登録された標準音声パターンと被験音声パターン間の距離を計算するのに、そのパラメータの分散に対応した距離荷重を行うことにより話者照合結果が向上することを実験により導いている。<sup>(6)(7)(8)</sup>

また話者の標準音声パターンを作るのに用いた音声の発声時期と被験音声発声時期の間隔が長くなると、同一話者音声でも標準音声、被験音声のパターン間距離が大きくなり良好な照合結果が得られないことが知られている。<sup>(4)(9)</sup> ここでは少数の話者について、特定単語4 /yoN/に対する話者の発声時期のずれが各種の音声スペクトルパラメータに対して与える影響の調査、な

\* 電子工学科 電子機器工学講座

\*\* 山武ハネウエル株式会社

らびに対比を行っている。

## 2 資料音声と分析

特定の数字列の音声を登録して標準音声とし、同一数字列を発声した場合の音声は、登録話者音声であるかどうかを照合する場合を想定し、まず日本語数字音声 4 /yoN/ について調査を行った。

### 2.1 音声サンプル

実験に使用した音声サンプルは数字音声 4 /yoN/ で、成人男子 8 名が各 8 回発声した合計 64 個の音声サンプルを用いた。音声信号は 3.4 kHz の低域通過フィルタを通過後 8 kHz、12 ビットで標準化・量子化された。8 名中 2 名については発声の経時変化を調べるため 1 年後に発声された各 7 個の音声サンプルが追加使用されている。

これら音声のスペクトルを記述するパラメータとしては線形予測係数  $a_k$ 、PARCOR 係数  $K_k$ 、対数断面積比  $L_k$  が用いられ、その時系列でスペクトルの時間パターンが表わされた。また発声の経時変化の検討にはケプストラム係数  $C_k$  も使用した。分析は AD 変換された音声サンプルに対し一次の差分処理をほどこした後、10 ms ごとに 30 ms のフレームを切り出し、ハニング窓を乗じ、自己相関法による 10 次 ( $k=1\sim 10$ ) の線形予測分析を行い上記係数を求めた。これら係数よりスペクトルの極 (ホルマント周波数に対応) 位置を Z 平面上に表わし、係数の時系列より極軌跡を求めると、時間に対し急激な変化が見られる個所があり、声道の動きと対比して誤った極を抽出していると考えられるので、これを取り除く目的で各パラメータ係数の 5 フレームにわたる移動平均値の時系列を用いることにより、スペクトルの時間変化の平滑化を行っている。

### 2.2 音声の時間整合

一人の話者が同じ単語を繰り返し発声した場合には、発声時間長など、そのスペクトルの時間構造は音声サンプルにより異なっており、登録された標準音声と被験音声を照合する場合に、より良いマッチングを行う為に時間軸の非線形伸縮を行う必要がある。ここでは二つの音声サンプルの時系列間の対応を計るのに DP による方法<sup>(10)</sup>が用いられた。

本研究では標準音声の時間軸は変化させず、被験音声の時間軸のみを非線形に伸縮させ、時間整合を計る DP マッチングが用いられている。

$$A \quad ; \quad a_1 \quad , \quad a_2 \quad , \quad \dots \quad , \quad a_i \quad \dots \quad a_1$$

$$B' \quad ; \quad b'_{j(1)} \quad , \quad b'_{j(2)} \quad , \quad \dots \quad , \quad b'_{j(i)} \quad \dots \quad b'_{j(0)}$$

$$B \quad ; \quad b_1 \quad , \quad b_2 \quad , \quad \dots \quad , \quad b_j \quad \dots \quad b_j$$

図 1 B の A への時間整合

いま標準音声のスペクトルパラメータ係数ベクトルの時系列 ( $a_1, a_2, \dots, a_i, \dots, a_1$ ) を  $A$ 、被験音声のスペクトルパラメータ係数ベクトルの時系列 ( $b_1, b_2, \dots, b_j, \dots, b_j$ ) を  $B$  とし、 $B$  を  $A$  に対し時間整合させ、整合後の時系列を  $B'$  で表わし、その時系列を ( $b'_{j(1)}, b'_{j(2)}, \dots, b'_{j(i)}, \dots, b'_{j(0)}$ )

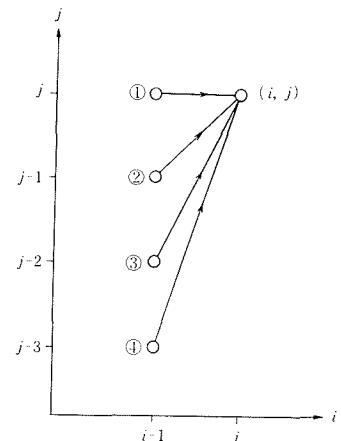


図 2 許容 DP パス

とする。これらの関係を図1に示す。DPパスは図2に示されるように4通りに制限され、パターン間類似度を表わす評価関数 $g(i, j)$ は式(1)の漸化式により決定された。

$$g(i, j) = d(i, j) + \min \begin{cases} g(i-1, i) & ; \text{①} \\ g(i-1, i-1) & ; \text{②} \\ g(i-1, i-2) & ; \text{③} \\ g(i-1, i-3) & ; \text{④} \end{cases} \quad (1)$$

$$d(i, j) = d(\mathbf{a}_i, \mathbf{b}_j) = \sum_{k=1}^P (b_{ki} - a_{ki})^2$$

ここで $d(i, j)$ は $\mathbf{a}_i$ と $\mathbf{b}_j$ の間の距離で、ユークリッド距離の二乗が用いられた。Pは分析次数である。また図2で水平方向へ続けて進める回数は2回に制限され、始端と終端の対応には次式の制限が設けられている。

$$1 \leq j(1) \leq 3, \quad J-2 \leq j(I) \leq J \quad (2)$$

$g(I, j(I))$ を最少にする経路を選択することにより $\mathbf{B}'$ が決定される。この時間整合により $\mathbf{B}$ は $\mathbf{A}$ に最も類似した時系列 $\mathbf{B}'$ に変換され、 $\mathbf{A}$ の*i*番目のフレーム(時刻*i*と記す)における係数ベクトル $\mathbf{a}_i$ に対して必ず $\mathbf{b}_{j(i)}$ が対応することになり、複数の同一単語音声に対し $\mathbf{b}_{j(i)}$ を求めその分布を調べる事が可能となった。

### 2.3 標準スペクトルパターンの作製

前節で述べたDPによる時間整合を用い、各話者が発声した複数の数字音声/yoN/より各話者の標準スペクトルパラメータ係数ベクトルの時系列(標準スペクトルパターンとよぶ)が求められた。まず各話者のS個の音声の中からその平均時間長に最も近い音声を選び、その音声のスペクトルパターンを仮の標準音声とし、他のS-1個の音声サンプルを時間整合させ、整合後のS個の長さの等しいスペクトルパラメータ係数ベクトルの時系列より、各時刻における各次数係数の平均値を成分とする平均スペクトルパラメータ係数ベクトルを求め、その時系列を標準スペクトルパターンとした。

### 2.4 スペクトルパラメータ係数ベクトルの分布

2.3で述べられた方法で、各話者の8個の音声(同一時期に発声、S=8)より標準スペクトルパターンを求め、それに本人の上記8個の音声サンプルの時間整合を行い、整合後の各時刻におけるスペクトルパラメータ係数ベクトルの分布が求められた。図3は話者Hについて求められた $\alpha_k(k=1 \sim 10)$ の各時刻*i*における分散 $\sigma_{aki}$ とその平均値 $\bar{\sigma}_{ai}$ を示す。各次数のパラメータ係数の分散は時刻により大幅に変動することが示されている。また

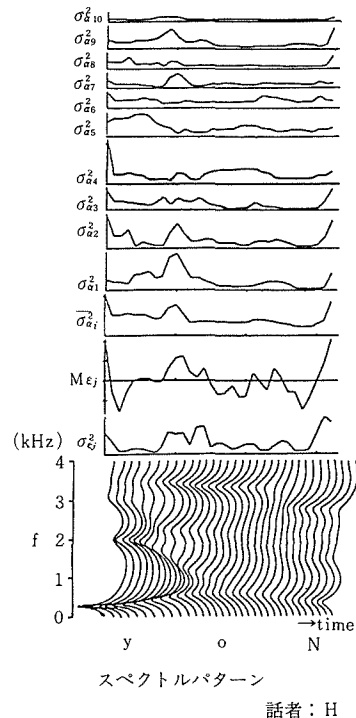


図3 線形予測係数 $\alpha_k$ の話者内変動の時間パターン(音声/yoN/)

/yo/の音韻の過渡区間および/N/の区間で分散の小さな傾向が見られる話者が多かった。スペクトルパラメータとして $K_k$ ,  $L_k$ を用いた場合についても同様な関係が認められている。

DPによる時間整合を行う場合に被験音声スペクトルパターンは非線形伸縮をうけるが、各時刻における伸縮度を調べるため式(3)に示される時間伸縮歪 $\varepsilon_{j(i)}$ を定義し、各資料について $\varepsilon_{j(i)}$ が求められた。

$$\varepsilon_{j(i)} = \frac{1}{2} [j(i) - j(i-1)] + \frac{1}{2} [j(i+1) - j(i)] \quad (3)$$

図3に示される $M_{\varepsilon_j}$ ,  $\sigma_{\varepsilon_j}^2$ は話者Hの標準スペクトルパターンにHの8個の音声サンプルを時間整合した場合に求められた $\varepsilon_{j(i)}$ の平均値と分散を示している。他の話者についても同様な値が求められたが $|M_{\varepsilon_j}|$ ,  $\sigma_{\varepsilon_j}^2$ は音韻の過渡区間において小さな値をとる傾向が認められた。

### 3 話者照合実験

8名の話者が一時期に発声した各話者それぞれ8個の単語音声/yoN/を用いて照合実験を行った。話者音声の標準スペクトルパターンと被験音声のスペクトルパターン間の距離計算を行い、距離がある閾値より小さい音声サンプルはその話者のものとし、閾値より大きいサンプルは他話者のものと判定する話者照合方式が通常想定されるが、実験では3.2で述べるように事後的に設定された閾値を用い、その閾値に対応する誤り率が求められた。

#### 3.1 重み付き距離

図3に見られるように標準スペクトルパターンに時間整合を行う場合、時刻により同一話者内の分散値 $\sigma_{ki}^2$ は異なった値をとる。したがって標準スペクトルパターンと被験音声のスペクトルパターン間の距離計算を行う場合、分散の小さな時刻におけるパターン間距離は重く荷重して算出する方が良好な照合結果が得られることが予想される。ここでは式(4)で示されるパターン間の重み付き距離を考え、重みを変化させた照合実験が行われた。

$$\left. \begin{aligned} D_{s1}(A, B) &= \frac{1}{\sum_{k=1}^p \sum_{i=1}^l w_{ki}} \sum_{k=1}^p \sum_{i=1}^l w_{ki} (b_{kj(i)} - a_{ki})^2 \\ D_{s2}(A, B) &= \frac{1}{\sum_{k=1}^p W_k} \sum_{k=1}^p W_k \frac{1}{\sum_{i=1}^l w_{ki}} \sum_{i=1}^l w_{ki} (b_{kj(i)} - a_{ki})^2 \\ w_{ki} &= \begin{cases} \textcircled{1} 1 \\ \textcircled{2} 1/\sigma_{ki}^2 \\ \textcircled{3} 1/\sigma_{ki}^4 \\ \textcircled{4} 1/\sigma_{ki}^6 \end{cases} & W_k &= \begin{cases} \textcircled{1} 1 \\ \textcircled{2} 1/\sum_{i=1}^l \sigma_{ki}^2 \end{cases} \end{aligned} \right\} \quad (4)$$

$w_{ki}$ は $k$ 次のパラメータ係数の時刻 $i$ におけるパターン間距離に対する重みである。 $W_k$ は $k$ 次のパラメータ係数に関する距離に対する重みである。また $\sigma_{ki}^2$ は標準スペクトルパターン作製に用いられた各話者の音声サンプルを本人の標準スペクトルパターンに時間整合した場合、各話者について求められた $k$ 次のパラメータ係数の時刻 $i$ における分散で、実際に使用されたスペクトルパラメータ $a_k$ ,  $K_k$ ,  $L_k$ に対し、それぞれ $\sigma_{a_{ki}}^2$ ,  $\sigma_{K_{ki}}^2$ ,  $\sigma_{L_{ki}}^2$ が用いられた。

$D_{s1}$ はパラメータ係数の次数にかかわらず同一分散を有する時刻に対して同一重みを与えるもので $D_{s2}$ において $W_k = \sum_i w_{ki}$ とした場合に相当する。 $D_{s2}$ は各次のパラメータ係数間の差に対し各次係数別、各時刻別に重み付けを行おうとするものである。 $D_{s1}$ において $w_{ki} = 1$ の場合、ならびに $D_{s2}$ の $w_{ki} = 1$ ,  $W_k = 1$ の場合は重みを全く考慮しない距離となり両者とも等しくなる。

### 3.2 照合結果

各話者の標準スペクトルパターンに対し、先に述べられた本人、他人計8人の音声サンプルについて、前述の時間整合が行われ各種の重み付き距離が求められた。

本人および他人音声サンプルのスペクトルパターンと標準スペクトルパターン間の距離の分布例を図4に示す。図中  $d_i$ ,  $d_i$  はそれぞれ本人、他人音声サンプルの距離の累積分布関数である。 $e_r = 1 - d_i$ ,  $e_a = d_i$  はそれぞれ横軸Dの値を閾値に選んだ場合の本人音声として拒否する本人拒否率、他人の音声を本人のものとして受理する詐称受理率である。 $e_r$ と $e_a$ の交点が事後的に求められ、対応する誤り率を照合誤り率Eとし、さらにE以外の照合結果を評価する測度として、式(5)であたえられる $P_1$ ,  $P_2$ が用いられた。

$$P_1 = D_2 / D_1$$

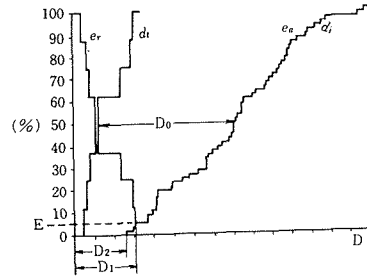
$$P_2 = D_0 / \sqrt{\sigma_i^2 + \sigma_j^2}$$

(5)

ここで $D_0$ ,  $D_1$ ,  $D_2$ は図4に示されるように分布 $d_i$ ,  $d_i$ の平均値の距離差、 $e_r$ ,  $e_a$ の横軸との交差距離にそれぞれ対応する。また $\sigma_i^2$ ,  $\sigma_j^2$ は分布 $d_i$ ,  $d_i$ の分散である。 $\sigma_i^2$ は $\sigma_j^2$ に比して図4に見られるように一般に大きいため $P_2$ は分布 $d_i$ により左右され、 $P_1$ は本人音声サンプル、他人音声サンプル中最も近接したサンプル間距離比を表わし、 $P_1$ ,  $P_2$ とも大きい程良い照合結果に対応する。

各話者8個の音声サンプルより最初に発声された4個の音声サンプルを選び、これらサンプルより求められた標準スペクトルパターンに対し、本人テスト音声として8個の音声サンプルを、詐称者音声として他の7名の話者それぞれ4個合計28個の音声サンプルを用いた照合実験が行われた。実験に使用されたスペクトルパラメータは $\alpha_k$ ,  $K_k$ ,  $L_k$ の3種類である。

表Iに照合結果を示す。表の値は話者8名についてそれぞれもとめられた結果の平均である。この表から距離尺度3 ( $D_{s2}$ で $W_k = 1 / \sum_i \sigma_{ki}^2$ の場合)を用い、 $w_{ki}$ の重みとして $w_{ki} = 1, 1/\sigma^2$ の場合に比較的良好な結果が得られることがわかる。表Iの実験では各話者の標準スペクトルパターンは8個の音声サンプル中最初に発声された4サンプルから求められたが、8個の全音声サンプルから標準スペクトルパターンを作製し、本人テスト音声として8個の音声サンプル、詐称者音声として他の7名の各8音声サンプルを使用した場合の結果を表IIに示す。表Iにくらべすべての値が良くなっており、特に $w_{ki}$ の荷重の大きい ( $1/\sigma^2 \sim 1/\sigma^6$ ) 場合の改善が著しい。しかし表IIに使用された標準スペクトルパターンの作製は、照合する音声サンプルより標準スペクトルパターンを造ることを意味し、実際には不可能であることを考慮すれば、 $w_{ki}$ は $1/\sigma^2$ が良いといえる。また標準パターンが照合結果に及ぼす影響は非常に大きく、その作製に関する検討については別途報告する予定である。また表I, 表IIの照合結果を考慮すればスペクトルパラメータは $\alpha_k$ ,  $K_k$ ,  $L_k$ のいずれを用いても同程度の照合結果が得られるといえる。



- D : パターン間距離
- E : 話者照合誤り率
- $e_r$  : 本人音声拒否誤り率
- $e_a$  : 詐称者音声受理誤り率
- $d_i$  : 本人音声分布関数
- $d_i$  : 詐称者音声分布関数

図4 パターン間距離(D)の分布と照合誤り

表Ⅰ 話者照合結果

表Ⅱ 話者照合結果

標準スペクトルパターンを4音声サンプルより作製した場合 標準スペクトルパターンを8音声サンプルより作製した場合  
 $1; D_{s1}, 2; D_{s2}(W_k=1), 3; D_{s2}(W_k=1/\sum_i \sigma_{ki}^2)$   $1; D_{s1}, 2; D_{s2}(W_k=1), 3; D_{s2}(W_k=1/\sum_i \sigma_{ki}^2)$

		$w_{ki}$		$1/\sigma^2$	$1/\sigma^4$	$1/\sigma^6$			$1$	$1/\sigma^2$	$1/\sigma^4$	$1/\sigma^6$	
E %	$\alpha$	1	3.13	1.34	3.57	8.93	E %	$\alpha$	1	0.70	0	0	
		2	3.13	1.79	4.24	4.69			2	0.70	0	0	0
		3	1.56	0.89	1.34	1.79			3	0.45	0	0	0
	K	1	2.01	4.02	4.91	9.82		K	1	0.70	0	0	0.89
		2	2.01	3.80	5.13	5.80			2	0.70	0	0	0
		3	0.89	3.35	4.69	5.36			3	0	0	0	0
	L	1	2.90	4.24	5.80	15.63		L	1	1.12	0	0	0
		2	2.90	2.90	4.69	5.58			2	1.12	0	0	0
		3	2.23	3.80	4.69	5.13			3	0	0	0	0
P <sub>1</sub>	$\alpha$	1	1.43	1.53	1.33	1.02	P <sub>1</sub>	$\alpha$	1	1.70	3.32	5.69	4.81
		2	1.43	1.48	1.51	1.53			2	1.70	2.84	4.20	4.94
		3	1.44	1.54	1.55	1.53			3	1.95	3.11	4.52	5.32
	K	1	1.39	1.57	1.24	0.66		K	1	1.75	2.94	4.25	4.32
		2	1.39	1.41	1.28	1.19			2	1.75	2.70	3.78	4.59
		3	1.43	1.48	1.33	1.22			3	2.00	2.91	3.94	4.81
	L	1	1.28	1.45	1.12	0.52		L	1	1.57	3.09	5.27	5.01
		2	1.28	1.34	1.19	1.04			2	1.57	2.69	4.02	5.00
		3	1.40	1.41	1.22	1.08			3	1.95	3.05	4.50	5.72
P <sub>2</sub>	$\alpha$	1	2.35	2.25	1.79	1.19	P <sub>2</sub>	$\alpha$	1	2.21	2.38	2.20	1.64
		2	2.35	2.48	2.21	2.08			2	2.21	2.27	2.07	1.84
		3	2.35	2.43	2.28	2.19			3	2.50	2.43	2.38	2.22
	K	1	2.47	2.30	1.57	1.29		K	1	2.36	2.34	1.91	1.13
		2	2.47	2.51	2.15	2.01			2	2.36	2.42	2.18	2.08
		3	2.33	2.44	2.05	1.90			3	2.18	2.41	2.45	2.23
	L	1	2.17	2.34	1.89	1.21		L	1	2.14	2.58	2.27	1.61
		2	2.17	2.25	2.13	1.94			2	2.14	2.20	2.21	2.15
		3	2.43	2.47	2.31	2.06			3	2.45	2.63	2.45	2.42

3.3 時間軸伸縮歪を用いた話者照合

図3に見られるように、ある話者の音声サンプルをその話者の標準スペクトルパターンに時間整合を計る場合には、その音声サンプルの時間軸は非直線伸縮歪をうける。その歪量は他話者の標準スペクトルパターンへ時間整合する場合の方が一般に大きいと考えられる。したがって時間軸伸縮歪の時間パターンには個人性

表Ⅲ 時間軸伸縮歪 $\varepsilon_i(t)$ による照合実験結果

情報が含まれており、話者照合に利用できることが予想される。一例として式(6)に示される時間歪距離 $D_T$ を用い表Ⅱに用いられた音声サンプルと標準スペクトルパターンを使用して話者照合を行った場合の結果を表Ⅲに示す。なおスペクトルパラメータとしては $\alpha_k, K_k, L_k$ を用いた場合の結果が示されている。

		$w_{ki}$	1	$1/\sigma^2$	$1/\sigma^4$	$1/\sigma^6$	$1/\sigma^8$
E %	$\alpha$	7.589	2.678	3.571	10.491	13.616	
	K	8.928	3.125	2.232	5.803	10.714	
	L	8.035	2.901	3.125	6.473	9.151	
P <sub>1</sub>	$\alpha$	0.790	0.978	0.713	0.429	0.286	
	K	0.846	1.221	1.254	0.817	0.510	
	L	0.783	1.250	1.570	1.382	1.293	
P <sub>2</sub>	$\alpha$	1.400	1.104	0.832	0.719	0.625	
	K	1.386	1.067	0.662	0.516	0.417	
	L	1.441	1.026	0.687	0.607	0.526	

$$D_T = \frac{1}{\sum_{i=1}^I w_i} \sum_{i=1}^I w_i (\epsilon_{j(i)} - M_{\epsilon_{j(i)}})^2 \tag{6}$$

$\epsilon_{j(i)}$ は式(3)で示される時間軸伸縮歪である。 $M_{\epsilon_{j(i)}}$ は $\epsilon_{j(i)}$ の話者内平均である。また式(6)中の距離荷重  $w_i$  は式(7)に示される5通りについて行われた。

$$w_i = \begin{cases} \textcircled{1} & 1 \\ \textcircled{2} & 1/\sigma_{\epsilon_{j(i)}}^2 \\ \textcircled{3} & 1/\sigma_{\epsilon_{j(i)}}^4 \\ \textcircled{4} & 1/\sigma_{\epsilon_{j(i)}}^6 \\ \textcircled{5} & 1/\sigma_{\epsilon_{j(i)}}^8 \end{cases} \tag{7}$$

$\sigma_{\epsilon_{j(i)}}^2$ は $\epsilon_{j(i)}$ の話者内分散である。

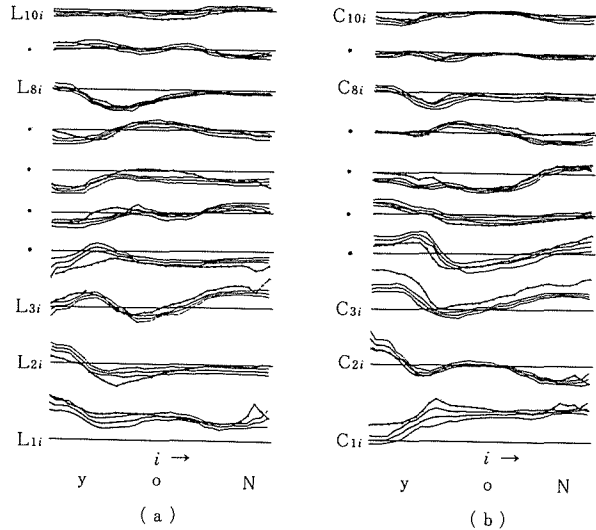
表IIIの結果は時間軸伸縮歪の話者照合への利用可能性を示している。

#### 4 話者音声の発声時期にもとづく変化

話者の識別, 照合ともに登録音声サンプルと被験音声サンプルの発声時期との差が大きくなるにたがって誤り率の増加することが報告されている。これは話者照合に用いられている音声パラメータで表わした話者音声, 発声時期により異なるため生じるものと考えられる。したがって発声時期が変わっても, 同一話者に対しなるべく変化しないような音声パラメータを利用することが望まれる。

そこでまず特定語音声を, 各種のスペクトルパラメータ係数ベクトルの時系列で表わした場合に, 発声時期の経過により, どのような変化が見られるかの調査を行った。

図5(a)は話者Hの音声/yoN/について, スペクトルパラメータとして対数断面積比 $L_k$ を用いて表わした場合の変化を示す。図は話者Hの8個の音声/yoN/を $L_k$ で表わし, 話者Hの標準パターンに整合した場合得られる各次係数値の各時刻における平均値 $M$ と標準偏差 $\sigma$ を求め,  $M$ と $M \pm \sigma$ の時間変化を実線で, 1年後に発声された話者Hの7個の音声を同様に整合して得られた平均値を附点実線で示したものである。



—— ; 最初の音声サンプルの平均値, 平均値±標準偏差  
 ..... ; 1年後の音声サンプルの平均値  
 話者H

図5 発声時期差1年の場合の音声サンプル間のパラメータの変化  
 (a)対数断面積比を使用した場合  
 (b)ケプストラム係数を使用した場合

図 (b) は図 (a) と同じ音声サンプルについて、音声スペクトルをケプストラム係数  $C_k$  で表わした場合について同様に求められたものである。実線と附点実線で表わされた平均値の間の変化幅  $\Delta M$  は、対応する時刻で求められた標準偏差に比べかなり大きな値をとる場合のあることが図からわかる。したがって、長時間にわたる発声を対象とする話者照合に対しては、この様に变化するパラメータをそのまま用いることにはかなり問題があるといえる。

表Ⅳ 各種パラメータ係数の比変化幅絶対値の平均

	$k$	1	2	3	4	5	6	7	8	9	10	mean
$\left  \frac{\Delta M}{\sigma} \right $	$\alpha$	1.72	1.71	3.70	2.83	2.00	0.97	2.35	1.35	1.01	1.98	1.96
	K	1.32	1.54	2.06	1.61	1.45	1.83	1.20	1.49	0.95	1.92	1.53
	L	1.22	1.58	2.16	1.67	1.40	1.88	1.19	1.56	0.95	1.95	1.56
	C	1.72	1.24	3.25	1.39	1.72	1.79	2.09	1.07	1.65	0.91	1.68

表Ⅳは  $\alpha_k, K_k, L_k$  ならびにケプストラム係数  $C_k$  など各種スペクトルパラメータで音声 /yoN/ を記述した場合に観測された各時刻における変化幅と標準偏差の比の絶対値  $|\Delta M/\sigma|$  の全発声区間について求められた平均値を示したものである。表は発声時期のずれが長期間にわたると、スペクトルパラメータとして  $K_k, L_k, C_k$  を用いた場合、平均で同一時期発声で見られるパラメータ係数変動の標準偏差の約 1.6 倍の変化幅が観測され、 $\alpha_k$  を用いた場合にはこれらに比べ比変化幅の平均が多少大きな値をとることを示している。また各パラメータ係数の比変化幅も次数により異なっている。しかし本表は 2 名の話者の特定語 /yoN/ の観測から得られたものであり、この結果よりこれ以上の結論を導くことは困難で、今後さらに検討を予定している。

## 5 む す び

以上に述べられた結果は数字音声 /yoN/ について少数の話者について行われた検討結果であり、普遍性のある結論を導くことはできないが、得られた結果を要約すると、

(1) 同一時期に発声された単語音声 4 /yoN/ についての話者照合では、スペクトルパラメータとして、線形予測係数  $\alpha_k$ , PARCOR 係数  $K_k$ , 対数断面積比  $L_k$  のいずれを用いても良好な照合結果が得られ、また話者間距離の算定には式(4)の  $D_{s2}$  を使い、重み  $w_{ki}, W_k$  を  $w_{ki}=1$  又は  $1/\sigma_{ki}^2$   $W_k=1/\sum_i \sigma_{ki}^2$  に選ぶのが適当である。また標準スペクトルパターンの選択により照合結果に大きな差異を生じる。

(2) 発声時期のずれが長期間にわたる場合には、スペクトルパラメータとして、 $\alpha_k, K_k, L_k$  ならびにケプストラム係数  $C_k$  のいずれを用いても、一時期に発声された音声サンプルで観測されるパラメータの話者内変動よりかなり大きな変化が見られる場合があり、 $K_k, L_k, C_k$  を使用した場合、平均で同一時期発声サンプルで観測される標準偏差の約 1.6 倍の変化幅が観測され、 $\alpha_k$  を使用した場合にはこれらより多少大きな値が観測された。

話者照合は、一般に長期間にわたって発声される音声を対象とすることが多いと考えられるため、4章に示された話者の発声変動に影響されることなく、良好な照合結果の得られるスペクトルパラメータの利用、照合方法の開発が必要であるが、これらは今後に残された問題である。

終わりに本研究に貢献された高橋成夫君、高木宏明君、実験とりまとめに協力頂いた齊川勝男技官、小田明君に深謝致します。

## 参 考 文 献

- 1) R. C. Lummis ; IEEE Trans. **AU-21**, No. 2, p. 80, 1973
- 2) 追江博昭ほか ; 日本音響学会研究発表会講演論文集 2 - 2 - 22 (昭 47-10)
- 3) 徳永靖夫ほか ; 日本音響学会研究発表会講演論文集 1 - 3 - 24 (昭 48-10)
- 4) 古井貞熙ほか ; 電子通信学会研究会資料, PRL78-30, 1978
- 5) 市川 薫ほか ; 日本音響学会誌 **35** (1979), No. 2, p. 63-69
- 6) 高木宏明ほか ; 昭和 53 年電気四学会北海道支部連合大会講演論文集 180
- 7) 広瀬秀治ほか ; 日本音響学会研究発表会講演論文集 2 - 2 - 10 (昭 54-10)
- 8) 広瀬秀治ほか ; 電子通信学会研究会資料, EA79-88, 1979
- 9) 古井貞熙 ; 電子通信学会論文誌 **57-A** (1974) No. 12, P. 880-887
- 10) 追江博昭ほか ; 日本音響学会誌 **27** (1971) No. 9, p. 483-490