

論文 / 著書情報
Article / Book Information

論題(和文)	単語の統計的パラメータによる話者認識
Title(English)	
著者(和文)	古井貞熙, 板倉文忠
Authors(English)	SADAOKI FURUI
出典(和文)	電子情報通信学会論文誌, Vol. J56-A, No. 11, pp. 717-724
Citation(English)	, Vol. J56-A, No. 11, pp. 717-724
発行日 / Pub. date	1973,
URL	http://search.ieice.org/
権利情報 / Copyright	本著作物の著作権は電子情報通信学会に帰属します。 Copyright (c) 1973 Institute of Electronics, Information and Communication Engineers.

UDC 534.78.001.36 : 612.789

単語の統計的パラメータによる話者認識

正 員 古井 貞熙†

正 員 板倉 文忠†

Talker Recognition by Statistical Features of Speech Sounds

Sadaoki FURUI† and Fumitada ITAKURA†, Regular Members

あらまし 種々の単語音声を、部分自己相関係数 (PARCOR 係数) と基本周波数の時系列で表現し、その有声音区間における平均値、そのまわりでの標準偏差、パラメータ間相関などの統計量を特徴パラメータとして、複数個の単語を組み合わせる方法を用いて話者認識実験を行った。その結果、3 か月おきの 4 時期にわたるサンプルを既知サンプルとし、その最終の時期から 3 か月後のサンプルを、4 単語を組み合わせて認識する場合、9 名から 1 名を選ぶ話者識別で平均 99.1%、詐称者が 36 名の話者照合で平均 99.2% の認識率が得られ、この方法の有効性が確認された。複数個の単語を組み合わせる効果の背景について、いくつかの点が明らかになった。認識誤りを生ずる原因となる、特徴パラメータの時期的変動特性について、種々の検討を行い、パラメータが、短期間ではきわめて安定しているが、時期がたつに従って、大きく変化する性質を、定量的に明らかにした。

1. まえがき

音声に含まれる個人性情報を抽出したり、抽出した情報を用いて、未知の音声の話者を機械的に認識しようとする“話者認識”の試みが、声紋やバンキングなどの問題に関連して、最近盛んに研究されるようになって来た⁽¹⁾。一般に話者認識の問題を取り扱う上で最も困難な点の一つは、ある程度長い期間をおくと、特徴パラメータが変化してしまうところにある^{(2),(6)}。

従来の話者認識の方法は、発声される内容に無関係な統計的パラメータを用いるものと、発声される内容をあらかじめ定めておいて、その動的パターンの時間正規化マッチングを行うものに大別することができる。前者に属するものには、長時間平均スペクトル⁽²⁾や、長時間音声の帯域フィルタ出力の相関行列⁽⁷⁾を用いる方法などがあり、パラメータを安定に抽出できる上、時期的変動を直接的に正規化できる特徴があるが、各音韻に特有な個人性情報をとり出し得ない難点がある。後者の方法による場合は、各音韻ごとの情報や動的特徴が使えるが、時間正規化に重点がおかれるので、

それ以外の時期的変動を直接的に考慮することがむずかしい。

本論文では、両者の特徴を生かす方法として、いくつかの単語を区切って発声し、それぞれの単語で時間的に平均化された統計的パラメータを、時期的変動を考慮した上で、複数個の単語について組み合わせるという方法⁽⁸⁾を試みる。複数個の単語を組み合わせる効果について、定量的検討を行うとともに、パラメータの時期的変動特性について、詳細な検討を行う。

2. 分析方法

図 1 に示すように、静かな部屋で発声された音声を 3.4 kHz 低域通過フィルタに通したのち、8 kHz でサンプリングし、11 bit で量子化を行う。10 msec 間隔で 32 msec の区間を抽出し、ハミングウィンドウを乗じてから、部分自己相関係数 (PARCOR 係数)⁽⁹⁾を用いた分析を行う。時刻 t における区間の音声を基本周波数と 1 ~ 12 次の PARCOR 係数を組み合わせた 13 次元ベクトル x_t で表現する。PARCOR 係数 k_{il} は、絶対値が 1 を越えることがないので、そのひん度分布を正規分布に近づけて、認識方法との整合をとるために、つぎのような変換⁽⁹⁾を行う。

$$x_t = (x_{1t}, x_{2t}, \dots, x_{13t}) \quad (1)$$

† 電通公社武蔵野電気通信研究所基礎研究部、武蔵野市
Musashino Electrical Communication Laboratory, N.T.T.,
Musashino-shi, 180 Japan
論文番号：昭 48-421[A-125]

$$x_{it} = \begin{cases} p_i : \text{基本周波数} & (i=1) \\ \frac{1}{2} \log \frac{1+k_{i-1,t}}{1-k_{i-1,t}} & (2 \leq i \leq 13) \end{cases} \quad (2)$$

$k_{i,t}$: 第 i 次のPARCOR係数

基本周波数は、逐次的に1~12次のPARCOR係数を導出した後の予測誤差の相関関数のピークの位置によって抽出し、そのピークの高さにしきい値を設けて、有声音区間の検出を行う⁽⁶⁾。

x_t に関して、単語における全有声音区間を用いて、平均値 μ 、標準偏差 σ 、共分散行列 Σ 、および相関行列 A を求める⁽³⁾。

$$\mu = \frac{1}{N} \sum_{t=1}^N x_t \quad (3)$$

$$\Sigma = (\sigma_{ij}) = \frac{1}{N-1} \sum_{t=1}^N (x_t - \mu)(x_t - \mu)' \quad (4)$$

$$\sigma = (\sigma_{11}, \sigma_{22}, \dots, \sigma_{13,13}) \quad (5)$$

$$A = (\lambda_{ij}) \quad \lambda_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}} \quad (6)$$

N : 有声音区間の区分数

λ_{ij} についても、式(2)と同じ変換を行う。

/ba:ngo:/, /namae:/, /bakuoN/, /ko:geN/

の4種類の単語を用いて実験を行う。この単語の選択に関しては、特別な意味はない。

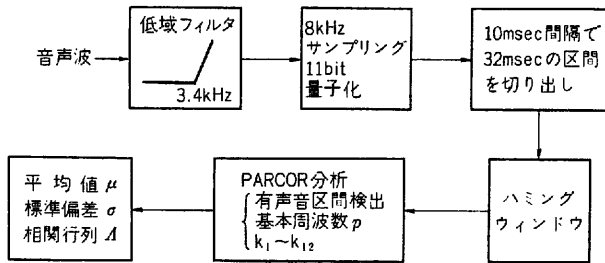


図1 分析方法

Fig.1-Block diagram of the analyzing method.

3. 認識方法

前章で述べた方法によって抽出したパラメータを用いて、話者識別および話者照合を行う。話者識別とは、未知音声があるか、あらかじめ定められた M 人の登録話者のうちの、いずれの話者によるものであるかを判定するもの（クラス数 = M ）であり、話者照合とは、未知音声がある登録話者 A のものであるか否かを判定するもの（各登録話者に対しクラス数 = 2）である。

1 単語を用いる場合の認識方法は、以下の通りであ

る。まず σ と A の各成分について、登録話者 9 名の既知サンプルを用いて、話者効果に関する分散分析を行う。話者効果の比較的大きい m 個の成分を選択し、 m 次元ベクトル y とする。

$$y = (y_1, y_2, \dots, y_m) \quad (7)$$

y を μ の成分と組み合わせて、 $13+m$ 次元の特徴ベクトル z とする。

$$z = (z_1, z_2, \dots, z_{13+m}) \quad (8)$$

$$z_i = \begin{cases} \mu : & (1 \leq i \leq 13) \\ y_{i-13} : & (14 \leq i \leq 13+m) \end{cases}$$

未知サンプルと、各話者の標準パターンに関して、時期的な変動を考慮した、つぎのような重みつき距離を設定し、これによって話者識別および話者照合を行う。

$$D(z_u, \theta_k) = |V|^{1/13+m} (z_u - \theta_k)' V^{-1} (z_u - \theta_k) \quad (9)$$

z_u : 未知サンプル

θ_k : 登録話者 k の標準パターン（既知サンプルの平均値）

V : (V_{ij}) : 時期的変動に関する共分散行列（既知サンプルから登録話者別に求めたものを平均化して用いる）

この距離は、特徴ベクトルが多次元正規分布に従うことを仮定し、その共分散行列の各成分の値が比例的

に変化できる条件のもとでの最大尤度によって、認識を行うことに相当する。 V の成分のうち、 σ や A の相関成分 (V_{ij} , $i \neq j$, i or $j > 13$) の値は、処理時間と記憶容量の削減のために 0 とおく。

複数個の単語を組み合わせる場合には、各単語について、式(9)で定義される距離を求め、これを代数平均する。話者識別の場合は、これらの距離に関して、未知サンプルから最も近い標準パターンの話者をもって、その未知サンプルの話者と判定する。話者照合の場合

は、標準パターンからの距離が、あるしきい値よりも小さいサンプルは、その登録話者によるものと判定し、それ以外のサンプルは、他の話者のものであるとして棄却する。ここではこのしきい値を、便宜上、2 種類の誤り率すなわち、受け入れるべきサンプルを棄却してしまう割合と、棄却すべきサンプルを受け入れてしまう割合が等しくなる値に事後的に設定し、このときの誤りを認識誤り率とする。

4. 認識実験

複数の単語を組み合わせる効果と、パラメータの時期的変動の効果を定量的に調べるため、種々の条件を変えて、6種類の認識実験を行う。各実験における既知サンプルと未知サンプルの時間的關係は、図2に示すとおりである。各時期に各話者は、1分間程度の間隔において、同じ単語を繰り返し3回発声している。

登録話者は男性9名である。話者識別の場合は、9名から1名を選ぶ形をとり、話者照合の場合は、9名の各登録話者に対して、棄却すべきサンプルとしてA(8名)、B(26名)およびC(36名)の3種類の話者(詐称者)によるサンプルを用いて実験を行う。Aは、登録話者9名のうちの、受け入れるべき本人以外の8名の話者、Bは、Aに他の18名を加えたもの、Cは、さらに10名を加えたものである。

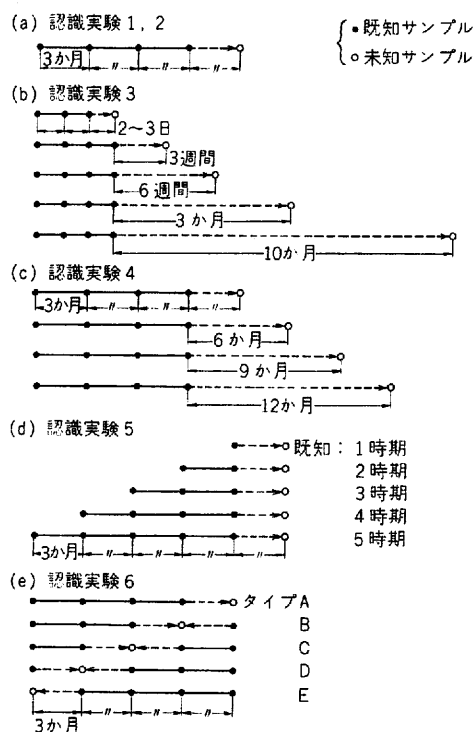


図2 各認識実験における既知サンプルと未知サンプルの関係

Fig.2-Relations between reference samples and unknown ones in each recognition experiment of six types.

4.1 認識実験1—3か月後のサンプルの認識—

図2(a)に示したように、3か月おきの4時期にわたるサンプルを既知サンプルとし、最終の既知サンプルから、3か月後のサンプルを未知サンプルとして、認識実験を行った。同様の実験を、4種類の単語に関して、それぞれ8通り行った。一回の認識に用いられる既知サンプルは、各登録話者の各単語について12サンプルである。各単語について、話者識別の未知サンプルは、合計216サンプル(9名×3回×8通り)、話者照合の場合の受け入れるべきサンプルは、同じく216サンプル、棄却すべきサンプルは、9名×3回×8とおりに(詐称者数)で、詐称者Cのとき7.776サンプルである。

σ とAのうちの認識に用いる成分の数 m を、0から20まで変えて、この特徴を用いる効果について調べた。図3は、4種類の単語と9名の登録者に関して、平均した結果を示したものである。 σ とAのうちの、話者効果の大きい20成分程度を用いることにより、 $m=0$ 、すなわち平均値 μ のみを用いる場合に比較して、誤り率がほぼ1/2に低下することがわかる。 $m=20$ としたときの平均認識率は、話者識別の場合96.7%、詐称者がCの話者照合の場合96.9%である。各単語による認識率は、話者識別、話者照合ともに95.7~97.7%の範囲におさまっており、単語による差は小さい。

話者照合の詐称者に関して、Aの場合と、BおよびCの場合とで、認識率に差がみとめられる。これは、含まれている話者の個人差とともに、つぎのような要因によるものと考えられる。すなわちAは、既知サンプルを用いて、話者の効果に関する分散分析を行う際に、用いられるサンプルの話者と共通しているが、BおよびCには、Aの他に18~28名の新しい話者が加

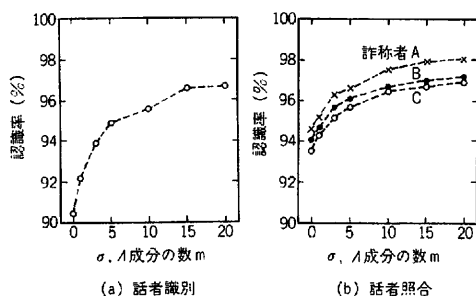


図3 認識実験1の結果
Fig.3-Identification and verification accuracy of experiment 1.

わっているためと考えられる。

話者識別の誤認識行列と、話者照合の登録話者別の誤り率を、単語ごとに調べたところ、誤りの生じやすい話者と生じにくい話者のあることと、単語によって誤りの傾向が異なることがわかった。後者の結果から、2種類以上の単語を組み合わせることにより、さらによい認識率の得られることが期待できる。

4.2 認識実験2—複数単語の組合せ—

4.2.1 結果 4単語を組み合わせたときの結果を、1単語の場合と対比して図4に示した。認識に用いる σ と A の成分の数 m の値にかかわらず、4単語を組み合わせることによって、1単語の場合の認識誤りをほぼ $1/2 \sim 1/3$ に減らすことができる⁽⁴⁾。4単語を組み合わせ、 $m=20$ としたときの、話者識別の誤認識行列、話者照合(詐称者C)の、話者別の誤り率、および詐称者による誤りの分布を図5に示す。話者識別の

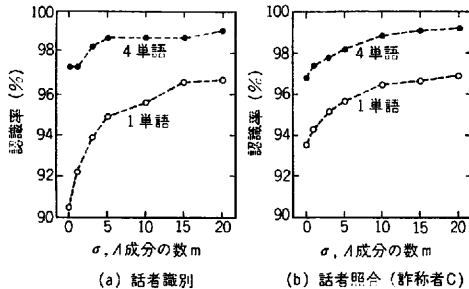
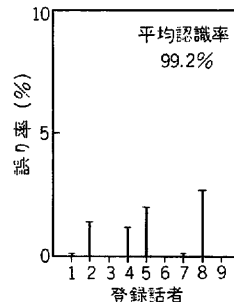


図4 認識実験2の結果
(1単語による場合との比較)

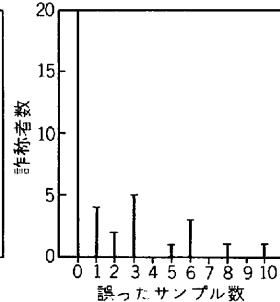
Fig.4- Identification and verification accuracy of experiment 2.

	1	2	3	4	5	6	7	8	9	認識率
1	24									100%
2		24								100
3			24							100
4				24						100
5					22	2				91.7
6						24				100
7							24			100
8								24		100
9									24	100
平均認識率										99.1%

(a) 話者識別の誤認識行列



(b) 話者照合(詐称者C)の登録話者別誤り率



(c) 話者照合(詐称者C)の詐称者による認識誤りの分布

図5 4単語を組合せ、 $m=20$ としたときの誤認識行列。

登録話者別誤り率および詐称者による誤りの分布

Fig.5-Confusion matrix, averaged error rate for individual talkers and error distribution among impostors.

平均認識率は99.1%，話者照合の平均認識率は99.2%である。

各詐称者の総サンプル数は216サンプルである。誤りのない詐称者が最も多く、最も誤りの多い詐称者は10サンプル(4.6%)が誤って登録話者のサンプルとして受け入れられてしまったことを示している。この10サンプルは、9名の登録話者のうちの誤りの多い3名に分布しており、特定な集中は見られない。

上に述べた結果は、詐称者が登録話者と同じ男性の場合のものであるが、女性8名のサンプルを詐称者として用いたところ、完全に棄却できることが確かめられた。

4.2.2 特徴ベクトルの総次元数と誤り率の関係

上の結果は、各単語で用いている特徴にかなり独立性があるために、複数個の単語を組み合わせることによって、用いる情報量がふえ、認識率がよくなることを示している。この効果に関する定量的検討として、単語を組み合わせたり、 m の値を変えたりしたときの、特徴ベクトルの総次元数と、認識誤り率の関係を調べた。

M は基本周波数と1～12次PARCOR係数の合計13成分からなるから、 σ と A から m 成分を選択して用いるとき、 n 単語の組み合わせなら総次元数は、

$$T = (13 + m)n \quad (10)$$

になる。たとえば、 $m=20$ 、 $n=4$ のとき、 $T=132$ である。

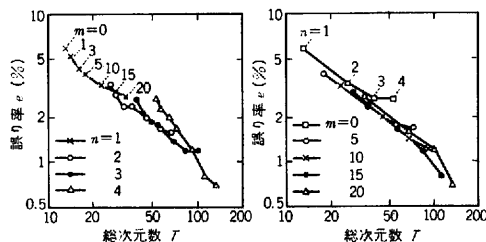
話者照合の場合の総次元数と誤り率の関係を図6に示す。 n の値でまとめたものを(a)に、 m の値でまとめたものを(b)に示した。総次元数、誤り率ともに対数スケールで示した。誤り率を e (%)で表わすと、両者

はほぼ、

$$eT = 100 \quad (11)$$

の直線で近似できる関係を有している。話者識別の場合も同様の結果が得られる。

図6(b)からわかるように、 m の値が10以下の場合には、単語数の増加に伴う誤りの減少がにぶる傾向が出ているが、 m の値が15、20の場合には、直線的に誤りが減少している。さらに単語



(a) 組合せる単語数でまとめたもの (b) σ , λ 成分の数でまとめたもの

図6 話者照合(詐称者B)の総次元数と誤り率の関係

Fig.6-Error rate vs. number of features.

数をふやしたときに、この直線性がどこまで保たれるかはわからないが、 m を15～20程度にすれば、単語数の増加によって、さらに誤り率は減少すると考えられる。たとえば、 $m=20$ とし、5単語を組み合わせるときには $T=165$ であるから、式(11)の直線性が保たれると、誤り率 $e=0.6\%$ となる。

4.2.3 組合せる効果の背景 4単語を距離上で平均するのではなく、4単語を音声波形上で時間的に接続したサンプル(2～3秒)を用いて、補足的な認識実験を行い、1単語のみによる場合、4単語を組み合わせる場合との比較を行った。その結果、1単語のみによる場合よりも、それを接続して長くした方が、認識率はやや良くなるが、単語ごとに区切って、距離上で平均する方が有効であることが確かめられた。

用いている特徴の、単語による独立性は、含まれる音韻とその長さなどが異なるための μ 成分の相違、 σ と λ の成分のうちの、分散分析によって選択される成分の相違などによるものと考えられる。

4.3 認識実験3—短期間の既知サンプルによる認識—

既知サンプルとして、ほぼ10日間という短かい期間に求めた3～4時期のサンプルを用いて、認識実験を行った。最終の既知サンプルから未知サンプルまでの時間間隔を、2～3日から10か月まで変えて、その効果を調べた。図7は、1単語による場合と、4単語の組み合わせによる場合の結果を示したものである。1単語の場合については、4種類の単語による結果を平均して示した。

1単語による場合、未知サンプルまでの時間間隔が2～3日のときには、ほぼ100%の認識が可能であるが、その時間間隔が長くなるに従って、単調に認識率が低下し、3か月程度になると、きわめて低い認識率しか得られなくなってしまう。4単語を組み合わせる

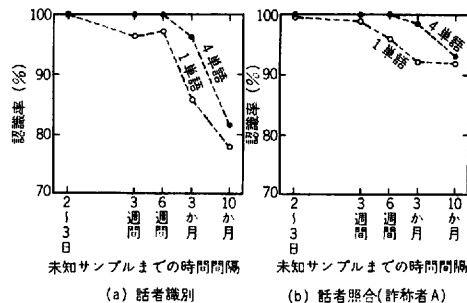


図7 認識実験3の結果($m=20$)

Fig.7-Identification and verification accuracy of experiment 3.

場合には、6週間程度の間なら100%の認識が可能であるが、この場合も、3か月程度たつと、認識率はかなり低下してしまうことがわかる。

4.4 認識実験4—長期間たった後のサンプルの認識—⁽⁵⁾

既知サンプルとしては、認識実験1および2と同様に、3か月おきの4時期にわたるものを用い、最終の既知サンプルから未知サンプルまでの時間間隔を、3か月から12か月まで4通りに変えて、認識実験を行った。1単語による場合と、4単語の組み合わせによる場合について検討した。異なる5時期の未知サンプルについて、同じ形の認識実験を行い、結果を平均して図8に示した。

いずれの場合も、未知サンプルまでの時間間隔が3か月から12か月まで長くなるに従い、認識率の低下する傾向が見られる。このことは、ここで用いている特徴が、3か月以上の長い期間内でも、ゆっくり変わってゆく性質を有することを示している。

4.5 認識実験5—既知サンプル数の効果—

最終の既知サンプルから未知サンプルまでの時間間

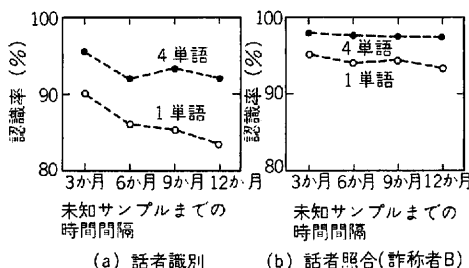


図8 認識実験4の結果($m=20$)

Fig.8-Identification and verification accuracy of experiment 4.

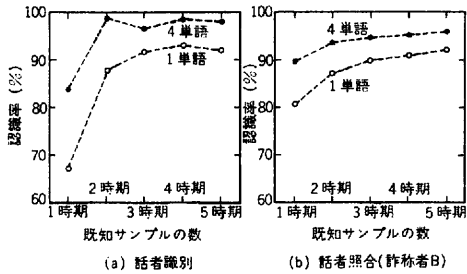


図9 認識実験5の結果 ($m=20$)
Fig.9- Identification and verification accuracy of experiment 5.

隔は3か月とし、既知サンプルを求める時期の数を、1～5時期(3か月おき)まで変えて、認識実験を試みた。結果を図9に示す。認識実験1, 2は、この“4時期”の場合に相当している。

既知サンプルとして、1時期のサンプルのみ(3サンプル)を用いた場合には、認識実験3で、未知サンプルまでの時間間隔を3か月とした場合よりも、さらに低い認識率しか得られず、2時期以上のサンプルを用いた場合との間に、大きな差が見られる。既知サンプルとしては、4時期程度のものを用いれば、ほぼ十分であることがわかる。

4.6 認識実験6—既知・未知サンプルの時間的順序関係—

サンプル間の時間間隔は3か月とし、図2(e)に示すように、既知サンプルと未知サンプルの時間的順序関係を5通りに変えて、認識実験を行った。認識実験1, 2は、このタイプAに相当する。

$m=0$ の場合の結果を図10に示す。いずれの場合も、タイプAの場合よりも、中程のタイプの方が良い認識率を示し、全体的に山形の結果を示している。このように、未知サンプルが、既知サンプルの前あるいは後にある場合よりも、中間にある場合のほうがよい認識率を示すのは、認識実験4の結果と同様に、3か月以上の長期間にわたっても、特徴がゆっくり変化する性質を有することを示しているといえよう。

認識実験3～6から、長期間たった後の未知サンプルに関して、よい話者認識を行うためには、ある程度長期間にわたった既知サンプルを必要とし、また、その蓄積を逐次更新してゆく必要があることがわかる。具体的には、3か月後のサンプルを認識するには、その3か月前からさかのぼって、3か月おきに4時期程度にわたったサンプルが必要となる。これらの結果は、

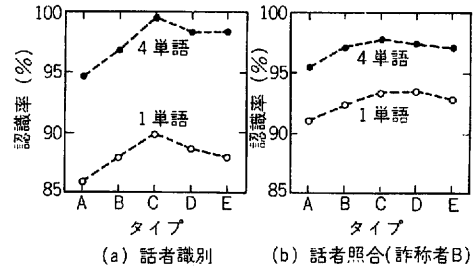


図10 認識実験6の結果 ($m=20$)
Fig.10- Identification and verification accuracy of experiment 6.

筆者らが、長時間平均スペクトルを用いた話者認識の検討を行った際に得られた結果⁽²⁾と類似している。

5. 考察と検討

5.1 パラメータの時間的変動

認識実験3～6の結果は、この方法で用いている特徴パラメータが、10日間程度の短い期間内ではきわめて安定しているが、3か月あるいはそれ以上たつと、ゆっくりではあるが大きく変化する性質を有することを示している。その変動の様子を μ 成分について調べた。

5.1.1 主成分分析 μ の長期間と短期間の変動を直接的に調べるために、各話者の各単語ごとに、長期間と短期間のそれぞれの場合について、主成分分析を行い、第1, 第2主成分からなる平面上に、各サンプル点を射影した。長期間のサンプルとしては、3か月おきに約3年間にわたって求めたサンプル(A, B, C, ..., L)を用い、短期間のサンプルとしては、図11に示すように、2～3日から3か月までの種々の間隔において約5か月間にわたって求めたサンプル(a, b, c, ..., g)を用いた。分析は、パターンの変動に関する共分散行列を、相関行列に変換してから行った。

結果の一部を図12に示す。図の下に、第1, 第2主成分の累積寄与率(総分散に対する2主成分の分散の割合)を示した。相関行列にもとづく分析なので、長期間と短期間の変動の大きさの比較はできないが、変動の特徴をつかむことができる。短期間の結果からわかるように、ほぼ10日間のb, c, d, eの点は、他の点と比べてかなり接近しており、パターンが、短い期間内では安定しているが、長い期間たつと大きく変わってくる性質を示している。長期間の場合の各

サンプル点は、3か月の間隔をおいて、かなりランダムに変動している。短期間の場合でも長期間の場合でも、その変動には、はっきりした一方向性のようなもの

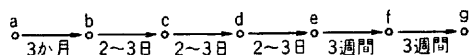


図11 短期間のサンプルの時間間隔
Fig.11-Time alignment of the short time span data.

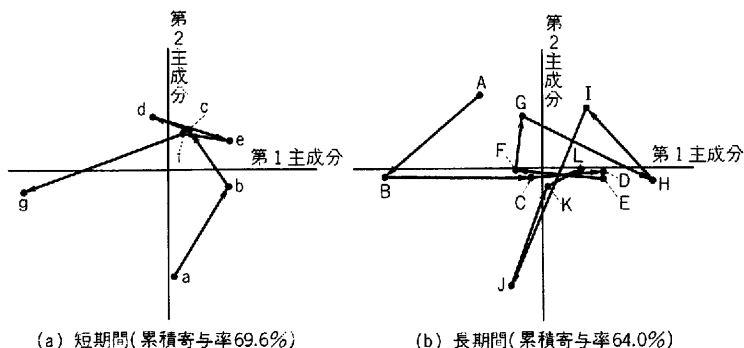


図12 μ ベクトルの短期間と長期間の変化
Fig.12-Long-term and short-term variations of the vector μ .

のは見出されない。

5.1.2 時間間隔と変化量の関係 種々の時間間隔をおいたときの、 μ の値の変化量を、ユークリッド距離(の二乗)を用いて求め、図13の実線で示した。4単語と9名の話者について平均した結果を示した。この図を見ると、時間間隔が9か月程度になるまでは、時間間隔が長くなるに従って、単調に変化量が大きくなる傾向があり、それ以上になると、時間間隔の長さに無関係に、ほぼ一定になっている。

ユークリッド距離(の二乗)は、近似的にパラメータの自己相関係数と対応づけることができる。図13の曲

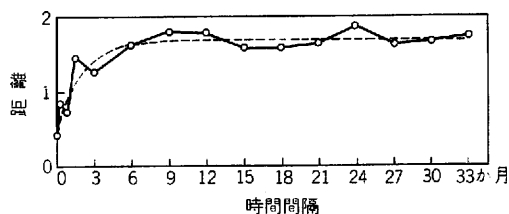


図13 時間間隔と μ ベクトルの変化量の関係
Fig.13-Amount of variation of vector μ vs. time passage.

線を指数曲線と見なして、あてはめを行ったところ、図の破線の結果が得られた。これは、不規則信号を、ほぼ2か月の時定数をもつRCフィルタに通した成分と、通さない成分の和からなる信号の、自己相関係数から求まる距離変化曲線である。前者の成分は、パラメータの時期的変動に対応し、後者の成分は、時期に無関係な変動に対応すると考えられる。 μ を構成する13成分のそれぞれについて、同様の検討を行ったところ、どの成分も同程度の時定数をもつ、類似の性質

を示した。

なお、 σ と λ の成分について、別に調べたところ、 μ 成分ほど顕著ではないが、同様の性質を有することが明らかになった。

5.2 認識に用いる σ と λ の成文を各単語で共通にした場合

これまででは、 λ と σ のうちの認識に用いる m 成分を認識を行うごとに、既知サンプルについて分散分析を行って選択していた。前述したように、このことが、単語を組み合わせる効果の原因の一つになっていると考えられるが、用い

る成分を各単語で共通に固定することができれば、特徴選択が不要になるとともに分析が簡単になる。

そこで、実際に4単語について用いる成分を共通に定めて、図2(a)の形で補足的な認識実験を行った。4単語に共通して比較的話者効果の大きい成分として、 σ からは基本周波数と1～10次のPARCOR係数に関する成分、 λ からは、基本周波数と1～10次のPARCOR係数との相関を示す成分を選んだ(合計21成分)。その結果、4単語を組み合わせるとき、話者識別で平均98.6%、詐称者Bの話者照合で平均98.3%の認識率が得られた。分散分析によって、単語別を選択する場合に比較して、0.5～1.0%の低下になっている。

6. むすび

種々の単語音声部分を部分自己相関係数(PARCOR係数)と基本周波数の時系列によって表現し、有声音区間におけるそれらの平均値、そのまわりでの標準偏差、パラメータ間相関などの統計量の特徴パラメータとして、複数の単語を組み合わせる方法により、話者の機械認識を試みた。

その結果、3か月おきの4時期にわたるサンプルを既知サンプルとし、その最終の時期から3か月後のサンプルを認識する場合に、4単語を組み合わせる方法により、9名から1名を選ぶ話者識別で平均99.1%、詐称者が36名の話者照合で平均99.2%の認識率が得られ、この方法の有効性が確認された。複数個の単語を組み合わせる効果の背景や、パラメータの時期的変動特性について、いくつかの点が明らかになった。その変動特性は、2か月程度の時定数をもつRCフィルタに通した不規則信号に類する性質を有している。

今後、時期的変動特性の解明、話者照合において、詐称者がさらに増加したときの効果、作り声の影響をはじめ、その他の種々の問題点について、検討を進めて行きたい。

謝辞 日ごろご指導いただく、当研究所野田基礎研究部長、斎藤第4研究室長、ならびに研究室の諸氏に感謝の意を表する。

文 献

- (1) 斎藤：“話者認識”，数理科学，116，p.49（昭48-02）。

- (2) 古井，板倉，斎藤：“長時間平均スペクトルによる話者認識”，信学誌，55-A，10，p.549（昭47-10）。
- (3) 古井，板倉：“スペクトルパターンの時間変動特性を考慮した話者認識”，音響学会講演論文集，2-2-20（昭47-10）。
- (4) 古井，板倉：“スペクトルパラメータによる話者認識におけるキーワード数と時期的変動の効果”，信学全大講演論文集，1258（昭48-03）。
- (5) 古井，板倉：“話者認識におけるスペクトルパラメータの時期的変動の効果”，音響学会講演論文集，2-1-18（昭48-05）。
- (6) 板倉：“統計的手法による音声の特徴抽出”，東北大学通研シンポジウム：音声情報処理，Ⅱ-5-1（昭46-02）。
- (7) K. P. Li and G. W. Hughes：“Talker differences as they appear in correlation matrices derived from spectra of continuous speech”，84th Acoust. Soc. Amer.，HH4（Nov.1972）。
- (8) P. D. Bricker, et al：“Statistical techniques for talker identification”，Bell Syst. tech. J.，50.4，p.1427（April 1971）。
- (9) R. A. Fisher：“Statistical methods for research workers”，Oliver & Boyd Ltd.（1958）。
- （昭和48年5月11日受付）