

論文 / 著書情報
Article / Book Information

論題	音声スペクトルの統計的パラメータに含まれる個人性情報
著者	古井 貞熙, 板倉 文忠
掲載誌/書名	研究実用化報告, Vol. 24, No. 5, pp. 959-975
発行日	1975,

音声スペクトルの統計的パラメータに 含まれる個人性情報*

古 井 貞 照**
板 倉 文 忠***

あらまし 種々の単語音声部分を自己相関係数(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、そのまわりの標準偏差、およびパラメータ間相関係数などの統計量に含まれる個人性情報の分析と、それらの統計量を用いた話者認識実験を行った。その結果、上記のパラメータには長期間たった後にも保たれる個人差があり、話者認識のためのパラメータとして用いることができること、これらのパラメータは、2カ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、ゆっくりした時期的変動特性を有すること、これらのパラメータから得られる個人性情報には、単語による独立性があるため、複数の単語に含まれる情報を組合せるのが有効であること、などが明らかになった。ほぼ10カ月にわたる学習サンプルを用い、4単語を組合せれば、3カ月後の未知サンプルの話者を99%以上の認識率で認識することができる。

1 ま え が き

音声の中には、その意味内容に関する情報のほかに、様々な情報が含まれている。その中の個人性情報、すなわち誰の声であるかという情報を抽出し、それを客観的なパラメータで表現することは、人間—機械系としての音声研究の一環として、音声認識の高度化や合成音の高品質化とも密接な関連を有し、将来は話者の機械認識という応用も考えることができる。

一般に音声波から個人性情報を客観的なパラメータで抽出するうえでもっとも困難な点の1つは、ある程度長い期間をおくと、同じ話者が同じ言葉を発声しても、それらのパラメータが変化してしまうところにある。われわれは、音声の個人性に関する研究の出発点として、以前に短文章の長時間平均(パワー)スペクトルをとり上げ、そのスペ

クトルパタンの長期間にわたる変動に着目して、実験的考察を行った。その結果、長時間平均スペクトルが、発声内容に無関係な話者の特徴を示すパラメータとして有効であることを確認するとともに、長時間平均スペクトルが種々の興味深い性質をもつことを明らかにした⁽²⁾⁽³⁾。

音声波からの個人性情報の抽出、あるいは話者認識に関する従来の方法は、発声される内容(言葉)に無関係な統計的パラメータを用いるものと、発声内容をあらかじめ定めておいて、その動的パタンの時間正規化マッチングを行おうとするものに大別することができる⁽⁴⁾。前者に属するものには、上述の長時間平均スペクトルや、長時間音声の帯域フィルタ出力の相関行列に着目した方法⁽⁵⁾などがある。これらの方法は、パラメータを安定に抽出できるうえ、パラメータの長期間にわたる変動特性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。後者に属するものには、ホルマント、基本周波数、エネルギーなどの時間変化パタンのマッチングを行うものがある。これらの方法では各音韻ごとの情報や動的な特徴が使えるが、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難しい。

* Individual Characteristics of the Statistical Features of Speech Sounds. By Sadaoki FURUI and Fumitada ITAKURA.

この研究は第4研究室で行われたものである。

** 基礎研究部第4研究室研究主任

*** 武蔵野電気通信研究所研究専門調査員

本論文では、両者の特徴を生かす方法として、いくつかの定められた単語を区切って発声し、それぞれの単語長の範囲内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うえで複数個の単語について組合せるという方法⁽¹⁰⁾を試みる。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とともに、複数個の単語を組合せる効果などについて定量的な検討を行う。

2 音波の分析

音声のスペクトルバタンを時系列としてみると、種々の時間的変動を伴っていることがわかる。その中には、いわゆる“ピッチのゆらぎ”をはじめとする細かい変動から、音韻の変化やイントネーションに対応する比較的ゆるやかな変化まで、多種多様なものが含まれている。このような変動特性の中に含まれる個人性情報を統計的パラメータとして抽出することを試みた。

2.1 PARCOR 分析

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、11 bit でサンプリングおよび量子化を行い、10 ms 間隔で 32 ms の区間を抽出する。この区間についてハミングウィンドウを乗じてから、1～12 次の PARCOR 係数（部分自己相関係数）⁽⁶⁾ $k_1 \sim k_{12}$ と基本周期 T を抽出する。基本周期は、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（変形自己相関関数）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⁽⁸⁾。

図2は、単語 /name/ の有声音区間を、上で述べた方法によって分析した結果の例である。上から順に、基本周期、全エネルギー、1～12 次の各 PARCOR 係数の時間変化パターンを示す。いずれも、平均値と変動幅を正規化してプロットしてある。右の (b) は、(a) と同じ話者が (a) から約 1 年後に発声したサンプルに関する結果である。

多数の話者が長期間にわたって反覆して発声したサンプルを、このようにして調べたところ、同じ単語を発声しても、各パラメータの変動幅の比較的大きい話者と小さい話者のあること、各パラメータが互いに話者によって特徴のある相関関係を有していることなどがわかった。PARCOR 係数は、近似的に声道断面積関数に対応づけることができる⁽⁸⁾から、基本周波数と PARCOR 係数間、あるいは PARCOR 係数相互間の相関は、発声器官の運動上の一般的な制約とともに、各話者固有の発声器官の運動上の特徴を反映していると考えられる。

2.2 統計的パラメータ

PARCOR 係数は相関係数の一種であり、その頻度分布は、平均値が ± 1 に近い場合には、極端に ± 1 近くにかたまって、非対称な偏った分布形となることがわかっている。また ± 1 に近い PARCOR 係数は、その値のわずかなずれが、声質に大きな影響を及ぼすことが経験的に知られており、このことは PARCOR 係数の定義からも容易に推察されることである。一般に相関係数 R は、つぎのような Z 変換を行えば、 Z はほぼ正規分布

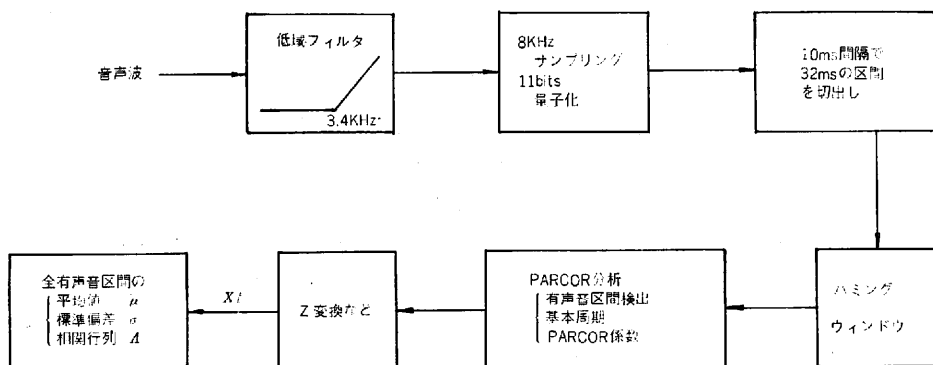


図1 分析方法

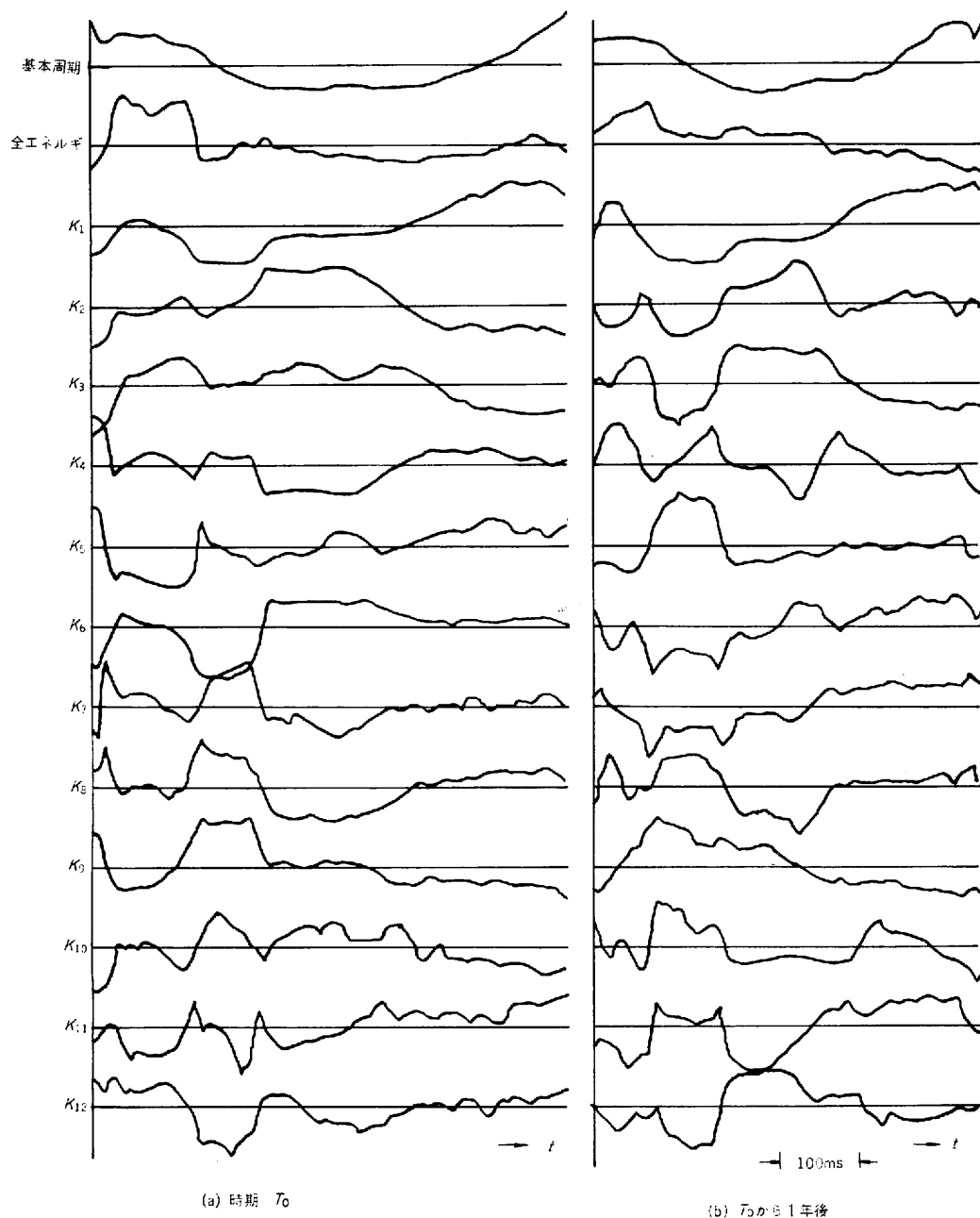


図 2 単語 /namae/ の有声音区間の分析結果 (話者 F)

に従うことが知られている⁽¹¹⁾ので、以下に述べる処理上の便宜のために、各 PARCOR 係数にこの変換を適用する。

$$Z = \frac{1}{2} \log \frac{1+R}{1-R} \quad (1)$$

一方基本周期 T については、多くの場合、周波数を

確率変数とする場合にはほぼ正規分布で近似できることが知られている⁽¹²⁾ので、

$$f = F_s / T \quad (2)$$

F_s : 標本化周波数 (8 kHz)

を求める。

時刻 t における Z 変換を施した PARCOR 係数

$k_{1t} \sim k_{12t}$ と基本周波数 f_t を組合せて、13次元ベクトル x_t とする。

$$\begin{aligned} x_t &= (x_{1t}, x_{2t}, \dots, x_{13t}) \\ &= (k_{1t}, k_{2t}, \dots, k_{12t}, f_t) \end{aligned} \quad (3)$$

単語の全有声音区間におけるベクトル x_t に関する統計的パラメータとして、平均値ベクトル μ 、標準偏差ベクトル σ 、共分散行列 Σ 、および相関行列 A を求める⁽⁴⁾⁽⁵⁾。

$$\mu = \frac{1}{N} \sum_{t=1}^N x_t \quad (4)$$

$$\Sigma = (\sigma_{ij}) = \frac{1}{N-1} \sum_{t=1}^N (x_t - \mu)'(x_t - \mu) \quad (5)$$

(' は転置を意味する)

$$\sigma = (\sigma_{11}, \sigma_{22}, \dots, \sigma_{13,13}) \quad (6)$$

$$A = (\lambda_{ij}), \lambda_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \quad (7)$$

N : 有声音区間の区分数

λ_{ij} ($i \neq j$) についても、さらに Z 変換を行う。

/bango:/, /namae/, /bakuon/, /ko:gen/ の4種類の単語を多数の話者が長期間にわたって発声したサンプルについて、これらの統計的パラメータを抽出し、これらのパラメータに含まれる個人性情報を分析する。上記の単語の選定に関しては、特別な意味はない。

3 分散分析

上述の方法によって抽出した統計的パラメータのおおの、有意な個人差があるか否かの検定を行うとともに、それらに含まれる個人性情報の大きさを調べるため、各パラメータについて話者要因に関する分散分析を行った。分析には、男性9名の話者がほぼ3カ月おきに3年間にわたって発声した4単語のサンプルを用いた。

図3は、各パラメータが有する話者効果の大きさ F を、危険率1%のときの有意水準値 $F_{0.1\%}$ で正規化して(危険率1%のときの有意水準値を1として)示したものである。 μ , σ , A のベクトルおよび行列のそれぞれについて、全成分の平均値を示した。この結果から、 μ ベクトルの成分が有する話者効果をもっとも大きいが、

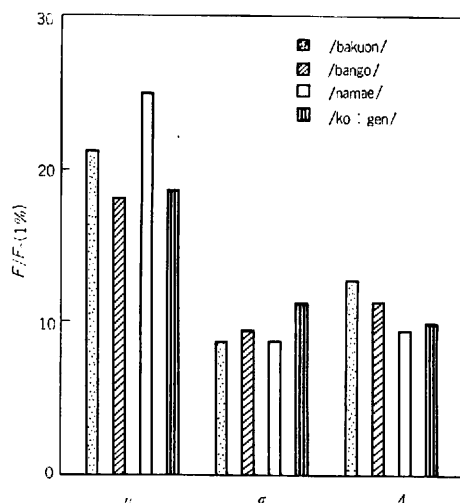
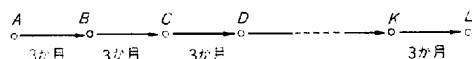
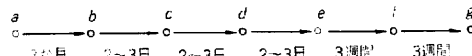


図3 各パラメータの話者効果



(a) サンプル α



(b) サンプル β

図4 2種類のサンプルの時間間隔

σ ベクトルおよび A 行列の成分も、きわめて大きな個人差を有することがわかる。

4 時期的変動

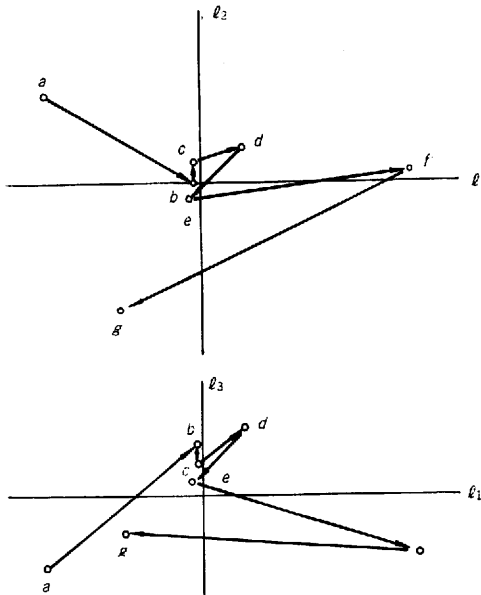
4.1 主成分空間への射影

μ ベクトル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、2組のサンプルを用いて主成分分析を行い、第1—第2主成分からなる平面、および第1—第3主成分からなる平面に各サンプル点を射影した。サンプル α は、図4(a)に示すように3カ月おきに約3年間にわたって求めたサンプル(A, B, ..., L)の組であり、サンプル β は、同図(b)に示すように、2~3日から3カ月までの種々の間隔を置いて約5カ月間にわたって求めたサンプル(a, b, ..., g)の組である。分析はバタンの変動に関する共分散行列を、相関行列

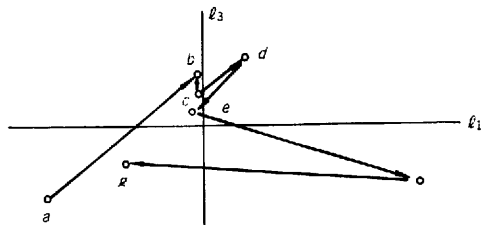
に変換してから行った。

結果の一部を図5に示す。(a)は話者Sの単語/ko:geN/に関する結果、(b)は話者Mの単語/baŋgo:/に関する結果、

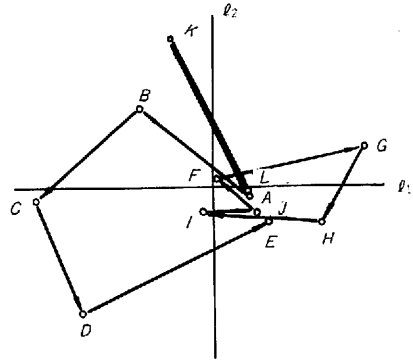
/:/に関する結果で、いずれも第1—第2主成分からなる平面への射影を上図に、第1—第3主成分からなる平面への射影を下図に示した。図の下に第1—第3



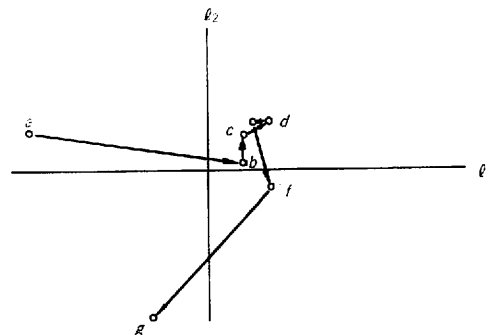
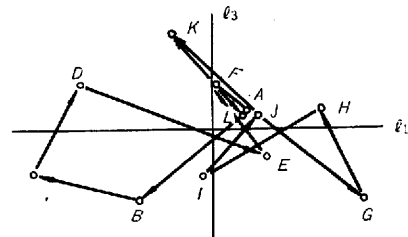
サンプルβ ($\gamma=77.0\%$)



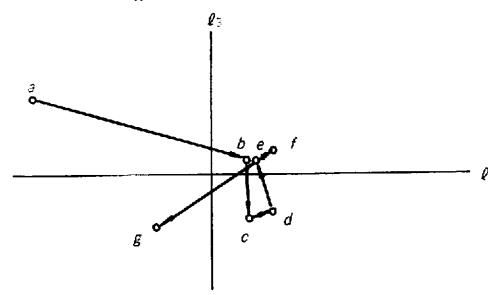
(a) 話者 S /ko:geN/



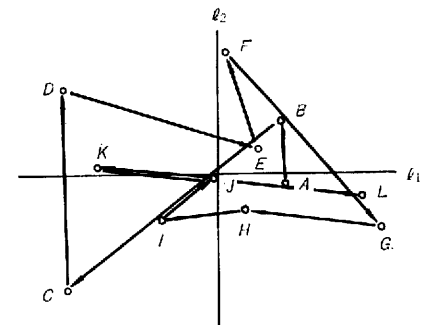
サンプルα ($\gamma=65.6\%$)



サンプルβ ($\gamma=71.2\%$)



(b) 話者 M /baŋgo:/



サンプルβ ($\gamma=71.2\%$)

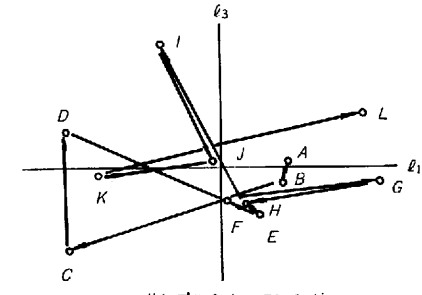


図5 第1～第3主成分からなる空間へのサンプル点の射影
 l_1 : 第1主成分, l_2 : 第2主成分, l_3 : 第3主成分
 γ : 第1～第3主成分の累積寄与率

主成分の累積寄与率 γ (総分散に対する 3 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、各国のスケールは必ずしも一致していないが、これらの図から、時期的変動の特徴をつかむことができる。

サンプル α の結果からわかるように、ほぼ 10 日間の b , c , d , e の 4 点は、他の点に比べて互いにか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。サンプル β の各サンプル点は、3 カ月の間隔をおいてかなりランダムに変動している。10 日間程度の短期間の場合でも、3 年程度の長期間の場合でも、その変動には、はっきりした 1 方向性のようなものは見出されない。

4.2 短期間と長期間のサンプルのばらつき

前節の結果から、10 日間程度の短い期間のサンプルのばらつきの大きさと、3 カ月以上の長期間のサンプルのばらつきの大きさに差のあることがわかった。このことを定量的に調べるため、約 2 年間にわたる長期間のサンプルの標準偏差と、10 日間程度の短期間のサンプルの標準偏差の比を求めた。ベクトル μ , σ , および行列 A の各成分ごとに [長期間のサンプルの標準偏差/短期間のサンプルの標準偏差] の値を求め、ベクトル μ および σ については、基本周波数と PARCOR 係数に関する成分を分け、PARCOR 係数に関する成分については全成分の平均値を求めた。行列 A については、全成分の平均値を求めた。

図 6 にその結果を示す。これから、 μ , σ , A のいずれも、ほぼ 10 日間の短期間の場合に比べて、ほぼ 2 年間の長期間の場合は、標準偏差が 1~1.5 倍に増大することがわかる。

4.3 時間間隔と変化量の関係

サンプルの時期的変動の性質をさらに詳しく調べるため、時間間隔の大きさと μ ベクトルの値の変化量との関係を調べた。変化量 ($d(\tau)$) は、下記のようにユークリッドノルムの 2 乗を用いて求めた。

$$d(\tau) = E[\|\mu_t - \mu_{t+\tau}\|^2] \quad (8)$$

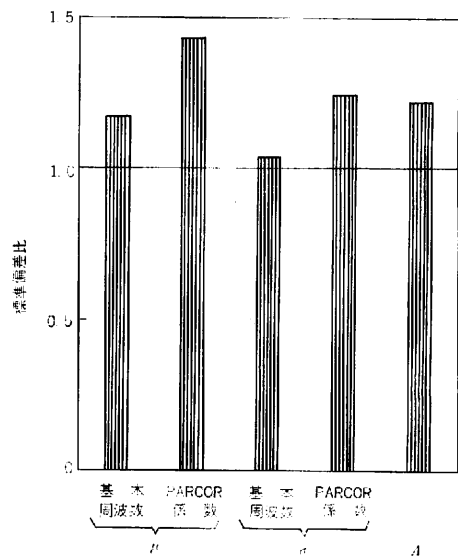


図 6 長期間と短期間のサンプルの標準偏差比

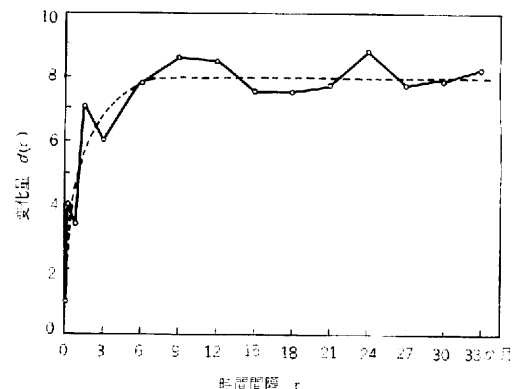


図 7 μ ベクトルの変化量と時間間隔の関係

種々の時間間隔に対して、多数のサンプルの組の変化量を求めて平均し、これをさらに 4 単語と 9 名の話者について平均した。結果を図 7 に示す。同じ時期に繰返し発声したサンプル間の変化量 (ばらつき) を 1 とした。

この図を見ると、時間間隔が 6 カ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、変化量は時間間隔の長さに無関係にほぼ一定になっている。

ユークリッドノルムの 2 乗の期待値は、かりに 1 次元空間で考えれば、パラメータの自己相関関数 ($\phi(\tau)$) と対応づけることができる。

$$d(\tau) = 2(\phi(0) - \phi(\tau)) \quad (9)$$

図7の曲線を指数曲線とみ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カ月の時定数をもつRCフィルタに通した成分（相関関数は指数関数となる）と、通さない成分（相関関数は $\tau=0$ で有限の値をもつインパルスとなる）の和からなる信号の、自己相関関数から求まる距離変化曲線に相当する。前者の成分はパラメータのゆっくりした時期的変動に対応し、後者の成分は時期に無関係な不規則変動に対応すると考えられる。ベクトル 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、どの成分も上の結果と類似の性質を示し、同程度の時定数をもつことが明らかになった。

なお、ベクトル σ と行列 A の成分について、別に調べたところ、ベクトル μ の成分ほど顕著ではないが、同様の性質を有することがわかった。

5 話者認識

節2.2で述べた方法によって抽出した統計的パラメータ μ 、 σ 、および A に含まれる個人性情報を評価し、時期的変動をはじめとする種々の性質をさらに明らかにするため、これらのパラメータを用いて、話者識別および話者照合実験を行った。話者識別とは、未知音声が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。

認識の判定に1単語のみを用いる場合の認識方法は以下のとおりである。まずベクトル σ と行列 A の各成分について、登録話者9名の学習サンプルを用いて、話者効果に関する分散分析を行う。 σ と A の全成分のうち、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。

$$y = (y_1, y_2, \dots, y_m) \quad (10)$$

y と μ の成分を組合せて $13+m$ 次元の特徴ベクトル z とする。

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

$$z_i = \begin{cases} \mu_i & (1 \leq i \leq 13) \\ y_{i-13} & (14 \leq i \leq 13+m) \end{cases} \quad (11)$$

各登録話者の標準パターンとして、各話者の学習サンプルの平均値を蓄えておき、未知サンプルと各登録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求める。

$$D(z_u, \theta_k) = |V|^{1/13+m} (z_u - \theta_k)' V^{-1} (z_u - \theta_k) \quad (12)$$

z_u : 未知サンプル

θ_k : 登録話者 k の標準パターン

$V = (V_{ij})$: 時期的変動に関する共分散行列

(学習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)

この値によって、話者識別および話者照合の判定を行う。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としたときの尤度に相当する。 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としておく。

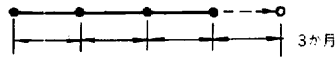
認識の判定に複数個の単語を組み合わせる場合には、各単語について式(12)で定義される距離を求め、これを全単語について代数平均する。

話者識別の場合は、上記の距離に関して、未知サンプルからもっと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小さい未知サンプルは、その登録話者によるものと判定し、それ以外の未知サンプルは、他の話者による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、すなわち本人のものとして受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、このときの誤り率を認識誤り率とする。

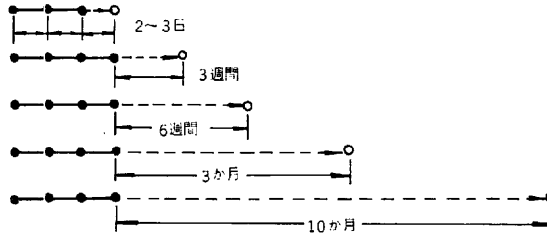
6 認識実験

パラメータの時期的変動の効果と、複数個の単語を

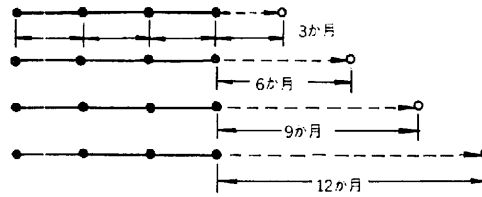
- (a) 統計的パラメータの有効性に関する実験 (節6.1)
 複数単語を組み合わせる効果に関する実験 (節6.2)



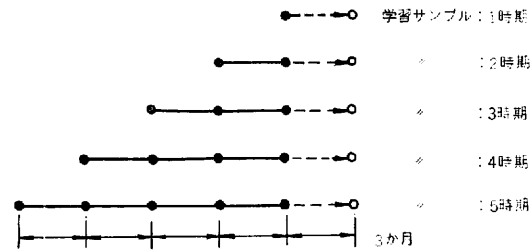
- (b) 短期間の学習サンプルによる認識 (項6.3.1)



- (c) 3か月以上たった後のサンプルの認識 (項6.3.2)



- (d) 学習サンプル数の効果に関する実験 (項6.3.3)



- (e) 学習・未知サンプルの時間的順序関係の効果に関する実験 (項6.3.4)

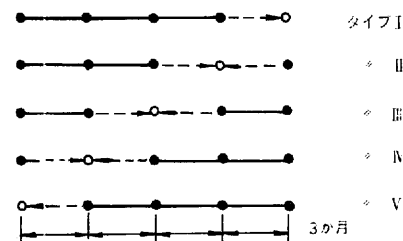


図8 各認識実験における学習サンプルと未知サンプルの時間関係

● 学習サンプル, ○ 未知サンプル

組合せる効果を定量的に調べるとともに、パラメータに含まれる個人性情報の大きさを評価し、個人差を表現するパラメータとしての有効性を確かめるため、節2.3で述べた4種類の単語を多数の話者が発声した音声サンプルを用い、種々の条件を変えて、6種類の認識実験を行った。各実験における学習サンプルと未知サンプルの時間的關係は、図8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、同じ単語を繰返し3回発声している。

登録話者は成人男性9名である。話者識別の場合には、9名から1名を選ぶ形をとり、話者照合の場合には、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)、C(36名)およびD(111名)の4組の話者(詐称者)によるサンプルを用いて実験を行った。Aは登録話者9名のうちの受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、Dはさらに75名を加えたものである。詐称者はいずれも成人男性である。

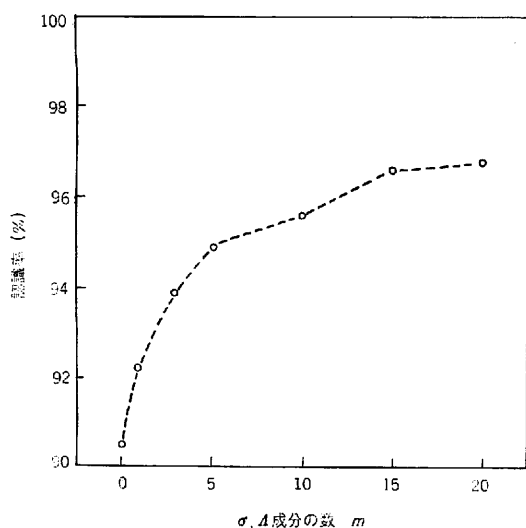
6.1 統計的パラメータの有効性

— 3か月後のサンプルの認識 —

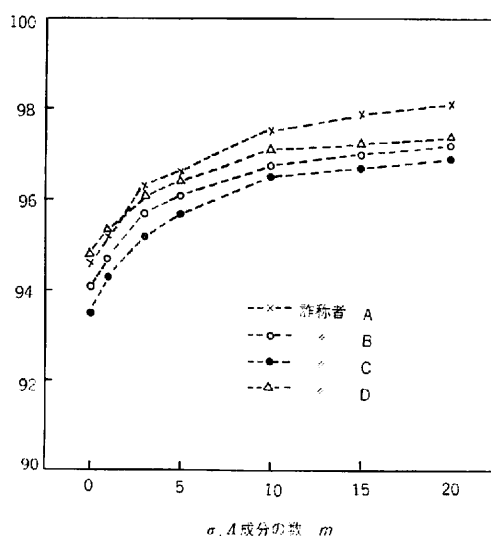
図8(a)に示したように、3か月おきの4時期にわたるサンプルを各登録話者の学習サンプルとし、最終の学習

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。異なる8時期(3か月間隔)の未知サンプルに対して、同じ形式の実験を行った。1回の認識に用いられる学習サンプルは、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)、話者照合の場合の登録話者のものとして受入れるべきサンプルは同じく216サンプル、他の話者のものとして棄却すべきサンプルは9名×3回×8通り×(詐称者数)で、詐称者Dのとき23976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9は4種類の単語と9名の登録話者に関して平均した結果を示したものである。 m の値が増加するに従って、認識誤りが指数関数的に減少し、 $m=20$ 、すなわち 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、 $m=0$ 、すなわち平均値ベクトル 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96.8%、詐称者がDの話者照合の場合97.3%である。各単語による認識率は、話者識別、話者照合ともに96.3~98.3%の範囲におさまっており、単語による差は小さい。以上の結果は、本論文で用いている統計



(a) 話者識別



(b) 話者照合

図9 単語の統計的パラメータの有効性

的パラメータ μ , σ および A が, 個人性情報を表現し, 話者の機械認識を行うためのパラメータとして有効であることを示している。

$m \geq 10$ の範囲で, 話者照合の詐称者に関して, A の場合とその他の B , C , D の場合とで認識率にやや差がみとめられる。これは, 詐称者の中に含まれている話者の個人差の偏りとともに, 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, σ と A の成分につ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学習サンプルの話者と共通しているが, B , C , および D には, A の他に 18~103 名の新しい話者が加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, 話者識別の誤りの組合せや, 話者照合の誤りを生じやすい話者が, 単語によって必ずしも共通していないことがわかった。この結果から, 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される。次節でこの検討を行う。

6.2 複数単語を組合せる効果

学習サンプルと未知サンプルの時間的關係は前節

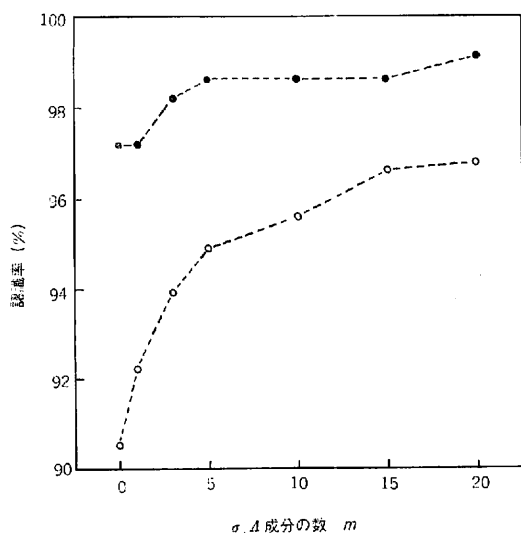
(節6.1)と同じとし, 4 単語を組合せて認識の判定を行った。認識実験の結果を, 1 単語を用いる場合(節6.1の結果)と対比して, 図10に示した。認識に用いる σ と A の成分の数 m の値にかかわらず, 4 単語を組合せることによって, 1 単語のときに比較して, 認識誤りをほぼ $1/2 \sim 1/4$ に減らすことができる⁽⁹⁾。4 単語を組合せ, $m=20$ としたときの平均認識率は, 話者識別で 99.1%, 話者照合(詐称者 D)で 99.3% である。

6.3 パラメータの時期的変動の効果

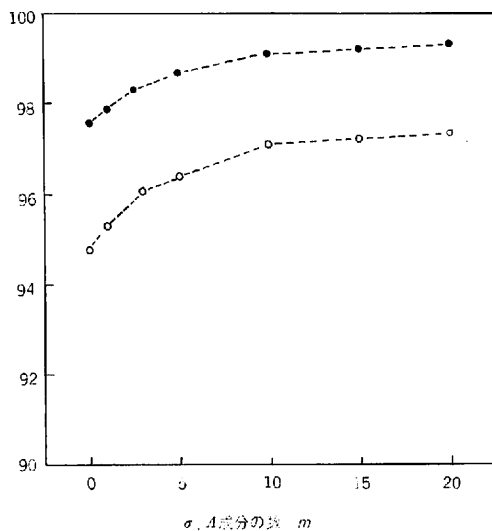
6.3.1 短期間の学習サンプルによる認識

学習サンプルとして, ほぼ 10 日間という短い期間内に求めた 3~4 時期のサンプルを用いて, 認識実験を行った。最終の学習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 カ月まで変えて, その効果を調べた。図 11 は, 1 単語による場合と, 4 単語の組合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



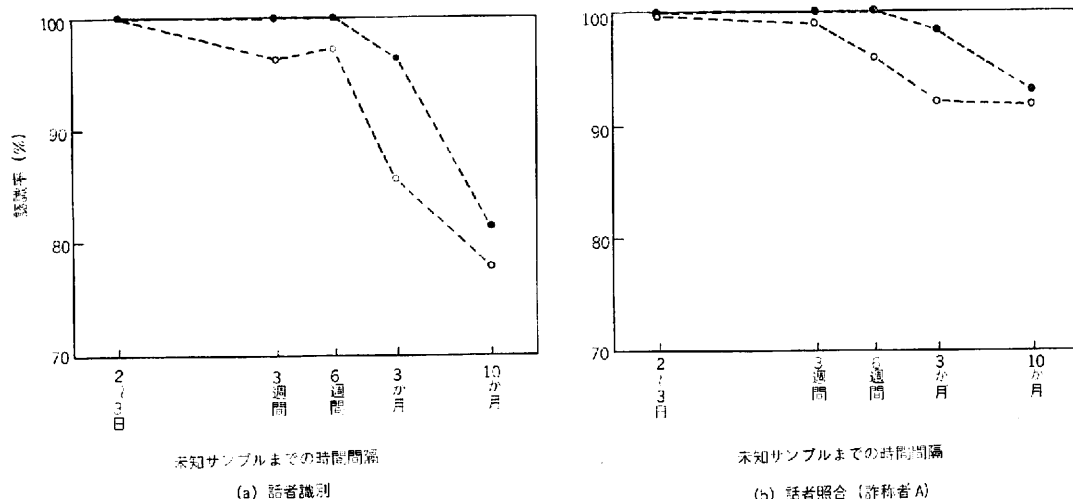
(a) 話者識別



(b) 話者照合 (詐称者 D)

図 10 複数単語を組合せる効果

----- 4 単語組合せ
- - - - - 1 単語

図 11 短期間の学習サンプルによる認識結果 ($m=20$)

----- 4単語組合せ
 ----- 1単語

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単語を組合せる場合には、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。

このように、学習サンプルが10日間程度の短期間のものである場合に、未知サンプルまでの時間間隔が3か月程度になると急激に認識誤りが増加するのは、章4で明らかになったパラメータの比較的ゆっくりした時期的変動に起因していると考えられる。この結果と節6.1および節6.2の認識結果の比較から、3か月程度以上の長期間たったのちの未知サンプルを高い認識率で認識するためには、ある程度長期間にわたった学習サンプルを用いる必要があることがわかる。

6.3.2 3か月以上たった後のサンプルの認識

学習サンプルとしては、節6.1および節6.2の認識実験と同時に、3か月おきの4時期にわたるものを用い、最終の学習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単語の組合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12に示した⁽⁷⁾。

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6か月まで長くなると、誤り率が2倍程度に増加するが、さらにそれ以上長くなっても誤り率はほとんど増加せず、ほぼ一定になっている。この結果は、節4.3で得られた時間間隔とパラメータの変化量との関係によく対応している。

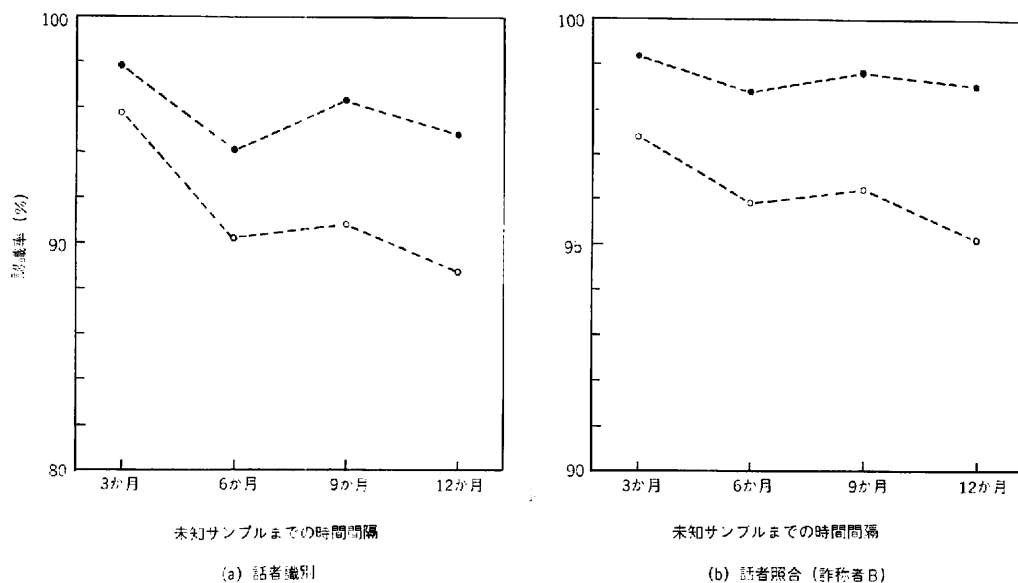
6.3.3 学習サンプル数の効果

最終の学習サンプルから未知サンプルまでの時間間隔は3か月とし、学習サンプルを求める時期の数を、1~5時期(3か月おき)の範囲で変えて、認識実験を試みた。結果を図13に示す。節6.1および節6.2の認識実験は、この“4時期”の場合に相当する(実験に用いている未知サンプルは若干異なっている)。

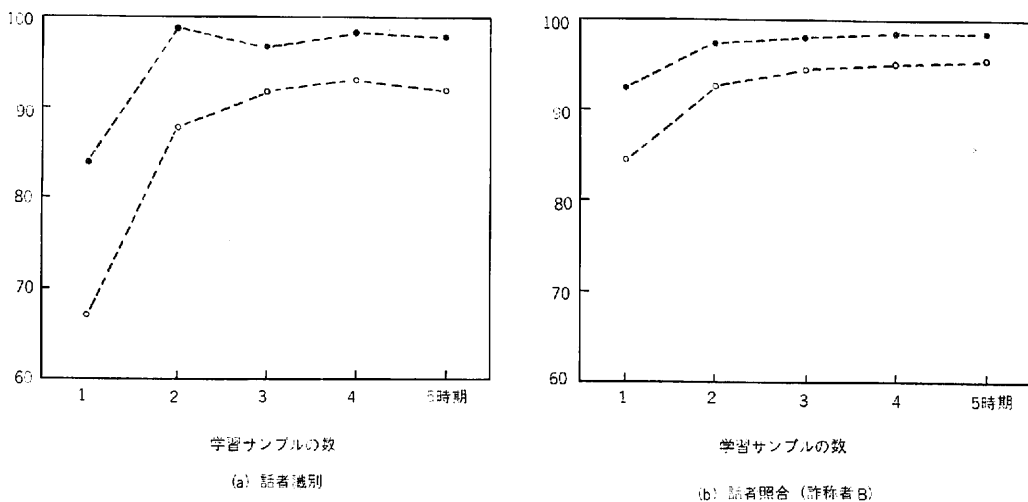
学習サンプルとして1時期のサンプルのみ(3サンプル)を用いた場合には、項6.3.1の認識実験(学習サンプルとしてほぼ10日間にわたる4時期のサンプルを使用)で未知サンプルまでの時間間隔を3か月とした場合よりも、さらに低い認識率しか得られず、2時期以上のサンプルを用いた場合との間に大きな差がみられる。学習サンプルとしては、4時期(ほぼ10か月)程度のものを用いれば、十分であることがわかる。

6.3.4 学習・未知サンプルの時間的順序関係の効果

サンプル間の時間間隔は3か月とし、図8(e)に示すように、学習サンプルと未知サンプルの時間的順序関係

図 12 3 カ月以上たった後のサンプルの認識結果 ($m=20$)

----- 4 単語組合せ
 -o-o- 1 単語

図 13 学習サンプル数の効果 ($m=20$)

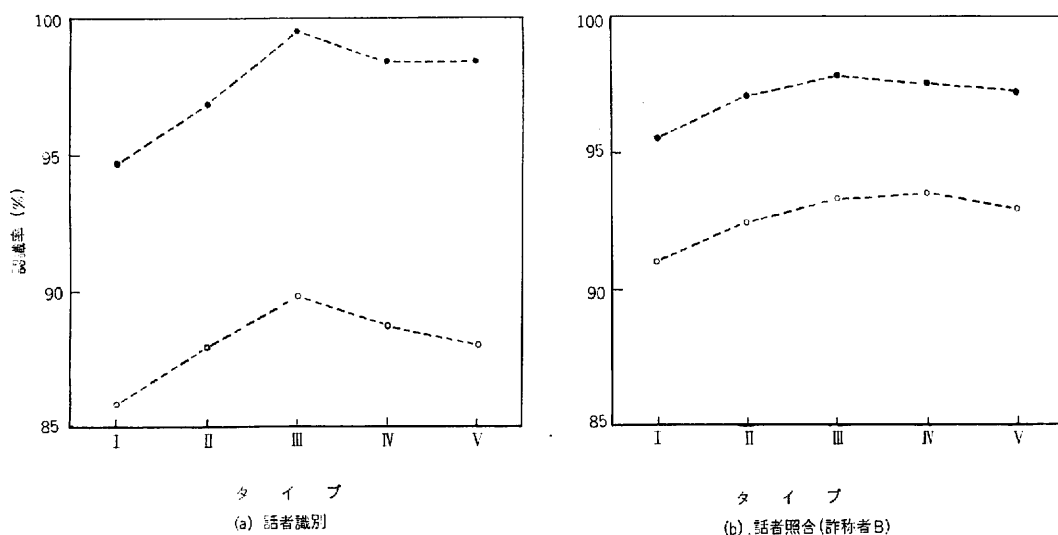
----- 4 単語組合せ
 -o-o- 1 単語

を 5 通りに変えて、認識実験を行った。節 6.1 および節 6.2 の認識実験は、このタイプ I に相当する（認識実験に用いている未知サンプルは一部異なる）。

$m=0$ の場合の結果を図 14 に示す。いずれの場合も、タイプ I あるいは V の場合よりも、中ほどのタイプの方がよい認識率を示し、全体的に山形の結果を示している。このように、未知サンプルが学習サンプルの前あるいは後にあ

る場合よりも、中間にある場合の方がよい認識率を示すのは、項 6.3.2 の認識実験や節 4.3 の結果と同様に、ここで用いている特徴が 3 カ月以上の期間でも（6 カ月程度までは）、時間とともに平均的にもとの値から遠ざかるような変化を示しているといえよう。

以上の項 6.3.1～項 6.3.4 の認識実験の結果から、長期間たった後の未知サンプルに関してよい話者認識率

図 14 学習・未知サンプルの時間的順序関係の効果 ($m=0$)

----- 4単語組合せ
 ---o--- 1単語

を得るためには、ある程度長期間にわたった学習サンプルを必要とし、また多くの場合に、その蓄積を逐次更新していく必要のあることがわかる。具体的には、3カ月後の未知サンプルを認識するには、その3カ月前からさかのぼって、3カ月おきに4時期程度にわたったサンプルが必要となる。

6.4 考察と検討

6.4.1 話者による認識結果の相違

節 6.1 および節 6.2 の認識実験の条件で $m=20$ とした場合について、話者識別の誤認識行列、および話者照合の登録話者別の誤り率と詐称者別の誤り率を調べた。表 1 は、認識の判定に 1 単語を用いる場合と、4 単語を組合せて用いる場合の、話者識別の誤認識行列を示したものである。図 15 は同様に 1 単語を用いる場合と 4 単語を用いる場合の、話者照合(詐称者 D)の登録話者別の誤り率を示したものである。これらの結果からわかるように、認識誤りを生じやすい話者と生じにくい話者は、話者識別と話者照合で共通しており、話者 2, 4, 5, 8 は前者に、話者 1, 3, 6, 7, 9 は後者に属する。

図 16 は詐称者 D (111 名) を用い、4 単語を組合せる方法による話者照合の、詐称者による誤りの分布を

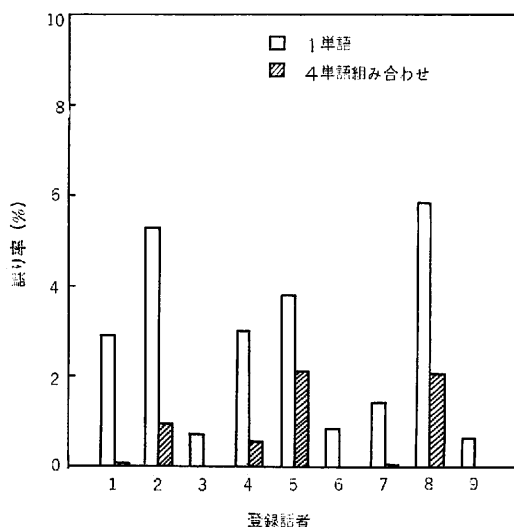


図 15 話者照合の登録話者別誤り率(詐称者 D)

示したものである。各詐称者の総サンプル数は 216 サンプルである。誤りの全くない詐称者がもっとも多く、111 名中 72 名 (64%) をしめる。誤りのもっとも多い詐称者は 12 サンプル (5.6%) が誤って登録話者のサンプルとして受入れられてしまっている。誤りが 3 サンプル (1.4%) 以下の詐称者は 97 名で、全体の 87% である。

ここで述べた結果は、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

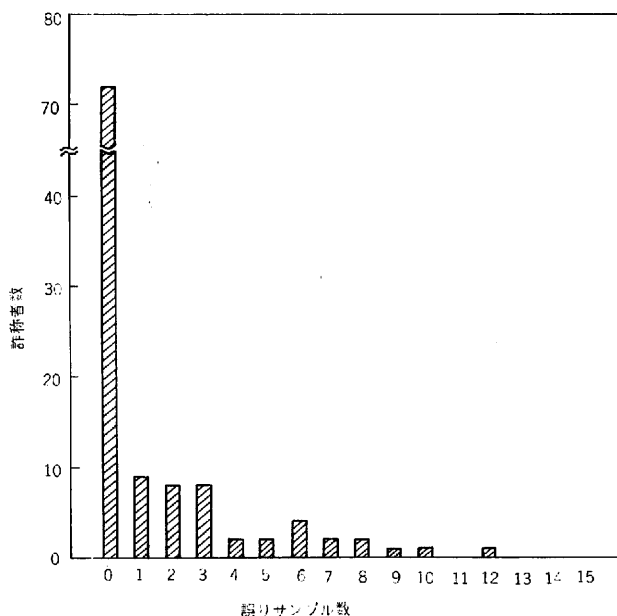


図 16 話者照合の詐称者による誤りの分布
(詐称者 D, 4 単語組合せ)

表 1 話者識別の誤認識行列

(a) 1単語											
入	出	1	2	3	4	5	6	7	8	9	認識率(%)
1		96									100
2			92			2			2		95.8
3			1	95							99.0
4			2		91		3				94.8
5		1	1			88	3		3		91.7
6						1	95				99.0
7								95		1	99.0
8			1			3	2	1	88	1	91.7
9										96	100
平均 認 識 率 (%)											96.8

(b) 4単語組み合わせ											
入	出	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

て用いたところ、完全に棄却できることが確かめられた。

6.4.2 複数単語の組合せの効果について

4 単語を標準ボタンと未知サンプルとの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った。その結果を、1 単語のみによる場合、および 4 単語を組合せる場合の結果と比較して図 17 に示す。この図から、1 単語のみによる場合よりも、それを接続した時間的に長いサンプルを用いた方が、認識率は

やよくなるが、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わかる。

この結果は、各単語で用いている特徴にかなり独立性があるために、複数の単語を組合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。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 Δ 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。この結果は未知の音声波か

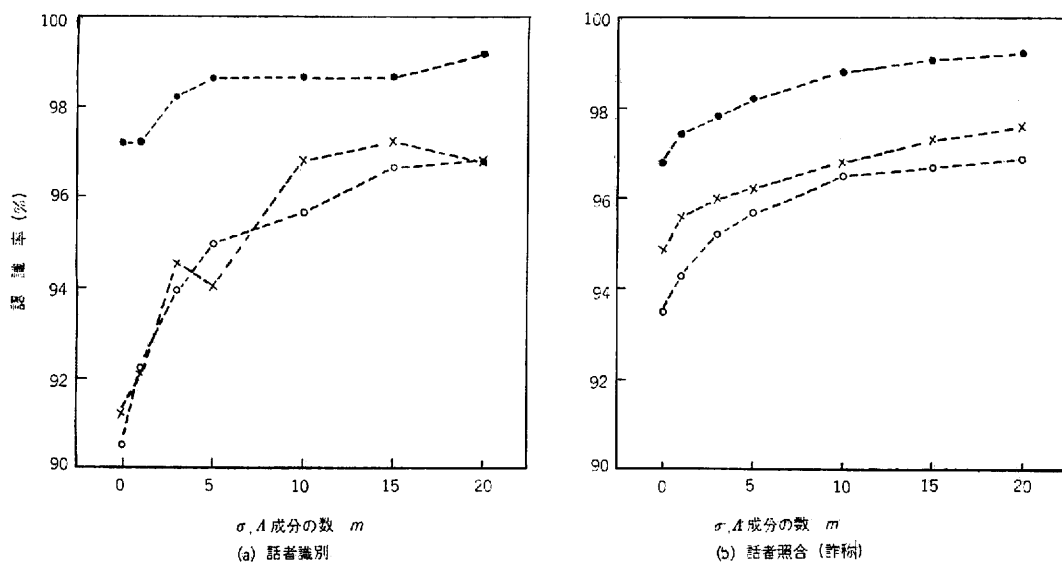


図 17 4 単語を接続する効果 (1 単語, 4 単語組合せとの比較)

--X-- 4 単語接続
 ---o--- 4 単語組合せ
 ——— 1 単語

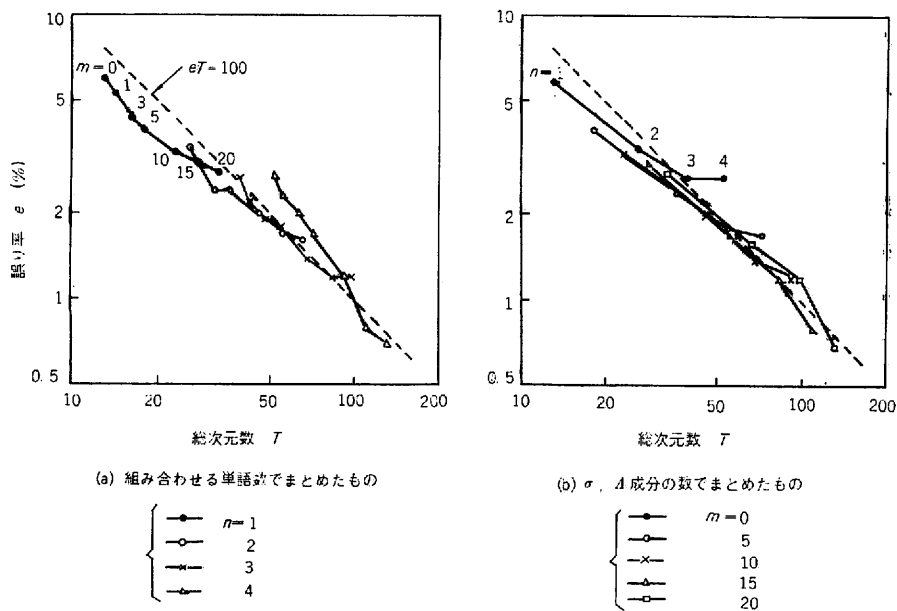


図 18 話者照合の総次元数と誤り率の関係 (詐称者 B)

ら個人性情報を能率よく抽出するためには、音韻情報をあらかじめ抽出しておくことが有効であることを示している。

6.4.3 特徴ベクトルの総次元数と誤り率の関係

複数個の単語を組合せる効果に関する定量的検討

として、単語数と m の値を変えたときの、特徴ベクトルの総次元数と認識誤り率の関係を調べた。 μ ベクトルは基本周波数と 1~12 次の 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 単語の組合せなら総次元数は

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

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

話者照合（詐称者 B ）の場合の総次元数と誤り率の関係を図 18 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、

$$eT = 100 \quad (14)$$

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 18(b) からわかるように、 m の値が 10 以下の場合には、3～4 単語で単語数の増加に伴う誤り率の低下がにぶる傾向が出ているが、 m の値が 15 あるいは 20 の場合には、直線的に誤り率が低下している。さらに単語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、さらに誤り率は低下すると考えられる。たとえば、 $m=20$ とし、5 単語を組合せるときには、 $T=165$ であるから、式 (14) の直線性が保たれるとすると、誤り率 $e=0.6\%$ となるものと推定できる。

6.4.4 認識に用いる σ と A の成分を各単語で共通にした場合

これまで、 A と σ のうちの認識に用いる m 個の成分を、認識を行うごとに学習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せる効果の原因の 1 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 8(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（個人差）の大きい成分として、 σ ベクトルからは基本周波数と 1～10 次の PARCOR 係数に関する成分、 A 行列からは、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。このときの m の値は 21 になる。認識実験の結果、4 単語を組合せるとき、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。この値は、分散分析によって単語別を選択する場合に比べて、話者識別で 0.5%、話者照合で 1.0% の低下になっている。

7 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、および PARCOR 係数と基本周波数相互間の相関係数などの統計量に含まれる個人性情報の分析を行った。さらに、複数の単語に関する上記のパラメータを組合せる方法を用いて、話者認識実験を行った。

その結果、つぎのようなことが明らかになった。

(1) 上記の統計的パラメータには、長期間たった後にも保たれる個人差があり、話者識別および話者照合のためのパラメータとして用いることができる。

(2) これらのパラメータは、2 カ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、ゆっくりした時期的変動特性を有する。

(3) このため、3 カ月程度たった後の未知サンプルの話者を高い認識率で認識するためには、各登録話者について、過去 10 カ月程度にわたる学習サンプルが必要になる。

(4) これらのパラメータから得られる個人性情報には、単語による独立性がある。このため、複数の単語に含まれる情報を組合せることにより、単語数に反比例して、認識誤り率が低下する。

(5) 3 カ月おきの 4 時期にわたるサンプルを各登録話者の学習サンプルとし、その最終の時期から 3 カ月後のサンプルを認識する場合、4 単語を組み合わせる方法を用いれば、9 名から 1 名を選ぶ話者識別で平均 99.1%、詐称者が 111 名の話者照合で平均 99.3% の認識率を得ることができる。

(4) に示した複数の単語を組合せる効果は、個人性情報を能率よく抽出するためには、あらかじめ未知音声の音韻情報を抽出しておくことが有効であることを示している。発声内容が未知の音声から、音韻と話者の情報をいかにして分離して抽出するかという問題は、今後の興

味あるテーマになると思われる。

また、パラメータのゆっくりした時期的変動特性を解明して、その原因をつきとめ、変動しにくい個人性情報を分離抽出することは、話者認識における学習サンプルの収集期間の短縮のためだけでなく、個人性情報の本質に近づくためにも、検討しなければならない問題であろう。

今後、これらの問題のほか、話者による認識率の相違、作り声の影響などについても検討を進めていく予定である。

謝 辞

日ごろ、ご指導ごべんたついただく武蔵野電気通信研究所基礎研究部野田部長、第4研究室斎藤室長、ならびに熱心に討論していただく第4研究室的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，No. 116，p. 49，1973.
- (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，No. 10，p. 549，1972.
- (3) 古井・板倉・斎藤：音声の長時間平均スペクトルに含まれる個人性情報，通研実報，23，No. 6，p. 1199，1974.
- (4) 古井・板倉：単語の統計的パラメータによる話者認識，信学誌，56-A，No. 11，p. 717，1973.
- (5) 古井・板倉：スペクトルパタンの時間変動特性を考慮した話者認識，音響学会秋季講演論文集，2-2-20，1972.
- (6) 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，No. 1258，1973.
- (7) 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，音響学会春季講演論文集，2-1-18，1973.
- (8) 板倉：統計的手法による音声の特徴抽出，東北大通研シンポジウム：音声情報処理，II-5-1，1971.
- (9) 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.，HH 4，1972.
- (10) P. D. Bricker et al：Statistical Techniques for Talker Identification，B. S. T. J.，50，No. 4，p. 1427，1971.
- (11) R. A. Fisher：Statistical Methods for Research Workers，1958，Oliver and Boyd Ltd.
- (12) 杉本・藤井・関口：音声の基本周波数の変動特性および選択信号素子の音声信号による誤動作，通研実報，11，No. 12，p. 2205，1962.

(1974, 12, 6 受付)