

論文 / 著書情報
Article / Book Information

論題	音声の長時間平均スペクトルに含まれる個人性情報
著者	古井 貞熙, 板倉 文忠, 斉藤 収三
掲載誌/書名	研究実用化報告, Vol. 23, No. 6, pp. 1199-1210
発行日	1974,

音声の長時間平均スペクトルに含まれる 個人性情報*

古 井 貞 熙**
板 倉 文 忠***
斎 藤 収 三****

あらまし 音声の個人性パラメータとして、短文章の長時間平均（パワー）スペクトルをとり上げ、1年半以上の長期間にわたって多くの話者について、そのスペクトルパターンの変動を調べるとともに、その変動特性を考慮した話者認識を試みた。その結果、そのパターンは2〜3日から3週間程度の間ではきわめて安定しているが、その後時期がたつにつれて、同じ発声者によるものでも、ゆっくり大きく変化することが明らかになった。その変動特性は、1カ月半程度の時定数を有する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。この長期間にわたる変動を考慮した適切な距離尺度を設定すれば、3カ月以上の長期間をおいた後でも、話者の認識が十分に可能であることが確認され、長時間平均スペクトルパターンに、安定した個人性情報が含まれていることが確かめられた。

1 ま え が き

音声の中に含まれる個人性情報を抽出したり、その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ているようになってきた⁽¹⁾。一般に個人性情報を定量的かつ客観的なパラメータで記述しようとするとき、もっとも困難な点の1つは、ある程度長い期間をおくと、それらのパラメータが変化してしまうところにある⁽²⁾⁽³⁾。しかし、個人性パラメータの長期間にわたる変動特性を、積極的にとり上げて検討した例はほとんどなく、話者認識実験においても、既知サンプルから未知サンプルまでの時間間隔の効果を定量的に調べたり、数カ月程度の長期の時間間隔において、

認識実験を行なった例はきわめて少ない⁽⁴⁾⁽⁵⁾。

本論文では、発声内容にほとんど依存せずに個人性を表現するパラメータとして、短文章の長時間平均（パワー）スペクトル^{(6)~(8)}をとり上げ、そのスペクトルパターンの長期間にわたる変動に着目して行なった実験的考察の結果について述べる。具体的な方法としては、1年半以上の期間にわたり、多くの発声者について、長時間平均スペクトルパターンの変動を調べるとともに、話者認識実験における、既知サンプルと未知サンプルの時間間隔と認識率との関係などを通じて、長時間平均スペクトルの時期的変動特性を定量的に明らかにする。同時に、その変動特性を考慮した方法により、長時間平均スペクトルによる話者認識の可能性を検討する。

2 長時間平均スペクトルの測定

2.1 測定方法

図1に示すように、しゃ音した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 で量子化を行なう。順々に32

* Individual Characteristics of the Longtime Averaged Speech Spectrum. By Sadaoki FURUI, Fumitada ITAKURA and Shuzo SAITO.

この研究は第4研究室で行なわれたものである。

** 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

**** 基礎研究部第4研究室長

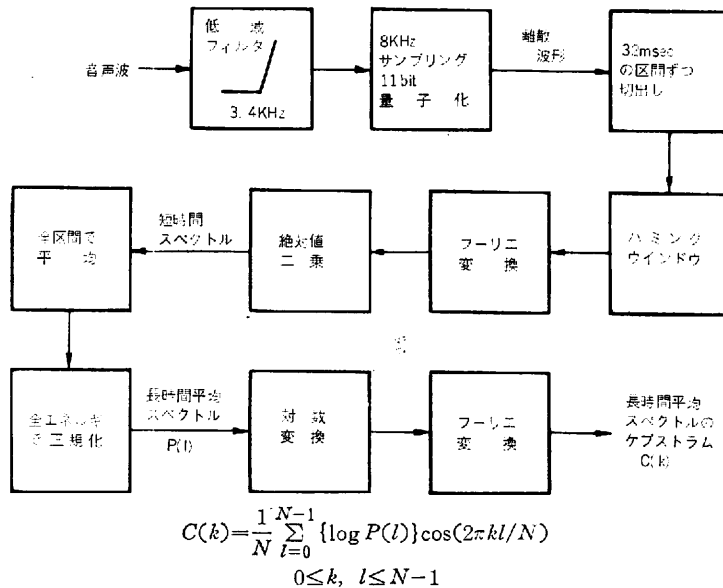


図 1 長時間平均スペクトルおよびケプストラムの測定方法

ms の区間 (256 点に相当) をとり出して、短時間パワースペクトル (密度) に変換する。この際、切出しによるひずみの混入を防ぐため、時間領域でハミングウィンドウをかけ合せておく。ハミングウィンドウは、つぎのように表わすことができる。 T_m は切出し区間の長さ (32 ms) である。

$$w(\tau) = 0.54 + 0.46 \cos \frac{\pi(2\tau - T_m)}{T_m}$$

$$0 \leq \tau \leq T_m$$

32 ms の区間ごとに求めた短時間パワースペクトルを、音声波形全体について平均して、長時間平均スペクトルとする。

つぎに、この長時間平均スペクトルを全エネルギーで正規化し、対数変換したのちにさらにフーリエ変換を行なって、ケプストラム⁽⁹⁾とする。定義から容易に知られるように、ケプストラムの低次の成分は、対数スペクトルパターンの概形を示し、高次の成分は微細構造に関与する。したがって、ケプストラムに変換することにより、両者を分離して取扱うことができる。またケプストラムは、対数スペクトルに関して線形直交変換を施したものであるから、ケプストラム領域での距離は、対数スペクトル領域での距離に比例する。

発声する内容は童話の一節 (10 秒程度) で、発声

速度、ピッチ (声の高さ)、発声レベルなどの条件については、発声者にとくに指示を与えないこととする。測定はおもに男性 9 名について、数日～3 カ月の時間間隔をおき、反覆して行なう。

2.2 測定結果

図 2 に長時間平均スペクトル (対数) の測定例を示す。(a) は話者 (発声者) によるパターンの相違を示すために、4 名の話者のそれぞれについて、ほぼ半年にわたるサンプルの平均パターンを示したものである。(b) は長期間にわたるパターンの変動を示すために、話者 S の、ほぼ 8 カ月にわたる 4 時期のパターンを示したものである。

3 話者認識

3.1 認識モデル

長時間平均スペクトルのケプストラムに関して、その時間的変動を考慮した認識モデルを構成し、話者識別および話者照合を行なう。ここで話者識別とは、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

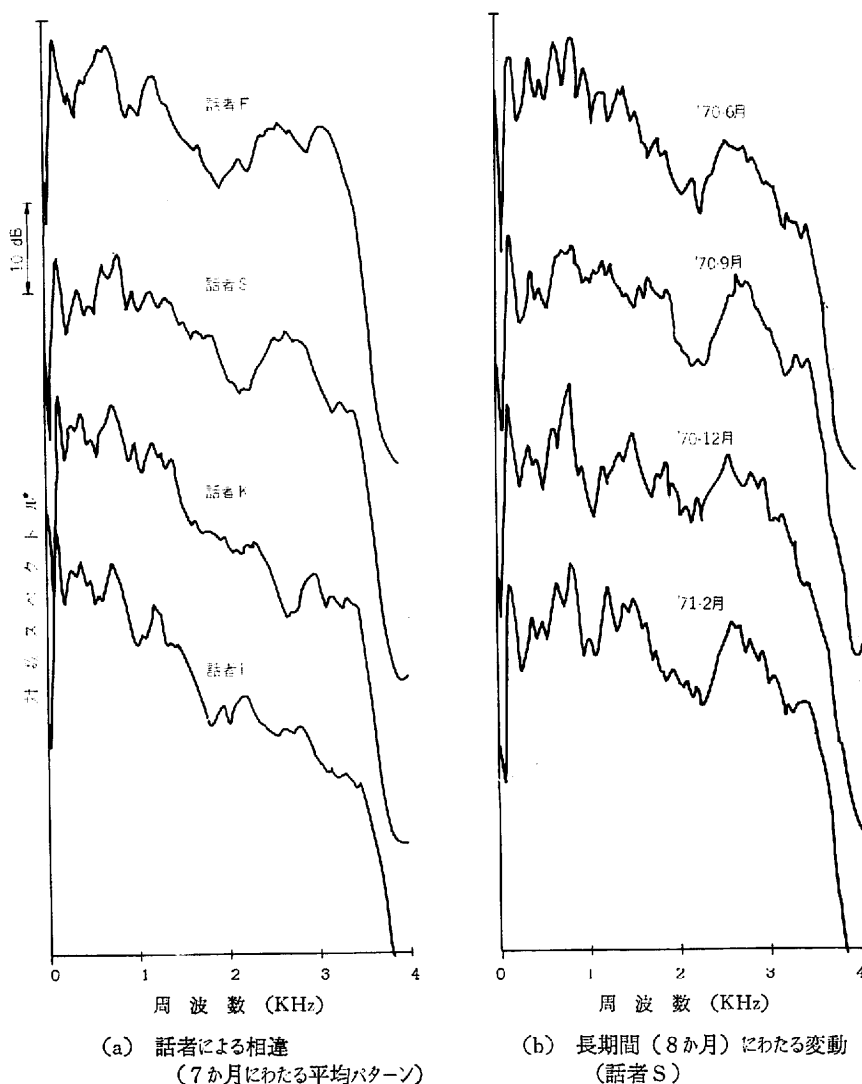


図2 長時間平均スペクトルの測定例

ものであるか否かを判定するもの(各登録話者についてクラス数=2)である。判定はつぎのような距離にもとづいて行なう⁽¹⁰⁾。

$$S(C, \{D_{rj}\}) = |V_r|^{1/N} (C - \bar{D}_r) V_r^{-1} (C - \bar{D}_r)^t \quad (1)$$

ここで、 C : 未知サンプルのケプストラム (N 次元ベクトル)

D_{rj} : 登録話者 r の j 番目の既知サンプルのケプストラム

$\bar{D}_r = \frac{1}{n} \sum_{j=1}^n D_{rj}$: 登録話者 r の既知サンプルの

平均値 (標準パターン)

V_r : 登録話者 r の共分散行列 ($N \times N$ 行列)

n : 各登録話者の既知サンプルの数

この距離は、各登録話者に属するサンプルの分布を、多次元空間における球に近づける線形変換を行なったのち、ユークリッド距離 (の2乗) に相当している。またこの距離にもとづく認識方法は、特徴ベクトルが多次元正規分布に従うことを仮定して、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、認識を行なうことに相当する。

ここで式(1)が意味をもつためには、 V_r が正定値であ

ることが必要であり、このためには、特徴ベクトルの次元 N と、既知サンプルの数 n に関して、 $N \leq n$ という条件が満たされなければならない。長時間平均スペクトルおよびケプストラムの値は、節 2.1 の定義およびこれらの偶関数性から明らかなように、実質的に 129 点について算出される。ケプストラムの成分のうち、スペクトルパターンの比較的微細な構造に関与する高次の成分は、認識のうえでの重要度が低いと考えられるので、かなり高次の $2/3$ の成分を無視して、低次の $1/3$ の成分を用いるとしても、次元数 N は 43 になり、かなり大きい。一方、時期的変動特性を反映した共分散行列 V_r を求めるためには、各登録話者について、ある程度の期間において既知サンプルを集めることが必要であるから、既知サンプルの数 n を大きくするのは、実際上容易なことではない。したがって、 $N \leq n$ という条件を満たすのは難しく、かりにこの条件が満足されても、 N と n との差が小さいときには、 V_r は大きな誤差を含む形で求まることが予想される。この対策として、各話者の共分散行列が類似していることを仮定し、 V_r のかわりにつぎのような行列を用いる。

〔方法Ⅰ〕

$$V_I = (V_{Iij}) = \frac{1}{m} \sum_{r=1}^m V_r \quad (2)$$

m : 登録話者の数

〔方法Ⅱ〕

$$V_{II} = (V_{IIij}) = \begin{pmatrix} V_{I11} & & 0 \\ & V_{I22} & \\ 0 & & V_{INN} \end{pmatrix} \quad (3)$$

〔方法Ⅲ〕

$$V_{III} = (V_{IIIij}) = \begin{pmatrix} V_{Ia} & 0 \\ 0 & V_{IIa} \end{pmatrix} \quad (4)$$

$V_{Ia} = (V_{Iij}): 1 \leq i, j \leq M, (M \times M \text{ 行列})$
 $V_{IIa} = (V_{IIij}): M+1 \leq i, j \leq N,$
 $((N-M) \times (N-M) \text{ 行列})$

方法Ⅱは、ケプストラムの各成分間の相関が小さいことを仮定して、 V_I の対角要素以外を 0 とおいたものである。方法Ⅲは、方法Ⅰと方法Ⅱの中間の形で、ケプストラムの低次の M 成分 ($1 < M < N$) に関しては各成分間の相関を考慮し、それよりも高次の成分に関して

は相関を無視するものである。

認識の際にパターンの時期的変動特性を考慮する必要性を確認するために、〔方法Ⅳ〕として、既知サンプルの平均値とのユークリッド距離 (の 2 乗) を用いる方法を試み、方法Ⅰ、Ⅱ、およびⅢの場合との比較を行なう。

話者識別の場合は、これらの距離に関して、未知サンプルからもっとも近い標準パターン (既知サンプルの平均値) 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し、それ以外のサンプルは、ほかの話者によるものであるとして、棄却するという方式を用いる。なおこのしきい値は、便宜上 2 種類の誤り率 (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。

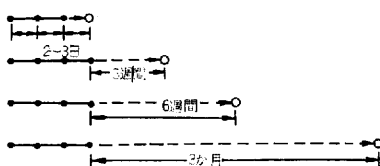
3.2 既知サンプルと未知サンプルの設定

既知サンプルの数と期間、既知サンプルから未知サンプルまでの時間間隔などが、認識率に及ぼす効果を調べるために、既知サンプルと未知サンプルを図 3 および表 1 に示すように設定して、4 種類の認識実験を行なう。

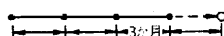
話者識別の場合の登録話者の数は 9 名とし、9 名のうちから 1 名を選ぶ形をとる。各登録話者の既知サンプルとして、各時期ごとに 3 回繰返して発声したサンプルを用い、未知サンプルとして、9 名が 3 回ずつ発声した 27 サンプルを用いる。話者照合の場合は、9 名の登録話者それぞれについて、受入れるべきサンプルとしてその話者が発声した 3 サンプルを用い、棄却すべきサンプルとして、登録話者と異なる 8 名または 26 名の話者 (詐称者) がそれぞれ 3 回ずつ発声した 24 または 78 サンプルを用いる。8 名の詐称者のグループを以後詐称者 α と呼び、26 名の詐称者のグループを詐称者 β と呼ぶ。 α 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、 β は α に他の 18 名を加えたものである。

認識に用いるケプストラムの次元 N および方法Ⅲの M の値は、予備実験の結果などをもとに、表 2 のように定めた。

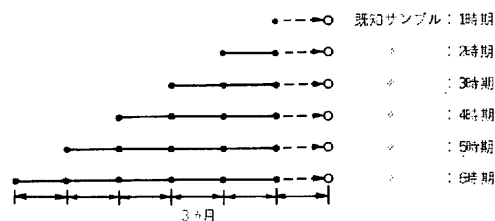
【認識実験1】



【認識実験2】



【認識実験3】



【認識実験4】

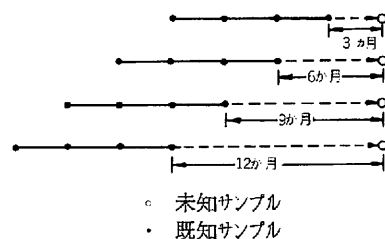


図 3 各認識実験における既知サンプルと未知サンプルの関係

表 1 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

	既 知 サ ン プ ル	既知サンプルから未知サンプル までの時間間隔	実験に用いる未知サンプルの数
認識実験1	2～3日おき3～4時期	2～3日から3カ月まで4通り	各4条件について1時期
認識実験2	3カ月おき4時期	3カ月	異なる4時期
認識実験3	1時期から3カ月おき6時期まで6通り	3カ月	各6条件について異なる4時期
認識実験4	3カ月おき4時期	3カ月から12カ月まで4通り	各4条件について1時期(共通)

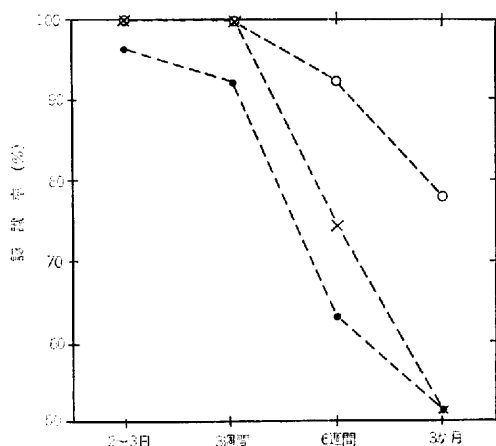
表 2 認識実験に用いるケブストラムの次元

	話 者 識 別	話 者 照 合
方法 I	N=30	N=30
方法 II	N=64	N=40
方法 III	N=30 M=10	N=30 M=10
方法 IV	N=64	N=40

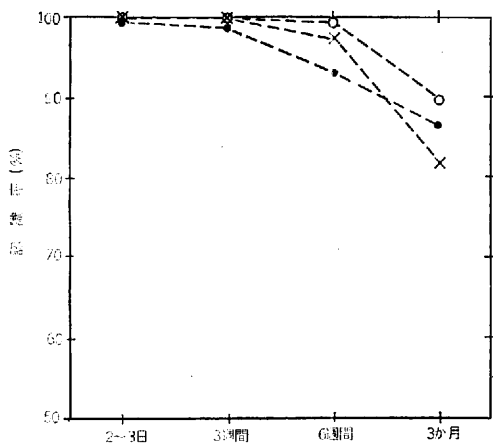
4 認識結果

4.1 認識実験1——短期間の既知サンプルによる認識——

図4は、既知サンプルから未知サンプルまでの時間間隔を変えたときの、9名の話者に関する平均認識率を、



未知サンプルまでの時間間隔
(a) 話者識別



未知サンプルまでの時間間隔
(b) 話者照合 (詐称者 α)

--x-- 方法 I
--o-- // II
--o-- // IV

図 4 認識実験 1 の結果

方法 I, II, IV の場合について示したものである。この実験結果について、各話者ごとの認識率を用いて、話者、認識方法 (I, II, IV)、既知サンプルから未知サンプルまでの時間間隔、およびこれらの交互作用に関する分散分析を行なった。その結果、話者識別と話者照合に共通して、時間間隔の効果および時間間隔と話者の交互作用の効果が、危険率 0.5% で有意差あり、話者および認識方法の効果が、危険率 1% で有意差あり

表 3 認識実験 2 の結果

	方法 I	方法 II	方法 III	方法 IV
話者識別	88.0%	89.8%	90.8%	76.0%
話者照合 (詐称者 α)	94.5	95.4	95.5	93.1
話者照合 (詐称者 β)	93.4	92.3	93.5	89.3

と認められた。

これらの結果から、既知サンプルが 10 日間程度の短期間のものである場合、既知サンプルから未知サンプルまでの時間間隔が 2~3 日、あるいは 3 週間程度に短いときには認識率が高いが、時間間隔が長くなるに従って単調に認識率が低下してしまい、3 カ月程度になると、きわめて低い認識率しか得られなくなってしまうことがわかる。また分散分析の結果から、話者によって長時間平均スペクトルパターンの安定性に差のあることがわかる。

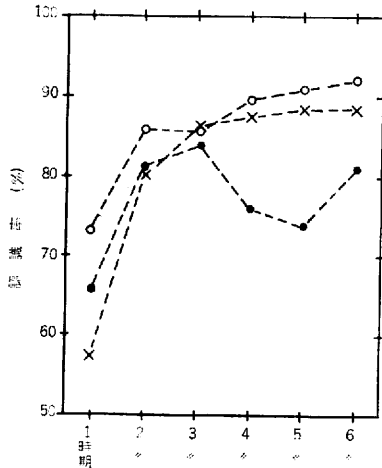
4.2 認識実験 2 —— 長期間の既知サンプルによる認識 ——

表 3 に、各認識方法を用いたときの、話者識別と話者照合の結果を示す。長期間にわたるデータを既知サンプルとして用いて、その変動を考慮した重みつき距離を設定することにより、最終の既知サンプルから未知サンプルまでの時間間隔が 3 カ月程度になっても、よい認識率の得られることがわかる。各認識方法のうちでは、方法 III による場合がもっともよい結果を示しているが、方法 I, II および III の差は小さい。

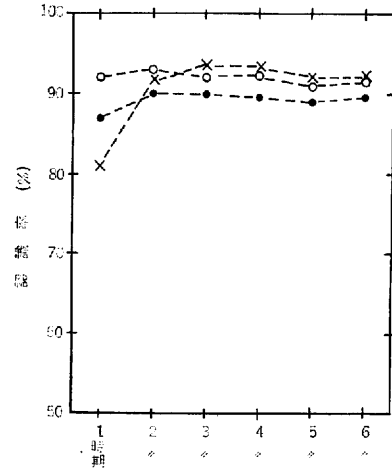
詐称者が α の場合と β の場合とで、話者照合の認識率に差がみとめられる。本質的に認識率に影響を及ぼすのは、詐称者の数ではなく、その声質であるから、詐称者のバリエーションをさらに多くした認識実験を行なえば、認識率は一定の値に収束するものと考えられる。

4.3 認識実験 3 —— 既知サンプルの数の効果 ——

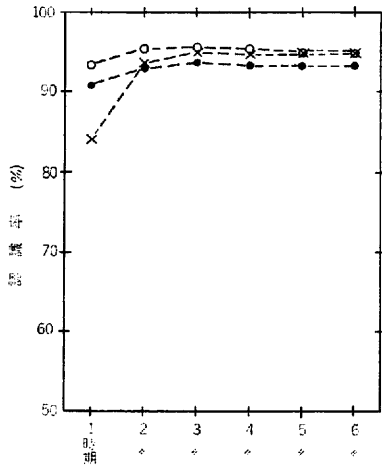
図 5 は、9 名の話者に関する平均認識率を、種々の認識方法の場合について示したものである。話者識別の結果を (a) に、話者照合の詐称者 α の場合の結果を (b) に、詐称者 β の場合の結果を (c) に示す。



既知サンプルの数
(a) 話者識別



既知サンプルの数
(c) 話者照合 (詐称者 β)



既知サンプルの数
(b) 話者照合 (詐称者 α)

図 5 認識実験 3 の結果

分散行列の推定誤差が極端に大きいためと考えられる。

4.4 認識実験 4 —— 長期間たった後のサンプルの認識 ——

結果を図 6 に示す。最終の既知サンプルから未知サンプルまでの時間間隔が 3 カ月から 12 カ月まで長くなった場合、重みづけ距離を用いなかったときには、時間間隔の増加に伴う認識率の低下がみられるが、方法 I および II を用いた場合については、そのような傾向はみられない。したがって、認識実験 2 で得られた結果のうち、方法 I ~ III の場合の結果は、未知サンプルまでの時間間隔を 3 カ月から 12 カ月程度にまで延長した場合についても、通用するものと考えられる。

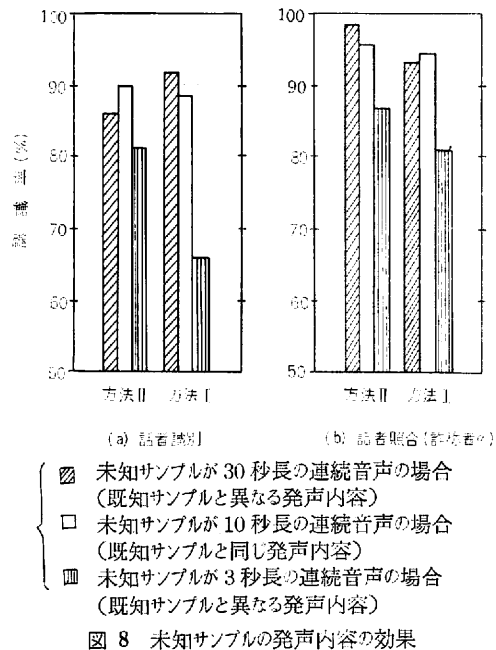
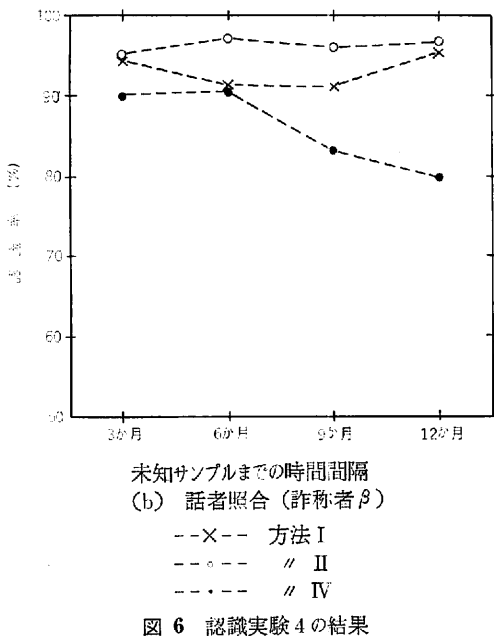
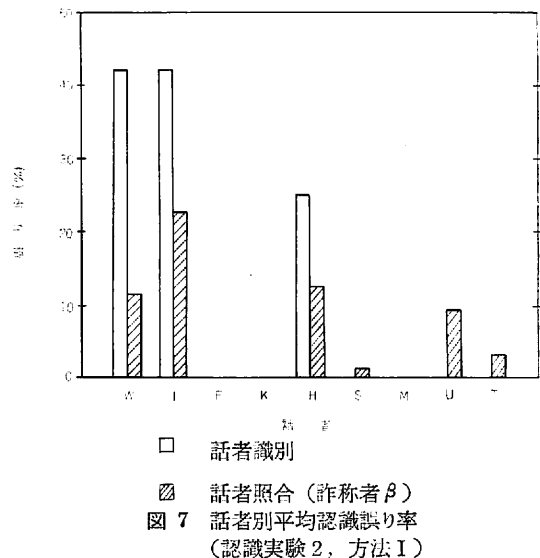
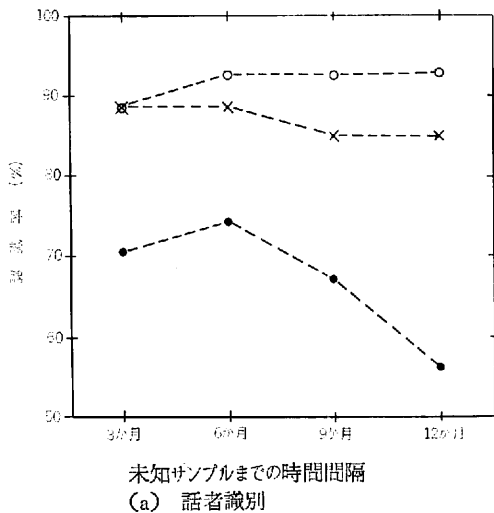
4.5 検討と補足実験

4.5.1 話者による相違について

認識実験 2 において方法 I を用いた場合の認識結果を、話者別に集計して図 7 に示した。この結果からわかるように、話者による誤り率 (認識率) の相違はいちじるしく、その分布の様子は、話者識別と話者照合とで、きわめて類似している。

これらの結果から、話者識別と話者照合に共通して、距離尺度の設定法にかかわらず、既知サンプルが 1 時期のみのときには、認識実験 1 の、3 カ月後の未知サンプルに関する結果と同程度の低い認識率しか得られず、よい認識を行なうためには、ある程度長期間にわたったサンプルが必要であることがわかる。また、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 カ月おきの 4 時期程度にわたる既知サンプルがあれば、ほぼ十分であることがわかる。

既知サンプルが 1 時期の場合に、方法 I の結果が重みづけを全く行なわない方法 IV の結果よりも悪いのは、共



4.5.2 音声サンプルの長さ

話者の特徴パラメータとして長時間平均スペクトルを用いる利点の1つは、発声内容にあまり影響されない点にある。しかし、そのように音韻系列の影響が無視できるためには、どの程度の長さがあればよいかという点は、必ずしも明らかではない。これまでに用いているサンプルは、10秒長程度の連続音声であるが、これが必要十分な長さであるかどうかを調べるため、つぎのような補足的実験を行った。

既知サンプルにはこれまでと同じものを使い、未知サンプルとして、異なる内容の30秒長程度と3秒長程度の2種の連続音声を用いて、既知サンプルと同じ内容の連続音声を用いた場合の結果と比較した。3カ月おきに4時期にわたって求めた既知サンプルを用いて、その最終サンプルから3カ月後のサンプルを認識する場合 (認識実験2と同じ条件) の結果を図8に示す。

30秒長のサンプルの場合は、10秒長のサンプルすなわち既知サンプルと同じ文章の場合と、ほぼ等しい認識率

が得られているが、3秒長のサンプルの場合は、それに比較してきわめて低い値しか得られていない。これは、10秒以上のサンプルによるパターンに対して、3秒程度のサンプルによるパターンが大きく異なっていることを示している。上の結果から、長時間平均スペクトルパターンを求めるサンプルとして必要な長さは、ほぼ10秒程度であると考えられ、本論文で用いているサンプルは、発声内容が無視できるといふ条件を、ほぼ満たしていると考えられる。

4.5.3 理想的な重みづけによる場合

距離尺度の設定に用いる共分散行列 V_I , V_{II} , および V_{III} の値を、既知サンプルのみでなく、未知サンプルも加えて求め（標準パターン $\bar{D}_r$ の算出には未知サンプルを含めない）、これを用いて認識実験2の形で実験を行なったところ、ほぼ100%の認識率が得られた。この結果から、理想的な重みづけが行なえれば、100%程度の認識率を達成することが可能であること、すなわち長時間平均スペクトルパターンには、十分な個人性情報が含まれていることがわかる。

4.5.4 話者照合の判定のしきい値

話者照合における判定のしきい値を、あらかじめ全登録話者に共通に、一定の値に設定した場合について、補助的に検討した。その結果、これまでのように事後に最適値に設定する場合に比べて、認識率は若干低下するが、その値は3%程度におさえられる見通しが得られた。

実際の話者照合の場合には、2種類の誤り率に対する要求（重要度）にしたがって、しきい値を決定することになる。

5 長時間平均スペクトルパターンの変動

5.1 認識実験の結果から

認識実験1および3の結果は、長時間平均スペクトルパターンが、同時期あるいは10日間程度の短期の間ではきわめて安定しているが、その後3カ月程度まで時期がたつに従って、ゆっくり大きく変わってくることを示している。また認識実験4の方法IVによる結果は、3カ月以上の長期間にわたっても、そのパターンが時期の経過に従って、元の位置から離れるような方向にゆっくり変わ

ってゆくことを示している。

しかしその変動は際限のないものではなく、認識実験2, 3, および4の結果、ならびに項4.5.3の結果が示すように、3カ月～12カ月の時期をおいたサンプルでも、長期間（9カ月程度以上）の変動を考慮した適切な変換を行なえば、各話者固有の領域におさめることができる。

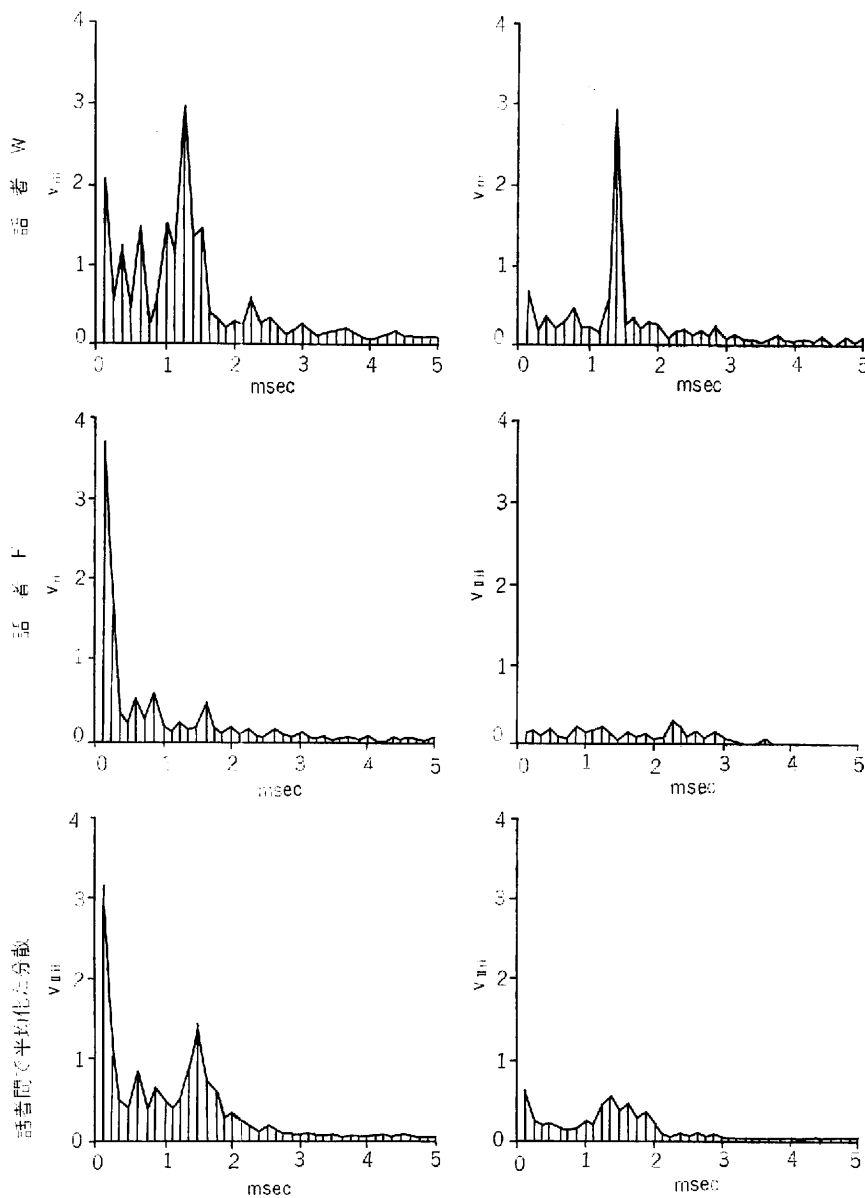
5.2 ケプストラムの各成分の分散

長時間平均スペクトルのケプストラム各成分の時期的変動の大きさを評価するため、短期間と長期間の2種の時期にわけて、それらの分散を求めた。短期間の変動を調べるサンプルとしては、2～3日おきに4回測定したスペクトルデータを用い、長期間の変動を調べるサンプルとしては、3カ月おきに6時期にわたって測定したデータを用いた。図9は、2名の話者のおのおのの分散のパターンと、9名の話者について平均化したパターンを示したものである。

この検討によって、たとえば話者Wにみるように、短期間と長期間においてほぼ等しい次数の付近に、特異なピークを生ずる話者の多いことがわかった。またすべての話者に共通して、長期間になるに従い、低次の成分の分散が大きくなる傾向がみとめられた。これは、長期間になるに従い、長時間平均スペクトルパターンの全体的な形の変動が、比較的細かい構造の変動に比べて大きくなることを示している。

5.3 主成分分析

長時間平均スペクトルパターンの、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、主成分分析を行なった。分析には、長時間平均スペクトルのケプストラムの0～12次の成分を用い、それらの成分からなるベクトルの時期的変動に関する共分散行列を、相関行列に変換してから分析を行なった。長期間のサンプルとしては、3カ月おきに約2年間にわたって求めたサンプルを用い、短期間のサンプルとしては、図10に示すように、2～3日から3カ月までの種々の間隔をおいて、約5カ月間にわたって求めたサンプルを用いた。



(a) 長期間(15か月)の場合

(b) 短期間(約10日間)の場合

図 9 長時間平均スペクトルのケプストラムの各成分の分散

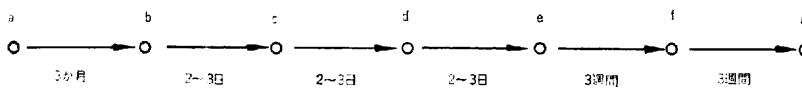


図 10 短期間のサンプルの時間間隔

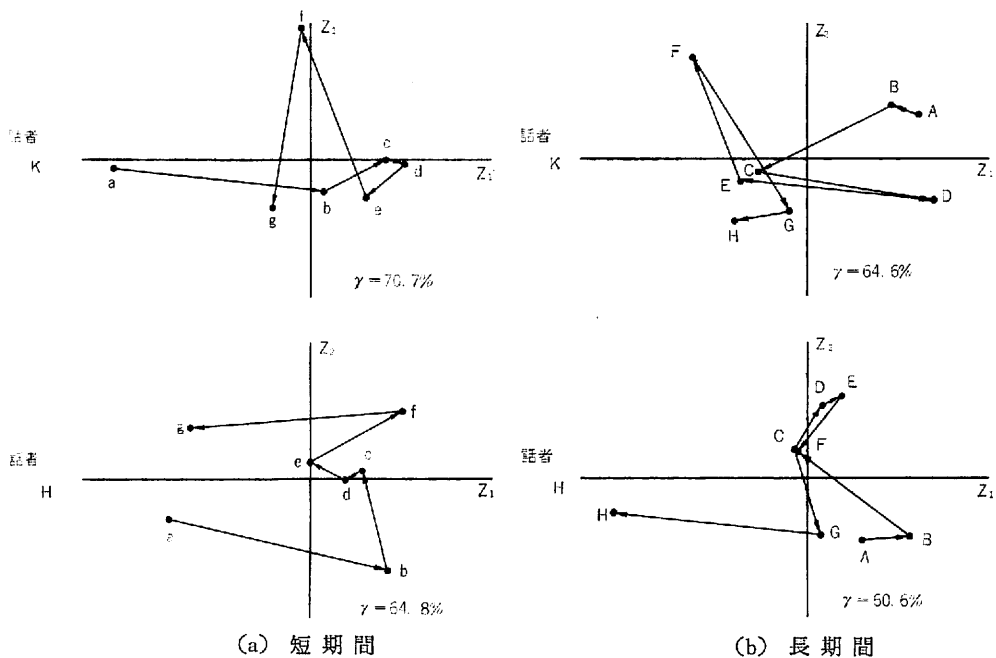


図 11 長時間平均スペクトルの短期間と長期間の変化(主成分分析結果)

z_1 : 第1主成分, z_2 : 第2主成分, γ : 累積寄与率

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結果の一部を, 図11に示す。短期間の結果を(a)に, 長期間の結果を(b)に, 2名の話者について示した。横軸が第1主成分(第1主軸), 縦軸が第2主成分(第2主軸)である。各図の下に, 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カ月の間隔を置いて, 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は見出されない。

5.4 時間間隔と変化量の関係

長時間平均スペクトルのケプストラムの0~63次の64成分を用いて, 各サンプルを64次元空間内の一点と考

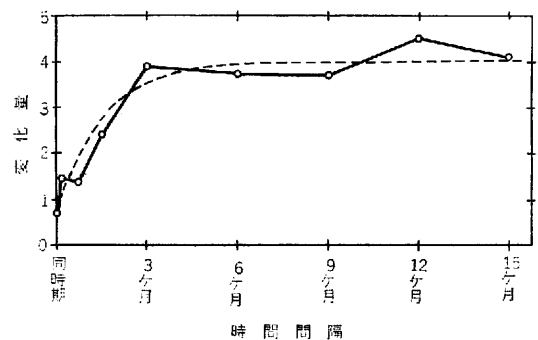


図 12 長時間平均スペクトルのケプストラムの変化量と時間間隔の関係

え, 種々の時間間隔をおいたときの変化量を, ユークリッド距離(の2乗)を用いて求めた。図12は, その結果を9名の話者に関して平均したものである。

この図をみると, 時間間隔が3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, 時間間隔の長さに関係にはほぼ一定になっている。この結果は, 長時間平均スペクトルパターンが, 3カ月程度までの間では, 時間の経過とともに, 元のパターンから平均的に遠ざかる方向にゆっくり変化する, 多次元ランダムウォーク的傾向を有す

る変化をするのに対し、3カ月以上の長い期間を置いてサンプリングしてみると、ある平均パターンを中心に不規則に変動する性質を有することを示している。これは、認識実験1～4で得られた種々の結果とよく対応している。

ユークリッド距離（の2乗）は、近似的にパラメータの自己相関係数を反転したものと対応づけることができる。図12の曲線を指数曲線とみなして、あてはめを行なったところ、同図の破線の結果が得られた。これは、不規則信号を、ほぼ1.5カ月の時定数を有するRCフィルタに通した成分と、通さない小振幅の成分と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、パターンの時期的変動に対応し、後者の成分は、時間間隔の長さに無関係な変動に対応すると考えられる。

6 む す び

これまでの実験と考察により、長時間平均スペクトルパターンは、2～3日から3週間程度の間では、きわめて安定しているが、その後時期がたつに従って、同じ発声者によるものでも、ゆっくり大きく変化する性質を有することが明らかになった。その時期的な変動特性は、1.5カ月程度の時定数を有する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ことがわかった。

認識実験を通じて、そのパターンの長期間にわたる変動を考慮した適切な距離尺度を設定すれば、3カ月以上の長期間をおいた後でも、話者の認識が十分に可能であることが確認され、長時間平均スペクトルパターンに、安定した個人性情報が含まれていることが確かめられた。

3カ月おきに4時期にわたって測定した長時間平均スペクトルを既知サンプルに用いて、その最終の時期の既知サンプルから3カ月後の未知サンプルを認識する場合、9名から1名を選ぶ話者識別で平均90.8%、詐称者が26名の話者照合で平均93.5%の認識率が得られた。

今後、長期間にわたる長時間平均スペクトルパターンの変動特性の解明、その特性をさらに適確にとり入れた距離尺度の設定法、長時間平均スペクトルパターンを決定している要因の追求、各音韻に特有な個人性情報を用いた話者認識方法⁽¹¹⁾との比較をはじめ、その他の種

々の問題点について検討を進めていきたい。

長時間平均スペクトルの時期的変動特性は、音声スペクトルパターンの平均的な性質を示していると考えられ、本論文の解析にもとづく結果は、他のスペクトルパラメータによる話者認識方法においても、考慮すべきものであると考えられる。

謝 辞

口ごろ、ご指導いただく当研究所野田基礎研究部長、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，No.116，p.49，1973.
- (2) W. A. Hargreaves and J. A. Starkweather：Recognition of Speaker Identity, Language & Speech, No.6, p.63, 1963.
- (3) P. D. Bricker et al.：Statistical Techniques for Talker Identification, Bell Syst. tech. J., 50, No.4, p.1427, 1971.
- (4) S. K. Das and W. S. Mohn：A Scheme for Speech Processing in Automatic Speaker Verification, IEEE Trans. AU, AU-19, No.1, p.32, 1971.
- (5) J. E. Luck：Automatic Speaker Verification Using Cepstral Measurements, J. Acoust. Soc. Amer., 46, No.4-2, p.1026, 1969.
- (6) 落合・福村：音声通信における基本品質について，通研実報，8，No.2，p.1469，1959.
- (7) H. Mangold：Die Individuellen Unterschiede bei Langzeitmessungen am Sprachsignal, NTZ, No.6, p.364, 1969.
- (8) G. Ramishvili：Some Problems in Automatic Identification of Spoken Utterances and Speakers, STL-QPSR, 2, No.3, p.36, 1970.
- (9) B. P. Bogert et al.：The Quefrency Analysis of Time Series for Echoes, Proc. of the Symposium on Time Series Analysis, p.209, 1963, John Wiley & Sons, New York.
- (10) 古井・板倉・斎藤：長時間平均スペクトルによる話者認識，電通学誌，55-A，No.10，p.549，1972.
- (11) 古井・板倉：単語の統計的パラメータによる話者認識，電通学誌，56-A，No.11，p.717，1973.

(1974, 2, 8 受付)