

論文 / 著書情報
Article / Book Information

論題(和文)	音声スペクトルの時間変動特性に着目した話者認識
Title(English)	
著者(和文)	古井 貞熙, 斉藤 収三
Authors(English)	SADAOKI FURUI
出典(和文)	電子通信学会, PRL78-30, , pp. 71-80
Citation(English)	, PRL78-30, , pp. 71-80
発行日 / Pub. date	1978, 7
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) 1978 Institute of Electronics, Information and Communication Engineers.

音声スペクトルの時間変動特性に着目した話者認識

古 井 貞 熙・斎 藤 収 三

(武蔵野通研)

1978年7月10日

社団法人 電 子 通 信 学 会

音声スペクトルの時間変動特性に着目した話者認識

TALKER RECOGNITION BY DYNAMICAL
FEATURES OF SPEECH SPECTRA古井 貞熙 斎藤 収三
Sadaoki FURUI Shuzo SAITO

日本電信電話公社 武蔵野電気通信研究所

Musashino Electrical Communication Laboratory, NTT

Abstract: New talker recognition methods by dynamical features of speech spectra have been investigated. The time pattern of both fundamental frequency and log-area ratios is extracted from speech signal, and pattern matching with time-normalization is applied using weighted distance based on the long-term variability of these parameters. From the warping function of the time-normalization, some specific regions where the time structure is highly stable can be found in the time pattern. The measure of stableness in the specific region can be used effectively as one of the measures of personal information.

I. まえがき 音声スペクトルは発声器官の動きに伴って時間的に変化しており、その時間変動特性には個人差があると考えられる。筆者らはこれまで、話者認識のための特徴パラメータとして、主として音声スペクトルの統計的特徴をとり上げ、その長期間にわたる変化特性などに着目した種々の分析と、話者認識実験を行って来た。その結果、短文音声の長時間平均スペクトルや、単語音声の基本周波数および対数断面傾度パラメータ (PARCOR係数の $\tanh^2$ 変換) の統計的特徴が、話者認識のためのパラメータとして有効であることがわかり、後者の特徴を4種類の単語音声から抽出して用いれば、学習サンプルから3か月後の未知音声の話者を、99%の精度で認識できることが明らかにされた。しかしこれらの特徴パラメータには、同じ話者が同じ単語を発声した音声から抽出したものであっても、数か月～1年程度の期間で見ると、あまりに変化する性質があるため、3か月あるいはそれ以上の間隔を置いたサンプルの話者を高い精度で認識するためには、1年程度の長期間にわたって学習

を行う必要がある。学習サンプルの収集期間が短いと、未知音声との時期差が大きくなるに従って、誤り率が单调に増加する^{(1)~(3)}。

一方、音声スペクトルの時間パターンそのもののマッチングによって話者認識を行う試みも、これまでに多数報告されており、音声スペクトルを表現するパラメータとしては、フィルタバンク出力、ホルマント周波数⁽⁴⁾、相関係数⁽⁵⁾、PARCOR係数⁽⁷⁾⁽⁸⁾などが検討されている。しかしこれらの検討においては、特徴パラメータの時間的変動はほとんど考慮されていない。

本報告では、まず音声スペクトルの時間パターンと、特徴パラメータの長期間変動特性を考慮しながらマッチングする手法について提案し、この手法による話者認識実験と、これに関連した種々の検討を行う。この検討においては、これまでの統計的特徴による結果との対比を行い、統計的特徴の一部を組合せた認識方法についても検討する。次に音声スペクトルの時間変動特性に含まれる個人性情報を抽出する新しい方法を提案し、この方法の有効性を実験的に確かめる。

II 音声スペクトルの時間パターンによる話者認識

2.1. 音声分析

- ・音声サンプル：単語音声 /namee/, /baNgo:/
- ・話者と発声時期：男声9名が約3年の間に種類の時間間隔をわけて繰返し発声。

図1に示す通り、比較的長期間にわたるデータの組と比較的短期間のデータの組の2種類を用いた。各時期に各話者は、同じ単語を、1分

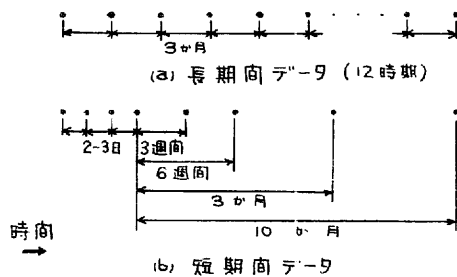


図1. 実験に用いた音声データ

程度の間隔をわけて繰返し発声している。

これらの音声を3.4kHzの低域通過フィルタに通したのち、8kHz、11ビットで標準化および量子化を行う。10ms毎に32msの区間(フレーム)を切り出し、ハミングウィンドウを乗じてから、1~12次の対数断面積比パラメータ(LAR)と基本周波数を抽出し、組合せて13次元の特徴ベクトルとする。LARを抽出したのちの残差の相関関数(変形相関関数)のピーク値とパワーに閾値を設けて、両者が閾値を越す区間を有声音区間とみなし、認識には有声音区間のみの時系列を用いる。

2.2. 話者認識 一般に話者認識は、話者識別と話者照合にわけることができる。前者は、入力音声からあらかじめ登録されている複数の話者のうちのいずれによるものかを判定するものであり、後者は、入力音声の、対象としている登録話者のものであるか否かを判定するものである。話者識別の場合は、入力音声との類似度が最も大きい標準パターンと話者を選び、話者照合の場合は、標準パターンとの類似度が最も大きい音声は本人のものと思われ、他は棄却する。本報告では、登録話者が9名の場合について、話者識別と話者照合の両方の実験を行う。

2.3. 時間軸非線形伸縮マッチング 同じ話者が同じ単語を繰返し発声して、そのスペクトルの

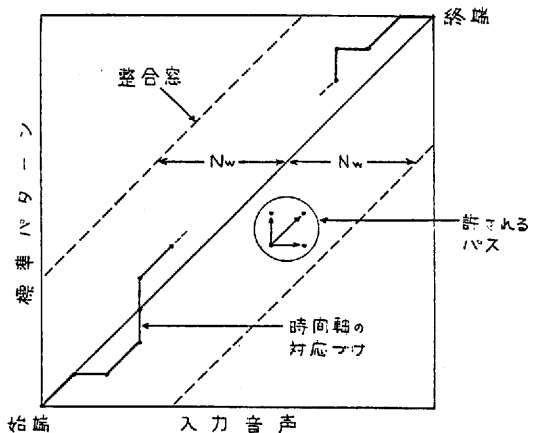


図2. 時間軸非線形伸縮マッチング

時間構造は種々に変化するので、これらの時系列間(標準パターンと入力音声)のマッチングを行うためには、時間軸の非線形伸縮を行う必要がある。ここでは、2つの時系列の時軸の最適対応を行う方法として、広く用いられているDPによる方法⁽⁹⁾を用いる。図2に示すように、対応する2つの時系列が作る平面上の、始端と終端を結ぶ対角線の両側に N_w フレームずつの整合窓を設けて、その中で対応づけを行い、DPパスの進む方向はこの平面上で、縦・横斜めの3方向のみとした。

2.4. 重みつき距離 2つの時系列の対応するフレーム間の距離には、特徴ベクトルの時期的変動を考慮した重みつき距離を用いる。入力音声の i 番目のフレームの特徴ベクトルを x_i 、話者 k の標準パターンの j 番目のフレームの特徴ベクトルを y_{tkj} であらわすと、両者の重みつき距離は、

$$d(t, t', k) = (x_t - y_{tkj})^T W (x_t - y_{tkj}) \quad (1)$$

ただし、 $W = \text{diag}(w_1, w_2, \dots, w_{13})$ で、 $\{w_i\}_{i=1}^{13}$ は次のようにしてもとめる。

話者 k の j 番目の学習サンプルの i 番目のフレームを z_{tkij} であらわす。 z_{tkij} の全有声音区間にわたる平均値 μ_{kj} をとめ、 μ_{kj} の全学習サンプルにおける分散 σ_{kj}^2 を計算する。

$$\mu_{kj} = \frac{1}{N} \sum z_{tkij} \quad (2)$$

$$\sigma_{kj}^2 = \text{Var} \mu_{kj} \quad (3)$$

σ_{kj}^2 を全登録話者について平均し、この各要素の逆数を w_i とする。

$$\bar{\sigma}^2 = \frac{\sum \sigma_k^2}{K} \quad (4)$$

$$W_i = 1/\bar{\sigma}_i^2 \quad (5)$$

2.5. 話者認識実験 図1に示した2種の音声データを用いて、話者認識実験を行う。長期間データを用いる場合は、各話者毎に、3か月毎の4時期(12サンプル)のデータを学習サンプルとし、その3か月後のデータを認識する。全部で12時期の音声のうち、未知音声としては8時期のものが使えるから、全話者の未知音声は、8時期×3回×9名=216サンプルとなる。

短期間データを用いる場合は、はじめの3~4時期のデータを学習サンプルとし、その後の各時期のデータを認識する。学習サンプルの収集期間は、長期間データの場合に約10か月、短期間データの場合に約10日間となる。

各話者の標準パターンは、学習サンプル間のDPマッチングにより、対応するフレーム毎に平均化を行い、作成する。2種類の単語とそれぞれ認識の判定に用いる場合のほか、両者の単語を組合せて判定に用いる場合についても実験を行う。

2.6. 長期間データによる認識実験の結果

話者識別において、DPマッチングの整合窓の窓幅を変えたときの、2種類の単語の全未知音声に関する平均認識誤り率を図3に示す。この図には、10数字音声の単語認識の場合の結果⁽¹⁰⁾と比較して示してある。整合窓の窓幅が狭いほど計算量が少なくなることと考慮すると、この結果から、窓幅の最適値は、話者識別と単語認識に共通して、ほぼ150ms程度であることがわかる。

整合窓の窓幅を200msとしたときの、話者識別と話者照合の結果を図4に示す。図の左端の点が、ここで検討している時間パターンによる認識結果であり、右端の点は、統計的パラメータ(単語有音声区画における基本周波数とLARの平均値と相関行列)を用いる場合の結果である。

この結果から、両者の特徴パラメータと識別および照合に共通して、/namae/よりも/baŋgo:/の誤り率の方がやや小さく、2種の単語を組合せると、誤り率は1単語の場合の平均誤り率の約1/2に低下することがわかる。1単語の場合の平均誤り率と、2単語の組合せのときの誤り率はいずれも、時間パターンによる場合よりも統計的パラメータによる場合の方がやや小さい。

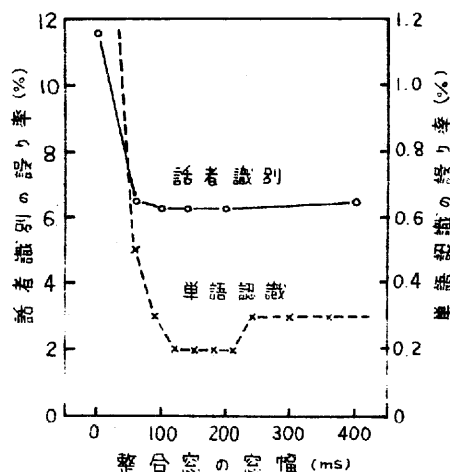
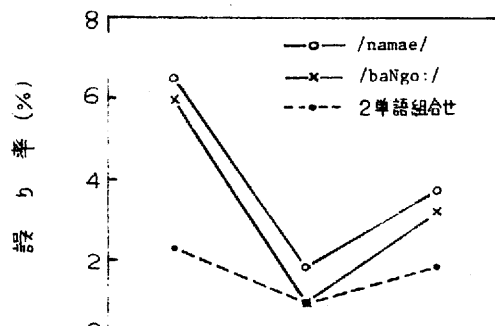
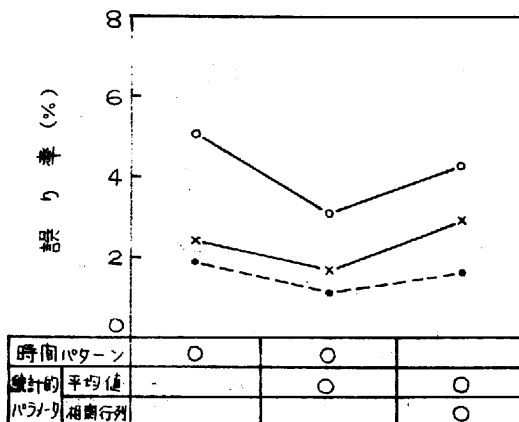


図3. 整合窓の窓幅と認識誤り率の関係
(長期間データによる話者識別の結果と、単語音声認識の結果の比較)



(a) 話者識別



(b) 話者照合

図4. 長期間データによる話者認識実験の結果
(○: 認識に用いる特徴パラメータ)

2.7 時間パターンと統計的パラメータの組合せ

時間パターンのマッキング結果に、統計的パラメータの一部である平均値ベクトルの重みつき距離（重みは時間パターンのマッキングの際に用いたものと同じ）を加えたものを認識の判定に用いて、この組合せ効果を調べた。両者の距離を加える際には、それぞれの距離の絶対値の差をバランスさせるように、適当な重みを与えた。このときの認識結果を、図4の中央部の点として示した。組合せ効果により、話者識別、話者照合 いずれの場合も、誤り率が大きく低下することがわかる。このときの1単語の場合の平均認識率は、話者識別で98.7%、話者照合で97.6%、2単語を組合せた場合の平均認識率はそれぞれ、99.1%と98.9%である。話者照合の登録話者別の誤り率を図5に示す。誤りを比較しやすい話者と生じにくい話者があるが、あまり極端な誤りの集中はみられない。

2.8 短期間データによる認識実験の結果 短期間データを用い、整合窓の窓幅を変えたときの、各時期の未知音声の話者識別結果を図6に示す。時間パターンのみを用い、統計的パラメータの組合せは行っていない。時期差が2-3日から10か月のいずれの場合も、窓幅が200ms以上のところで誤り率が最低となり、時期差とともに単調に誤り率が増加する傾向がみられる。

図7に、窓幅を200msとしたときの、話者識別

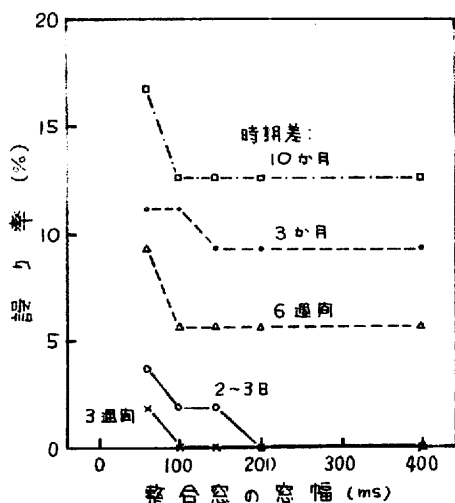


図6. 短期間データによる話者識別の結果

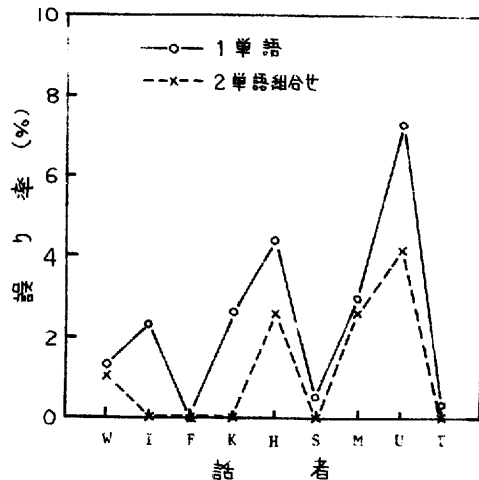


図5. 時間パターンと統計的パラメータの組合せによる話者照合の登録話者別誤り率

と話者照合の結果を示した。比較のために統計的パラメータによる場合の結果をよわせて示してある。この結果から、学習サンプルが10日間程度の短期間の場合には、未知音声との時期差が長くなるに従って単調に誤り率が大きくなり、時期差が10か月になると、誤り率は10%程度になってしまうことがわかる。この傾向は統計的パラメータの場合の結果と共通しているが、誤り率の増加の割合は、時間パターンの場合の方がやや小さい。

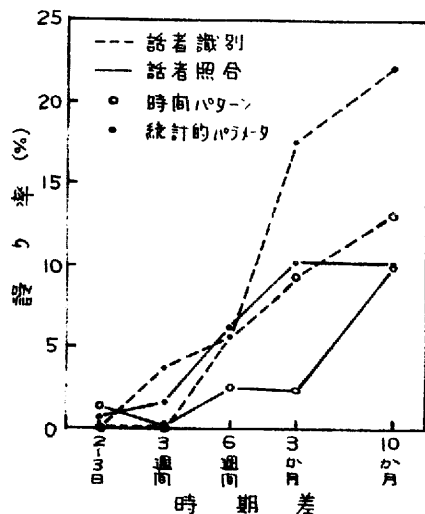


図7. 短期間データによる話者識別と話者照合の結果（整合窓の窓幅=200ms）

2.9. 学習サンプル数の効果 図1の長期間データを用いて、各話者の学習サンプル数が認識率におよぼす効果を調べた。2.6節の実験では、各話者の学習サンプルとして4時期の12サンプルを用いたが、ここでは1時期の1サンプルをそのまま標準パターンとして用いる場合と、3時期から1サンプルずつとり出して3サンプルによって学習を行う場合について話者識別実験を行った。実験には単語 /name/ を用いた。12サンプルを用いる場合を含めた結果を図8に示す。学習サンプル数の増加とともに単語に誤り率が低下し、この両者には対数平面上で線形な関係がある。

2.10. 重みつき距離と特徴ベクトルの有効性 DPマッチングを行う平面上での特徴ベクトルの距離として、これまでは、その時期的変動を考慮した重みつき距離を用いて来たが、ここでは単純なユークリッド距離を用いた話者識別実験を行って、重みづけの効果を検討した。実験には長期間データを用い、4時期の学習サンプルを用いた。結果は図9に示す通りで、2種のいずれの単語の場合も、重みづけを行わないと、誤り率が約1.5倍に増加することになる。

つぎに、これまでのようにLARの重みつき距離を用いずに、対数スペクトルのユークリッド距離を用いる場合の話者識別実験を行い、これまでの結果との比較を行った。平滑化した対数スペクトルの距離を計算する方法として、12次までのケプストラムのユークリッド距離を用いた。学習サンプルが3サンプルの場合の /name/ による認識実験の結果を図10に示す。重みつきLAR

を用いる場合に比べて、対数スペクトルのユークリッド距離を用いると、誤り率は約1.5倍に増加する。以上の結果は、ここで用いている特徴ベクトルおよび距離尺度の有効性を示している。

2.11. 非線形伸縮マッピングの効果 これまでは時間軸の対応づけに、端点を固定したDPマッチングを用いて来たが、ここでは端点をフリーにした非線形と線形の時間軸伸縮マッピングを行い、その結果とこれまでのものとを比較した。端点フリーの非線形伸縮の場合は、極端な対応づけを防ぐため、局所的に2倍以上の伸縮がみこられないよう制限を加えた⁽¹¹⁾。実験には図1の長期間データを用い、学習サンプルは4時期とした。認識の判定に1単語を用いる場合と、2単語を組合せ

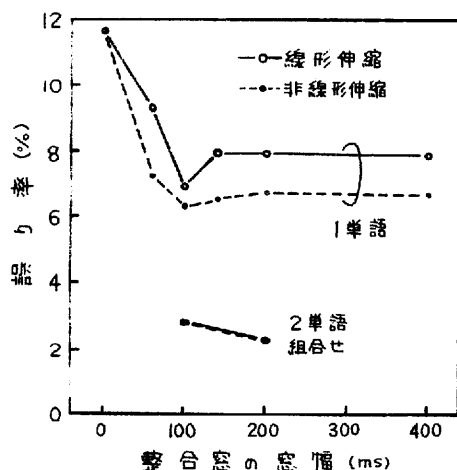


図11. 端点フリー線形/非線形伸縮マッピングによる話者識別実験の結果

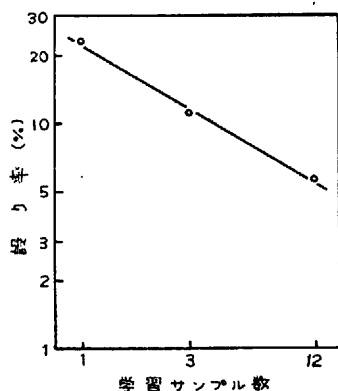


図8. 学習サンプル数の効果

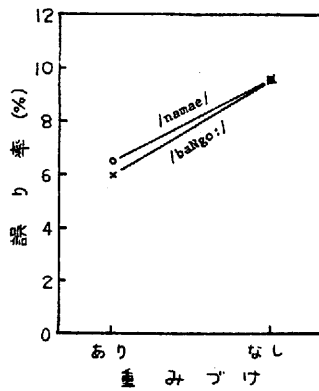


図9. 重みつき距離の効果

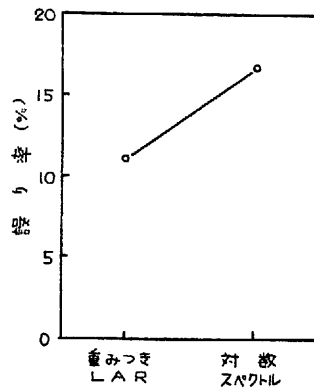


図10. 特徴ベクトルによる相違

る場合の話し者識別実験を行った。整合窓の窓幅を変えたときの誤り率を図11に示す。

認識の判定に1単語を用いる場合については、次のことがわかる。整合窓の窓幅が0のときは端点固定の線形伸縮に相当し、このときの誤り率は極めて大きい。非線形伸縮の場合の結果を図3の端点固定の場合と比較すると、最適窓幅(100ms)の場合の両者の誤り率は等しいが、端点固定の場合の方が最適窓幅の範囲が広く、主眼に端点と固定しようがないと言え、線形伸縮の場合、窓幅が100msより小さい非線形伸縮の場合に近い誤り率が得られているが、他の条件では誤り率10~20%と高くなる。

2単語と組む時、窓幅200msとしたときの誤り率は、端点固定非線形、端点フリー非線形、および端点フリー線形伸縮で等しい。

Ⅱ 安定度による動的な個人性情報抽出

3.1. DPパズによる時間軸の伸縮特性の検討 同じ話し者が同じ単語を連続して発声しても、そのスペクトルの時間構造は様々に変化するが、各話音に固有のあまり変動しない時間構造も存在すると思われる。この様子を男性9名による単語/namae/の長期間データを用いて、同一話し者の異なる時期のサンプル間と、異なる話し者のサンプル間の時間軸非線形伸縮マッチングの結果から(DPパズ)の結果により、定期的にみた、特徴ベクトルは基本周波数とLARの3次元で、DPマッチングは端点固定の方法で行った。

3か月おきの8時期の音声を用い、初めの3時期から1サンプルずつ各話し者の学習サンプルを取り出して、時間軸伸縮を行って33サンプルと重ね合わせ、各話し者の標準パターンを作る。残りのサンプルを入力音声として、標準パターンとの時間軸伸縮マッチングを行う。整合窓の窓幅は200msとする。

時間軸の対応づけの結果の一例(登録話し者がKの場合)を図12に示す。入力音声の話し者が標準パターンと同じ場合とは、DPパズの対角線からのずれがやや小さいほか、標準パターンのフレームによって、時間軸の伸縮があまりやさしいフレームと、比較的伸縮があまりにくく対角線方向に進む割合が大きいフレームがあることがわかる。後者は、7人の話し者に共通して、1単語中に3~4か所程度見出される。要

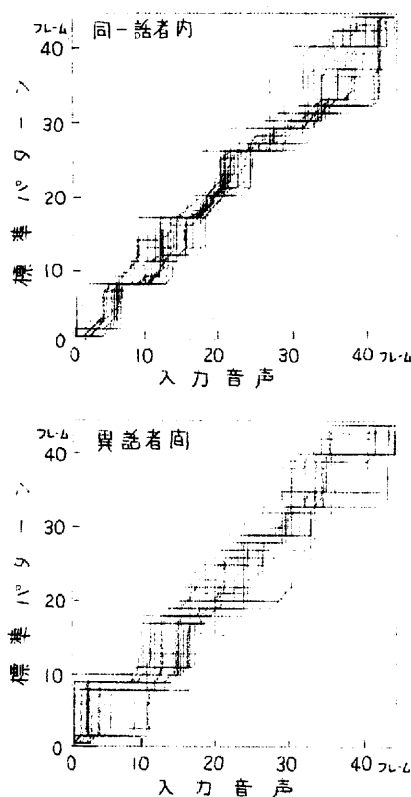


図12. 時間軸の対応づけの結果の一例 (/namae/, 標準パターン: 話し者K)

なる話し者との対応づけの場合には、このようなフレームの区別はあまり明確ではない。

3.2. 時間軸の安定度と特定領域 時間軸伸縮マッチングの結果、1フレームとの対応づけが2フレーム以下である場合、すなわち(2,m)の点を通るDPパズが図13に示すような通りか、あるいはかみ込み方をしているとき、(2,m)の点では伸縮がはたっていないものとし、前記でみた全DPパズについて、

$$(2, l, m) \quad (2, l, m+1)$$

標準パターンと各フレーム毎に、時間軸の伸縮の比率を割合を調べた。この割合を安定度と呼ぶ。(12)

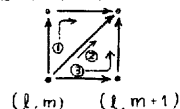


図13. (2,m)において伸縮なしに見える条件

その結果は図14に示す通りで、標準パターンと同一話し者の音声とマッチングを行ったときに

は、どの話者にも3~4か所の安定度が大きい区間のあることがわかる。標準パターンと異なる話者の音声とのマッチングの場合は、安定度はかなり小さくなる。そこで各話者について、安定度が50%を越す領域と取り出し、この領域をその標準パターンの特定領域と呼ぶことにする。

3.3. 音声スペクトルと特定領域との対応 上で定義した特定領域が、実際の音声のどのような区間に対応しているかを調べた。図12と同じ話者の /namae/ のスペクトル包絡とパワーの時間パターンと、特定領域に対応する区間を図15に示す。スペクトル包絡は、12次の最速スペクトル推定法によるものを示した。この結果からわかるように、特定領域はほぼ音韻の過渡区間に対応しており、この性質はすべての話者に共通している。

3.4. 時間軸の安定度による話者認識 2.5で述べた長期間データによる話者認識実験と同じ学習サンプルと未知音声の組合せによる話者識別実験により、時間軸の安定度の有効性を調べた。あらかじめ各話者の標準パターンにおける特定領域を調べておき、未知音声と各話者の標準パ

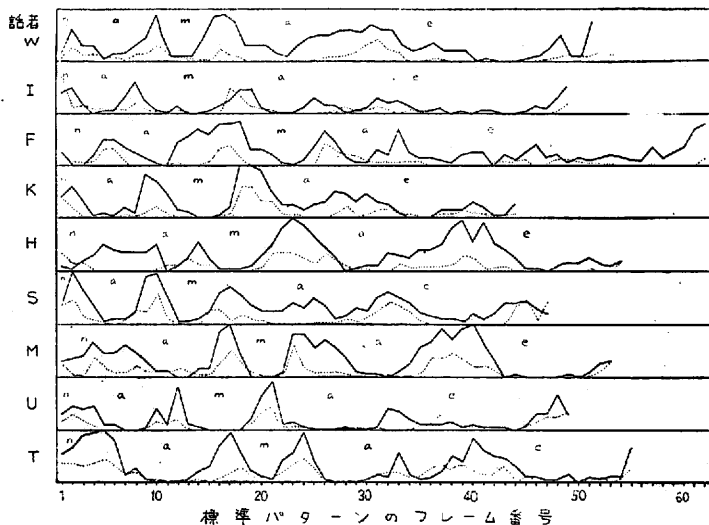


図14. 標準パターンの各フレームの安定度 (/namae/) — 同一話者内, 異話者間

ターンとのマッチングを行ったときに、標準パターンの全特定領域で伸縮のおこらない割合 (安定度) を、その標準パターンの音声との類似度として用いた。

この安定度を単独に話者の特徴として用いるときに、正しい話者が選ばれる割合を、図16に示す。この図には、安定度の値が大きい話者を順に選ぶときに、正しい話者が選ばれる累積確率が示してある。2種類の単語のそれぞれを用いる場合と、両者の安定度を平均した値を用いる場合 (2単語組合せ) の結果を示した。

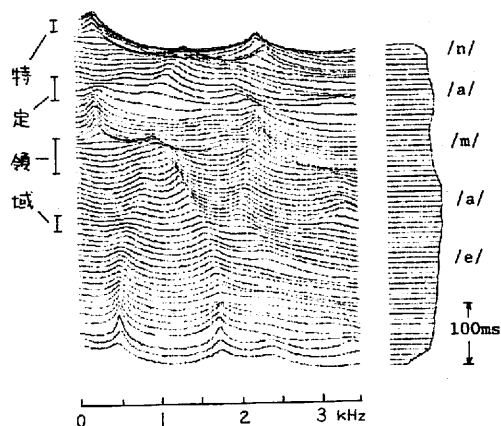


図15. 音声スペクトルと特定領域との対応 (話者K)

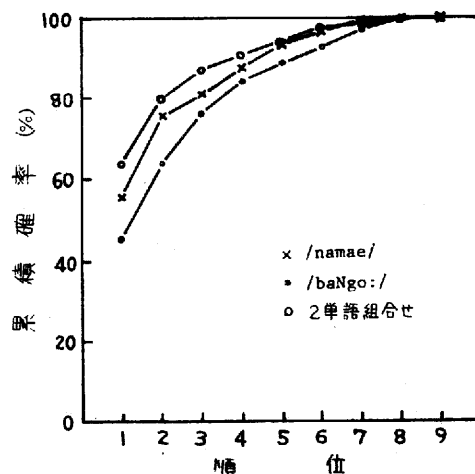


図16. 安定度による識別で正しい話者の選ばれる割合

正しい話者が第1位に選ばれる確率は、1単語の場合、平均して50%程度であり、第3位まで選べば、ほぼ80%の確率で正しい話者が抽出される。2単語を組合せの場合、これらの確率は10%程度向上する。この結果から、特定領域における安定度は、個人性情報を表現するパラメータとして単独に用いるには不十分であるが、他の特徴パラメータと組合せたり、補助的に用いれば、有効であると考えられる。

3.5. 時間パターンと安定度による話者認識 基本図数とLARの時間パターンのマッチングを基本とし、未知音声と各話者の標準パターンとのDPマッチング後の距離に関して、第1位と第2位の差がある閾値よりも小さいときには、前述の特定領域における安定度の大きい方を選ぶという方式で、話者識別実験を行った。時間パターンとのマッチング結果に統計的パラメータの一部である平均値ベクトルの距離を付加したもの(2.7参照)に、さらに安定度を組合せる場合についても実験を行った。学習サンプルと未知音声の組合せは、3.4と同様である。安定度を用いない場合の結果(図4(四))とあわせて、図17に結果を示す。

時間パターンに安定度を組合せることにより、1単語による判定の場合には、誤り率が組合せる前のほぼ3/5に低下し、2単語を組合せる場合は4/5に低下する。このことは安定度の有効性を示しており、このときの誤り率は、単語の統計的パラメータのすべてを用いる場合の誤り率にはほぼ等しい。時間パターンに平均値ベクトルを組合せて用いる場合は、さらに誤り率がかなり小さくなっており、安定度を付加する効果は見られない。

3.6. 短期間データによる話者認識 図1の短期間データを用いて、話者識別実験を行った。学習サンプルと未知音声の時期差と差を考慮し、安定度を単独に用いて、正しい話者の選ばれる割合を図18に示す。2種の単語の平均結果を示した。時期差が2~3日のときは、かなり大きな確率で本人が抽出され、図16の結果よりも大きな値を示しているが、時期差の増加とともに、値が低下する傾向がみられる。このことは安定度にも、図6~図7に示した音声スペクトル

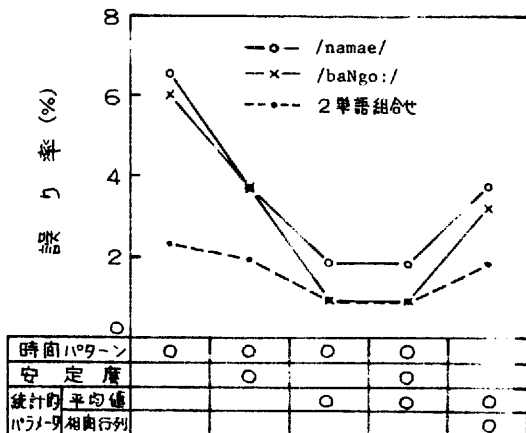


図17. 長期間データによる話者識別における安定度の効果

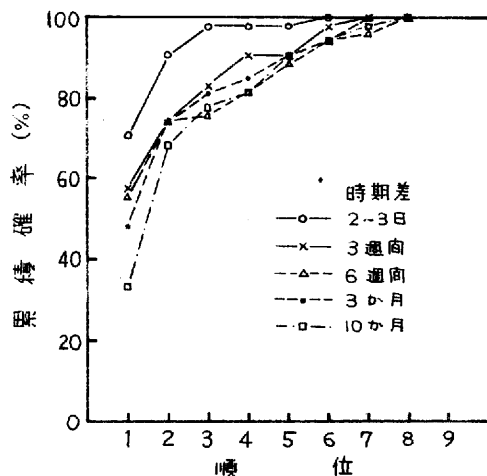


図18. 安定度による識別で正しい話者の選ばれる割合(学習サンプルが短期間の場合)

ルそのものの性質と同様の、短期的変動特性が存在することを示している。

次に、同じく短期間データを用いて、時間パターンと安定度による話者識別実験を行った。図19に、安定度を組合せない場合と統計的パラメータによる場合の結果をあわせて、学習サンプルと未知音声の時期差と誤り率の関係を示す。時間パターンに安定度を組合せることにより、時期差による誤り率の増加の度合いを、やや低下させることができる。

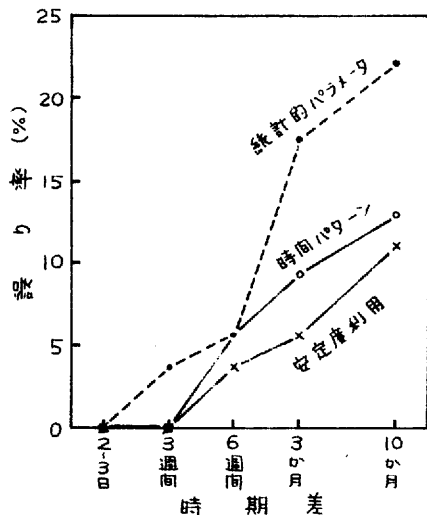


図19. 短期間データによる話者識別における安定度の効果

IV むすび

音声スペクトルの時間変動特性を用いた話者認識に関して種々の検討を行った。その結果、次のことが明らかになった。

(1) 音声スペクトルを表現する特徴パラメータとして、対数断面積比 (LAR) と基本周波数の時間パターンを用い、この特徴ベクトルの長期間にわたる時期的変動特性を考慮した重みつきDPマッチングを行うことにより、高い話者認識率を得ることができる。この認識率は、同数般的に平滑化したスペクトル包絡そのものの時間パターンを用いる場合よりも大きい。

(2) 時間軸の対応づけを行う2次元平面上の整合窓の窓幅は150~200msが最適で、端点は固定した方がよい。

(3) 時間パターンの類似度 (DPマッチングの結果) に、特徴ベクトルの単語全体における平均値 (従来の統計的パラメータの一部) の類似度を付加すると、認識率が大きく改善される。10か月程度の長期間の学習サンプルを用いて、3か月後の未知音声の話者識別および話者照合を行うと、1単語による判定の場合に約98%、2単語による判定の場合に約99%の認識率を得られる。

(4) 時間パターンのマッチングによる認識の場合にも、統計的パラメータによる場合と同様に、3か月以上の長期間の時期差をわたった未知音声

の話者と高い精度で認識するためには、10か月程度の長期間にわたる学習サンプルが必要である。学習サンプルが10日間程度の短期間のものであると、未知音声との時期差が3週間程度より大きくなると、その増加とともに認識率が急激に低下し、時期差が3か月以上になると、誤り率がかつた大きくなる。時期差による認識率の低下の度合いは、統計的パラメータの場合よりも時間パターンの場合の方がやや小さい。

(5) 同じ話者が同じ単語を発声しても、そのソフトウェアの時間構造は、各部分が非線形に伸縮するが、各話者に固有であり伸縮をしない区間もある。この安定な区間は、16音韻の週波区間に対応し、この区間を特定領域と定義する。

(6) 未知音声と各話者の標準パターンの、時間軸を対応付けたときに、各標準パターン内の特定領域で伸縮が起きている割合を安定度と名づけると、この値は未知音声と標準パターンの類似度の尺度の一種として用いることができる。この値も、時間パターンの類似度に加的に組合せて話者識別を行うと、時間パターンの類似度のみに用いる場合に比べて、誤り率を3/5に低下させることができる (1単語による認識の場合)。

今後さらに単語の種類と拡大して実験を行うとともに、音声スペクトルの動的特徴に含まれる個人性情報を抽出するさらに有効な方法の検討と進めてゆく予定である。

〔謝辞〕日頃熱心に討論して頂く研究室の諸氏に感謝する。

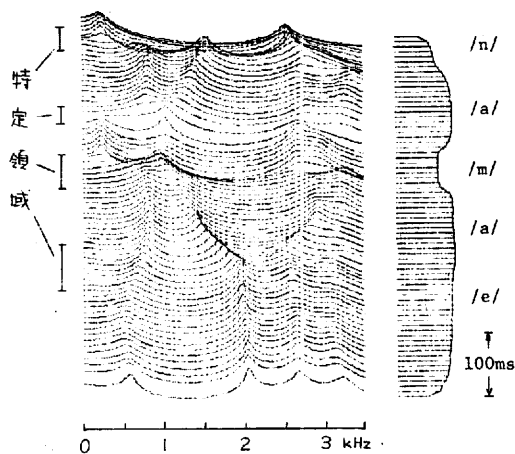
文

献

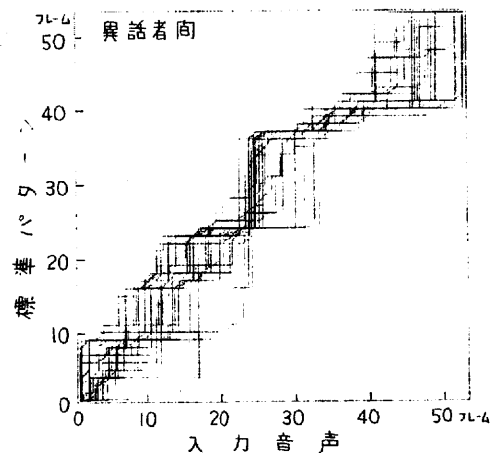
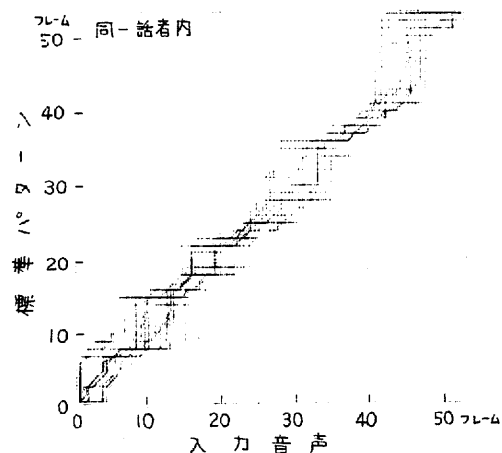
- (1) 古井・板倉・斎藤：長時間平均スペクトルによる話者認識，信学論，55-A，No. 10，p. 549，1972
- (2) 古井・板倉：単語の統計的パラメータによる話者認識，信学論，56-A，No. 11，p. 717，1973
- (3) 古井：音声の個人性パラメータの時期的変動と話者認識，信学論，57-A，No. 12，p. 880，1974
- (4) G.R. Doddington: Speaker verification, RADCR-TR-UI-963700-F, 1974

- (5) R.C.Lummis: Speaker verification by computer using speech intensity for temporal registration, IEEE Trans. AU-21, No.2, p.80, 1973
- (6) 迫江, 鵬田 4葉: 自己相関パラメータによる話者認識の検討, 音学講論(秋季), 2-2-22, 1972
- (7) 徳永, 平川, 杉原: 部分自己相関係数を用いた数字音声話者の認識実験, 音学講論(秋季), 1-3-24, 1973
- (8) 市川, 中島: 話者確認用パラメータの比較, 昭52信号情報全大, 227, 1977
- (9) 迫江, 4葉: 動的計画法を利用した音声の時間正規化に基づく連続単語認識, 音学誌, 27, No.9, p.483, 1971
- (10) 古井: 単語音声認識における時間正規化方法の検討, 信学研究所学習研究 PRL76-25 1976
- (11) F.Itakura: Minimum prediction residual principle applied to speech recognition, IEEE Trans. ASSP-23, No.1, p.67, 1975
- (12) 古井, 斎藤: 音声スペクトルの動的特徴を用いた話者認識, 音学講論(春季), 4-1-25, 1978

(付録) 標準パターンと話者がMの場合の, 時間軸の対応づけの例と, 特定領域.



付図2. 音声スペクトルと特定領域との対応 (話者M)



付図1. 時間軸の対応づけの結果の一例 (/namae/, 標準パターン: 話者M)