

論文 / 著書情報
Article / Book Information

論題(和文)	音声スペクトルの動的特徴を用いた話者認識
Title(English)	
著者(和文)	古井 貞熙, 斉藤 収三
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会1978年講演論文集, , , pp. 617-618
Citation(English)	, , , pp. 617-618
発行日 / Pub. date	1978, 5

4—1—25 音声スペクトルの動的特徴を用いた話者認識*

古井 貞照 斎藤 収三 (日本電信電話公社 武蔵野電気通信研究所)

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とともに変化しており、スペクトルの静的特徴とともにその動的特徴にも個人差があると考えられる。スペクトル包絡を対数断面積比(LAR)で表現し、基本周波数とLARの時間変化における相関特性などを用いて話者認識を行う試みについては、すでに報告したが⁽¹⁾、統計的特徴にとどまっており、動的特徴を積極的に話者認識に用いた試みは、まだほとんどない。

II 実験方法 同じ話者が同じ単語を繰返し発声しても、そのスペクトルの時間構造は様々に変化するが、各話者に特有であり変動しない時間構造も存在することが考えられる。この様子を、男性9名が3か月おきの8時期(各時期に3回)に発声した /namae/ と /baŋgo:/ の2種の単語音声を用いて、同一話者の異なる時期のサンプル間と、異なる話者のサンプル間の時間軸非線形伸縮マッチングの対応づけの結果によって調べてみた。各サンプルを3.4 kHzの低域フィルタに通したのち、10ms毎に32msのフレームをハミング窓によって切り出し、各時刻にもおける12次までのLARと基本周波数を求め、組合せて特徴ベクトル $\mathbf{x}_t$ とする。時間軸の対応づけは、 $\mathbf{x}_t$ に関する重みつき距離を用いたDPマッチングによって行う。重みには、 $\mathbf{x}_t$ の各要素ごとに、各単語の音声区間における平均値の8時期における分散の逆数を用いた。DPマッチングは、2種の音声サンプルが作る平面の、始点と終点を結ぶ対角線の両側に N_w フレームずつの窓を設けて、その中で行い、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。8時期の音声のうち、初めの3時期から1サンプルずつ学習サンプルをとり出し、時間軸伸縮を行いつつ3サンプルの平均化を行って、各話者の標準パターン時系列を作る。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った。

III 動的特徴の抽出 時間軸の対応づけの結果の一例を、図1に示す。 $N_w=20$ フレームとした。

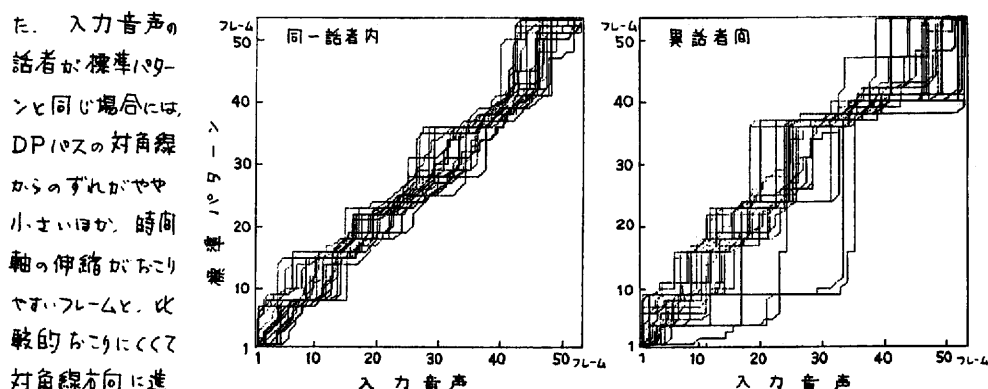


図1. 時間軸の対応づけの結果の一例 (/namae/, 標準パターン: 話者M)

フレームがあることがわかる。そこで、1フレームとの対応づけが2フレーム以下である場合には、伸縮がなっていないものとし、各話者の標準パターンの各フレーム毎に、伸縮のない割合(安定度)を調べた。/namae/の場合の結果を図2に示す。標準パターンと同じ話者の音声とマッチ

* Talker Recognition by Dynamical Features of Speech Spectra.
By Sadaaki Furui and Shuzo Saito (Musashino ECL, NTT)

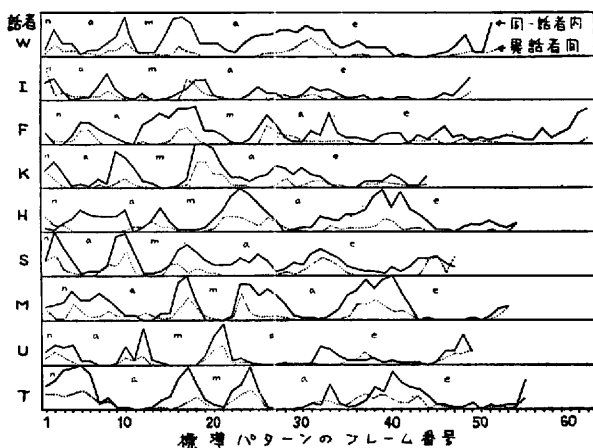


図2. 標準パターンの各フレームの安定度 (/namae/)

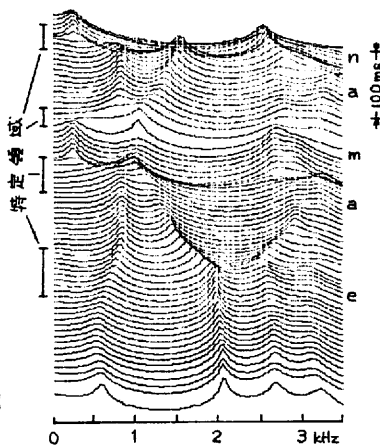


図3. 特定領域とスペクトルとの対応 (話者M)

ングを行ったときには、どの話者にも、3~4カ所の安定度の大きい区間のあることがわかる。異なる話者の音声とのマッチングの場合は、安定度がかなり小さくなる。各話者について、安定度が50%を越す領域をとり出し、この領域を特定領域と呼ぶことにする。この領域はどの話者についても、図3に示す例のように、ほぼ音韻の過渡区間に対応している。

Ⅳ 話者認識への適用 未知サンプルと各話者の標準パターンのマッチングを行ったときに、標準パターンの特定領域で伸縮のおこっていないフレームの割合 (安定度) を特徴として用いた話者識別実験を行った。9名の登録話者の学習サンプルとして、3ヵ月おきの4時期の音声を用い、3ヵ月後の音声の識別を行った。未知サンプルとして、9名が8時期に3回ずつ発声した216サンプルを用いた。Nw=10とした。安定度を単独に用いるときに、正しい話者が選ばれる割合を図4に示す。各単語の場合と、2単語の平均値を用いる場合の結果を示した。単語によって差があるが、この特徴の有効性がわかる。次に、DPマッチングの結果としての時系列としての距離に、安定度を組合せて話者識別実験を行った。DPマッチング距離に関して、 ϵ_1 位と ϵ_2 位以下の差がある閾値より小さい場合に、その閾値以内に入る話者のうち安定度の最も大きい話者を選ぶという方法を採用した。結果は図5に示す通りで、いずれの場合も安定度を用いる効果が見られ、2種類の単語を組合せ、安定度を補助的に用いれば、約98%の認識率を得ることができる。

V むすび 単語

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方法を検討し、話者識別実験によってその有効性を確認した。

[断片] 日頃無心に討論して頂く研究家の諸氏に感謝す。

[文献] 小古 邦彦: 信学論, 56-A, 11, p.717 (昭48)

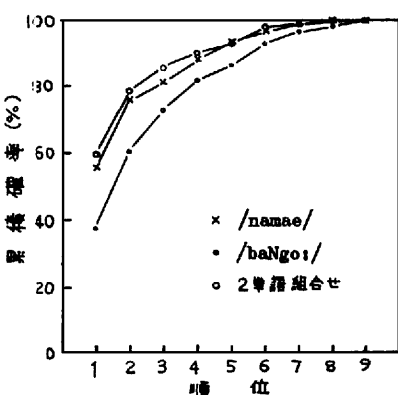


図4. 安定度による識別で正しい話者の選ばれる割合。

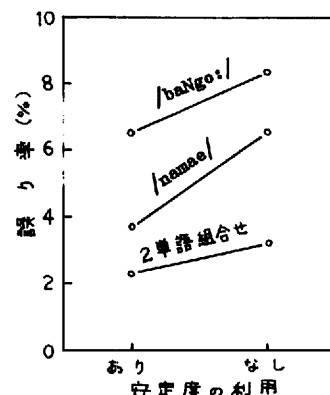


図5. 安定度を補助的に用いる効果。