

論文 / 著書情報  
Article / Book Information

|                   |                                     |
|-------------------|-------------------------------------|
| 論題(和文)            | 単語音声スペクトルの統計的特徴による話者認識の改良           |
| Title(English)    |                                     |
| 著者(和文)            | 古井貞熙                                |
| Authors(English)  | SADAOKI FURUI                       |
| 出典(和文)            | 日本音響学会1980年秋季講演論文集, , , pp. 397-398 |
| Citation(English) | , , , pp. 397-398                   |
| 発行日 / Pub. date   | 1980, 10                            |

古井 貞熙 (日本電信電話公社 武蔵野電気通信研究所)

**I まえがき** 音声波に含まれる個人性情報を用いて、その音声の話者を自動的に判定する話者認識の方法は、音声スペクトルの時間パターンを直接的に用いるものと、その時間パターンから抽出した統計的特徴を用いるものに分けることができる。一概にどちらが良いとは言えないが、統計的特徴を用いる方が、認識に必要な計算量が比較的少ないという利点がある。<sup>(1)(2)</sup>

ここでは、複数単語から抽出した統計的特徴を用いる新しい話者認識方法の有効性を確認した結果について述べる。

**II 特徴抽出と認識方法** 図1に、認識システムのブロック図を示す。単語音声は、音声の始端と終端を検出したのち、10ms毎に30msの区間を切り出して、各区間毎に有聲/無聲の判定を行い、有聲音区間のみの相関係数の時系列に変換する。全有聲音区間の平均相関係数を用いて一次系の逆フィルタを構成し、スペクトル等化処理を行う。このうち、基本周波数 $f_{0t}$ と10次までのケプストラム $\{C_{it}\}_{i=1}^{10}$ からなる11次元ベクトル $X_t$ の時系列に変換する。 $X_t = (C_{1t}, C_{2t}, \dots, C_{10t}, f_{0t})$ 。 <sup>(1)</sup>

この時系列から、各単語の全有聲音区間における平均値ベクトル $\mu$ 、標準偏差ベクトル $\sigma$ 、相互相関行列 $A$ 、およびフーリエ(コサイン)展開係数行列 $F$ を計算する。

これらの統計量のうち、 $\sigma$ 、 $A$ 、および $F$ ( $X_t$ の各要素ごとに1~6次の展開係数を計算する)の要素から、話者の分離に有効な60成分を選択し、 $\mu$ の全要素と組合せて、71次元の特徴ベクトルとする。の判定を行った(登録話者9名)。

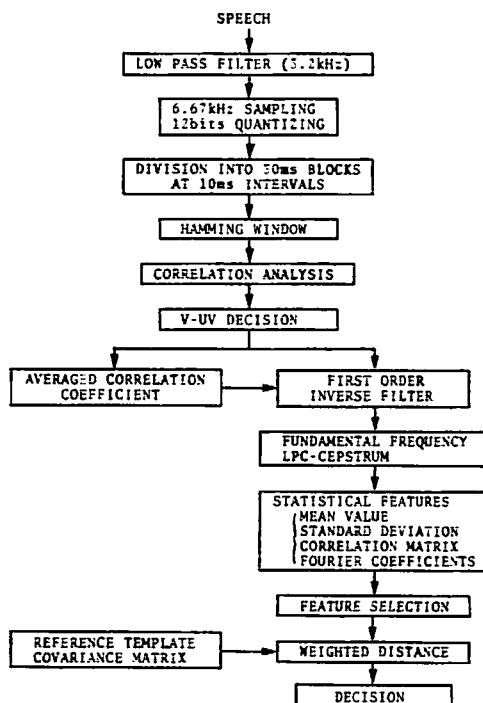


Fig.1. Block diagram indicating the principal operation of the speaker recognition system.

選択する成分は、多数話者の学習サンプルを用いた分散分析によって決定する。話者認識の判定は、入力音声の特徴ベクトルと各話者の標準パターンの、各成分の安定性を考慮した重みつき距離によって行う。

従来の方法<sup>(3)</sup>と異なる点は、スペクトルパラメータをLAR(対数断面積比)からケプストラムに変えた点と、特徴量にフーリエ展開係数を加えた点にある。

**III 認識実験の方法** 実験には、男性9名が3年間の長期間にわたって録音した/baŋgo:/, /namae/, /ko:geŋ/, /bakuon/の日本語4単語の音声サンプルを用い、4単語の特徴を組合せて話者識別と話者照合の判定を行った(登録話者9名)。

\* Improvement of a speaker recognition method using statistical features of speech spectrum extracted from word utterances. By Sadaaki FURUI (Musashino ECL, NTT)

**実験1** マイクロホンで収録した音声を用い、各話者毎に、3か月おきの4時期に収録した12サンプルを学習サンプルとして、その3か月後の音声と認識する実験を、多時期の音声サンプルを用いて行った。全登録話者の入力音声の総数は216サンプルである。話者照合の詐称者の音声としては、103名の男性話者が1回ずつ発声したものをを用いた。

**実験2** カーボン送話器を含む電話機及び擬似加入者線を通した音声と、同時に録音したマイクロホン音声を用いて実験を行った。学習は6か月おきの3時期(9サンプル)、入力音声との時期差は6か月とし、詐称者の音声として他の11名が3回ずつ発声したものをを用いた。全登録話者の入力音声の総数は54サンプルである。

**IV. 実験結果** 実験1で用いた9名の初めの4時期の音声によって各特徴の有効性の評価を行った結果を、F比の値によって図2に示した。Fの要素と面およびAの要素に分け、前者は屈折次数毎、後者は全要素の平均値を示した。比較のために、スペクトルパラメータとしてLARを用いる場合の結果をあわせて示した。この結果から、LARよりもケプストラムの方が有効性が高く、Fの要素の有効性は次数とともに低下することがわかる。図3に、71次元の特徴ベクトルを構成する各要素の割合を示す。図2と図3は共に、Fの要素が従来からの特徴面およびAの要素と同程度の有効性を有することを示している。

認識実験の結果を表1(実験1)と表2(実験2)に示す。表1の結果から、ケプストラムの統計的特徴による方法が極めて有効性が高く、極めて低い認識誤り率が得られることがわかる。実験2は実験1よりも実験条件が厳しいので、誤り率がやや大きくなっているが、本方法が、電話音声に対してもマイクロホン音声とほぼ同等の性能で動作することがわかる。

**V. おわりに** ケプストラムの統計的特徴を用いる新しい話者認識方法の有効性を確認した。

(謝辞) 日原御指事氏、当所野田基礎研究部長ならびに小池和典氏、有益な討論として当所研究室の諸氏に感謝する。

[文献] (1) 古井: 音声学序論, 1-7-2 (1980), (2) 古井: 音声学序論, 580-23 (1980), (3) 古井-板倉: 信学論, 56-A, p.717 (1977)

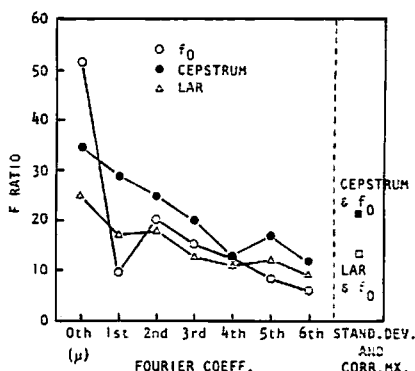


Fig.2. Effectiveness of each feature parameter set.

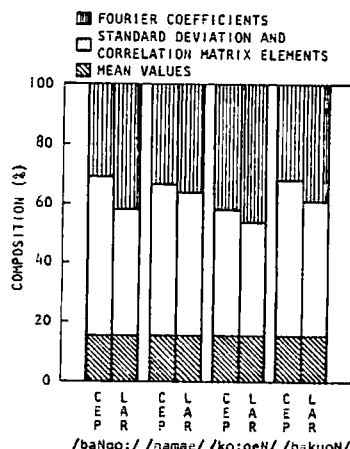


Fig.3. Composition of feature vectors for speaker recognition.

Table 1. Recognition error rates for experiment 1.

|          | IDENTIFICATION |                          | 0 %    |
|----------|----------------|--------------------------|--------|
|          |                |                          |        |
| LPC-CEP. | VERIFICATION   | A PRIORI FIXED THRESHOLD | 0.10 % |
|          |                | A POSTERIORI THRESHOLD   | 0.02 % |
| LAR      | IDENTIFICATION |                          | 0.46 % |
|          | VERIFICATION   | A PRIORI FIXED THRESHOLD | 1.21 % |
|          |                | A POSTERIORI THRESHOLD   | 0.51 % |

Table 2. Recognition error rates for experiment 2.

|      | IDENTIFICATION |                          | 0 %    |
|------|----------------|--------------------------|--------|
|      |                |                          |        |
| TEL. | VERIFICATION   | A PRIORI FIXED THRESHOLD | 2.44 % |
|      |                | A POSTERIORI THRESHOLD   | 0.49 % |
| MIC. | IDENTIFICATION |                          | 0 %    |
|      | VERIFICATION   | A PRIORI FIXED THRESHOLD | 2.44 % |
|      |                | A POSTERIORI THRESHOLD   | 1.36 % |