

論文 / 著書情報  
Article / Book Information

|                   |                                     |
|-------------------|-------------------------------------|
| 論題(和文)            | クラスタ化された短時間スペクトルを単位とする単語音声認識        |
| Title(English)    |                                     |
| 著者(和文)            | 菅村 昇, 古井 貞熙, 箱田 和雄                  |
| Authors(English)  | SADAOKI FURUI                       |
| 出典(和文)            | 日本音響学会1980年秋季講演論文集, , , pp. 391-392 |
| Citation(English) | , , , pp. 391-392                   |
| 発行日 / Pub. date   | 1980, 10                            |

# 1—1—17 クラスタ化された短時間スペクトルを単位とする 単語音声認識\*

○管村 昇 古井 貞熙 箱田 和雄  
(日本電信電話公社 武蔵野電気通信研究所)

## 1. まえがき

単語音声の認識方法として、これまで単語全体を一つのパターンとみなして単語単位の標準パターンを用いて認識する方法<sup>(1)</sup>と、音韻を単位とする標準パターンを作成し、このパターンを用いて単語辞書を使って認識する方法<sup>(2)</sup>などがある。認識対象の単語数が増加するにしたがって、前者は標準パターンを蓄えておく記憶容量が膨大なものとなり、マッチングの演算量も増加する。後者は、音韻標準パターンや単語辞書の追加にやや困難な点がある。

本稿では、両方式の中間的なものとして、言語レベルにとらわれない方法で、擬似的な音韻標準パターンを作成し、このパターンを用いて単語認識実験を試みに結果について報告する。

## 2. 単語認識アルゴリズム

### 2-1. 標準パターンの作成方法

標準パターンの作成方法としては、あらかじめ発声された単語音声をもとに、クラスタリングの手法を用いる<sup>(3)</sup>。この方法は、音声の物理的な特徴のみに着目する考えであり、言語体系に依存しない特長を有する。

### 2-2. 認識方法

認識方法のブロック図を図1に示し、以下その方法について説明する。

(1) 学習サンプルによって、標準パターン  $S_1, S_2, \dots, S_n$  をつくる。(図1-(b))

(2) 各単語の学習サンプルと標準パターン

$S_1 \sim S_n$  のマッチングを行い、各単語を  $S_1 \sim S_n$  の系列で表現する単語辞書を作成する。(図1-(d))

(3) 未知サンプルを分析し(図1-(a))各フレームごとに  $S_1 \sim S_n$  の標準パターンとの距離を計算して距離マトリクスをつくる。(図1-(c))

(4) 未知サンプルと標準パターンとのマッチングをDPによって行い(図1-(e))結果を出力する。(図1-(f))

本方法によれば、DPマッチングの距離としては、(2)で得られに各単語の標

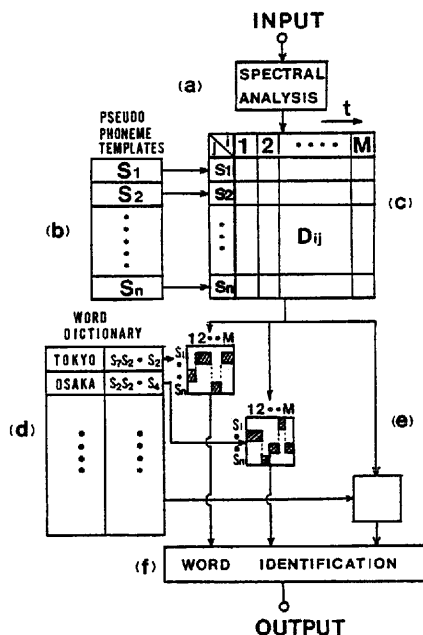


Fig.1 Block Diagram of Word Recognition System using Pseudo Phoneme Templates

\*Word Recognition Using Pseudo Phoneme Templates and the Dynamic Programming Technique  
by Noboru SUGAMURA, Sadaaki FURUI and Kazuo HAKODA (Musashino E.C.L., N.T.T.)

準)パタン系列を参照して、(3)で計算した距離マトリクスの要素を検索して用いるので、計算量を大幅に減少させることができる。たとえば認識対象とする単語数を $L$ 、未知サンプルのフレーム数を $M$ 、DPの窓幅を $N_w$ 、標準パタン数を $n$ とすると、単語全体を一つのパタンとして認識する方法では、1単語認識するのに必要な距離計算回数は、約 $M \times N_w \times L$ であるのに対し、本方法では、 $M \times n$ となる。したがって1フレームあたりの計算量の比は、 $R = n / (N_w \times L)$ となる。このことから、単語数が多くなればなるほど、本方法が有効であることがわかる。

### 3. 認識実験結果

表1に示すような実験対象、分析条件、DP手法で、認識実験を行った結果を図2に示す。図2には、本方法と単語全体を)パタンとした場合の認識率(a)を示している。実験で用いた標準パタン数は、128, 96, 64の3種類であり、これらのパタンを作成するために、2200~2600フレームのデータを用いている。この結果、標準パタン数が64になると認識率は急に低下するが、128では、全フレームのパタンを用いた場合とほぼ同等の結果が得られた。(誤認識数の差は、全話者で1サンプル)

つぎに距離の値にある閾値を設けて、正しい単語を棄却する誤りと、異なる単語を受理する誤りが等しくなるときの誤り率<sup>(4)</sup>によって認識機能の評価を行ったところ、図2(b)に示す結果が得られた。この結果、誤り率は、標準パタン数を96にしても、全フレームのパタンを用いた場合と同じであることがわかった。

以上の結果から、本方法で標準パタン数を128程度(全フレームの1/20~1/25)にしても認識精度は、ほとんど低下しないことが明らかとなった。

### 4. あとがき

簡単なクラスタリング手法を用いて作成した擬似的な音韻標準パタンで単語認識を行った結果について報告した。今後は本手法の有効な適用領域であると考えられる特定話者の大語い単語音声認識に適用することを試みる。また話者に依存しない不偏的な標準パタン構成法についても検討をすすめていく。

### 謝 辞

日頃御指導頂く、野田基礎研究部長、小池第四研究室長に感謝します

### 参考文献

- (1) 迫江, 千葉; 音言誌 27, 9, P.483 (1971)
- (2) 好田, 橋本, 斎藤; 信言論 55-D 3 P.186 (1972)
- (3) 菅村, 板倉; 音声研究会資料 879-08 (1979)
- (4) 古井, 板倉, 斎藤; 信言論 55-A 10 P.549 (1972)
- (5) S.Furui and A.E.Rosenberg; ICASSP '80 P.1060

Table.1 Experimental Condition

|         |                                |
|---------|--------------------------------|
| 実験対象単語  | 空巻名 67 単語                      |
| 発声者     | 男性 5 名 (各単語 学習用, テスト用 各 1 回発声) |
| 分析条件    | 8kHz サプリング<br>フレーム周期 15msec    |
| 特徴パラメータ | LPC ケプストラム                     |
| DP 手法   | 両端点フリー DP <sup>(5)</sup>       |

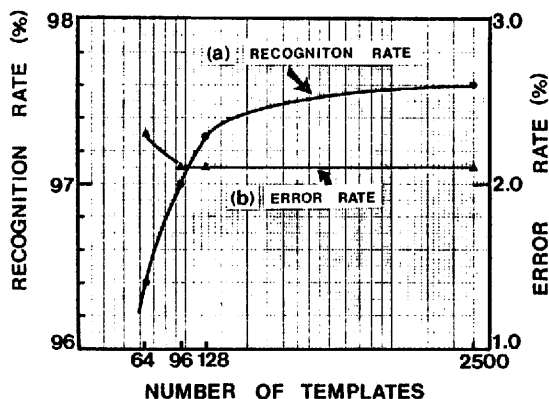


Fig.2 Recognition and Error Rate Using Pseudo Phoneme Templates