

論文 / 著書情報
Article / Book Information

論題(和文)	擬音韻標準パターンによる大語い単語音声認識
Title(English)	
著者(和文)	菅村 昇, 箱田 和雄, 古井貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	昭和56年度電子通信学会総合全国大会, , , pp. 5-332
Citation(English)	, , , pp. 5-332
発行日 / Pub. date	1981, 8
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) 1981 Institute of Electronics, Information and Communication Engineers.

擬音韻標準パターンによる 大語い単語音声認識

菅村 昇

稲田 和雄

古井 貞照

(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、クラスターリング手法により、短時間スペクトルの標準パターン(擬音韻標準パターン)を作成し、この標準パターンを用いて単語認識を行う方法については、すでに報告した⁽¹⁾。本稿では、この方法を大語いの単語音声認識に適用した結果について報告する。

2. 認識方法

認識方法のブロック図を図1に示し、以下その手順について説明する。

(1) あらかじめ発声された単語音声をもとに、クラスターリング手法を用いて、擬音韻標準パターン $S_1, S_2, \dots, S_n$ をつくる。⁽²⁾ (図1-(b))

(2) 認識対象単語の学習サンプルと標準パターンとのマッチングを行い、各単語を $S_1 \sim S_n$ で表現した単語辞書を作成する。(図1-(d))

(3) 未知単語音声をもとに(図1-(a)), 各フレームごとに $S_1 \sim S_n$ との距離を計算して距離行列をつくる。(C)

(4) 距離行列と単語辞書とを用いて、DPマッチングを行い(図1-(e)), 最小の距離を有する単語を出力する。

認識方法の特長 本方法は、単語全体を一つのパターンとして認識する方法と比べて、つぎのような特長をもつ。
(i) 単語辞書は、パラメータそのものではなく標準パターンの符号系列だけを蓄積しておくので、メモリが少なく済む。

(ii) 認識に要する距離計算の回数が、認識対象単語数に依存せず、少ない。たとえば、単語数を L 、未知サンプルのフレーム数を M 、DPの幅を N_w 、標準パターン数を n とすると、本方法と単語全体を用いる場合との1フレームあたりの計算量の比は、約 $R = n / (N_w \times L)$ となる。

以上のことから、単語数が多くなればなるほど、本認識方法が有効であることがわかる。

3. 実験結果

男性話者1名が発声した641都市名を認識対象単語とし、認識実験を行った。擬音韻標準パターンの作成には、1回目発声の641都市名のうち、無作為に抽出した64単語(フレーム総数2941)を用いて行った。フレーム周期は16ms、距離計算には、16次のケプストラムまたはWLRを用いた。標準パターンを128と256の2種類用意して、2回目発声の641都市名の認識実験を行ったところ、表1に示すような結果が得られた。表中には、単語全体の情報を用いて実験を行った結果についても示している。この結果、16次のケプストラムを用いた場合、標準パターン数128と単語全体の情報を用いた場合との認識率の差は0.6%であり、256個の標準パターンを用いた場合は、認識率はほとんど差がないことが明らかになった。

4. あとがき

本稿で述べた認識方法では、標準パターンの作成には、何ら言語的な情報を必要とせず、人手を介することもない。このような簡単

な方法で作成した比較的少数の標準パターンを用いて、大語いの認識を行っても認識率の低下は、わずかであり、本認識方法が、大語い単語音声認識に利用できる見通しが得られた。

謝 辞 日頃御指導頂く、野田基礎研究室長、小池第四研究室長に感謝します。また、認識実験に際し、御協力頂いた第四研究室、鹿野綱重、杉山社員に感謝します。

参考文献

- (1) 菅村、古井、稲田 音響学会論集 2-1-17 (1980-10)
- (2) 菅村、稲田 音声研究会資料 S79-08 (1979)
- (3) 鹿野 音響学会論集 1-1-10 (1980-10)
- (4) 杉山、鹿野 音声研究会資料 S80-13 (1980-5)
- (5) 鹿野、杉山 本論文集

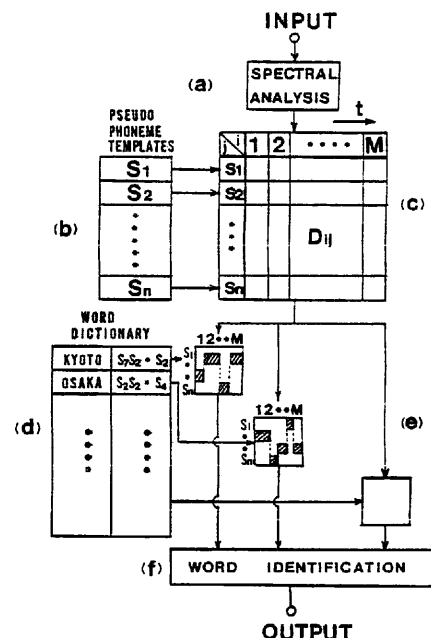


Fig.1 Block diagram of word recognition system using pseudo phoneme templates

Table.1 Recognition experiment results

Standard Patterns		Pseudo Phoneme Templates & Word Dictionary (Pattern Sequence)		Word Dictionary (Parameter)	
Number of Standard Patterns		128	256	All Frames	
Memory for Templates (Ratio)		0.03	0.04	1.00	
Distance Measure		CEP	WLR	CEP	WLR
		94.4	95.0	95.3	95.0
		97.4	97.5	97.5	97.7
		98.3	98.3	98.4	98.4
Recognition Rate (%)		94.4	95.0	95.3	95.0
Error Rate (%)		1.51	1.65	1.33	1.13