

論文 / 著書情報
Article / Book Information

論題(和文)	予備選択による大語い単語音声認識
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1981年春季講演論文集, , , pp. 551-552
Citation(English)	, , , pp. 551-552
発行日 / Pub. date	1981, 5

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

1. まえがき

これまでの単語音声認識の研究においては、認識対象語い数は、100語程度であることが多かった。対象語い数が数百から数千語の大語い単語認識においては、マッチングの効率化を図るために、マッチングすべき単語の予備選択を行うことが必要となる。本稿では、スペクトルパラメータの大局的な時間変動特性を利用した単語予備選択法について報告する。

2. 単語の予備選択法

単語の予備選択⁽¹⁾⁽²⁾に関しては、つぎのようなことが要求される。

- (1) 処理が簡単であること。
- (2) 予備選択を付加することによって認識率が著しく低下しないこと。

(3) 処理が話者に依存しないこと。

(4) 予備選択能力が高いこと(予備選択によって2次処理が大幅に減少できること)

このような観点から、本稿ではケプストラムのフーリエ展開係数を用いた予備選択処理について検討した。⁽³⁾641都市名を認識対象語いに選び、男性話者4名のデータをもとに実験を行った。単語認識の方法としては、先に提案したSPLIT法⁽⁴⁾を用いた。予備選択処理を付加した単語認識システムのブロック図を図1に示す。入力された単語音声スペクトル分析し、LPCケプストラムを抽出する。つぎにケプストラムのフーリエコサイン展開係数を求め、標準パターンとしてあらかじめ蓄えられている展開係数との二乗誤差を計算する。この値があらかじめ設定した閾値 θ 以下の単語のみを選択し、DPマッチングによる認識を行う。

3. 実験結果

まず話者1名を実験データにとり、予備選択実験によるマッチング回数の割合、認識率などの関係を調べた。ケプストラムの個数を10、フーリエ展開係

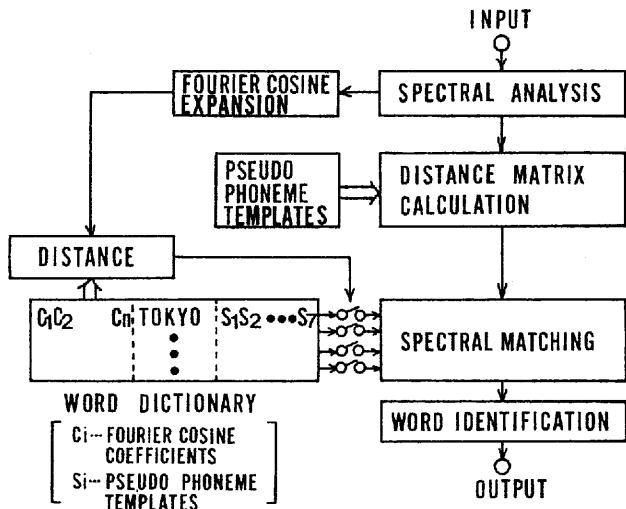


Fig.1 Block diagram of word recognition system

*Large Vocabulary Isolated Word Recognition With Pre-Selection.
by Noboru SUGAMURA, Sadaoki FURUI (Musashino E.C.L., N.T.T.) and
Kazuo HAKODA (Yokosuka E.C.L., N.T.T.)

数を7個に設定して実験を行った。

図2に、 θ と正解が予備選択単語に含まれる割合との関係(曲線A)、 θ とマッチング回数の割合(曲線C)、 θ と認識率との関係(曲線B)を示す。この結果 $\theta=2.0$ 以上では、予備選択によってほとんど認識率が低下しないことが明らかとなった。

つぎに $\theta=2.0$ と $\theta=1.75$ の場合のマッチング回数の割合と認識率との関係を話者4名のデータについて実験を行った。この結果を表1(A)に示す。表1には予備選択を行わないで単語単位の標準パターンを用いて認識を行った結果(B)と予備選択を行わないでSPLIT法を用いて認識を行った結果(C)についても示してある。また $\theta=2.0$ の場合のマッチング回数の分布(4名の平均)を図3に示す。

表1の結果から明らかのように、認識率をほとんど低下させることなく、マッチングの回数を約1/7にすることができる。またマッチング回数を1/10程度にしても、認識率の低下は0.5%(4名平均)程度であった。なお予備選択だけによる処理で、正解が10位以内に含まれる割合は、話者間平均で96.0%であった。

4.あしがき

ケプストラムのフーリエ展開係数を予備選択に利用する方法を提案し、641都市名单語認識において、その有効性を実験的に確認した。

謝 辞

日比谷徹博士、田中基礎研究部長、板倉第四研究室長に感謝します。また御討論して頂く、長野朝彦氏をはじめとする室員の皆様に感謝します。

参考文献

- (1) 中川, 内海, 坂井: 単語音声の大局, 局所両特徴による前読合法と個人差の学習法, 音声研究資料 S78-23
- (2) 藤崎, 成瀬, 保田, 佐藤: 単語音声認識における高度化と語音対応, 音声研究資料 S80-14
- (3) 吉井: SPLIT法による大規模的, 局所特徴の組合せによる単語音声認識, 音声学会論文集 1-18 (1980)
- (4) 吉井, 藤田, 吉井: フラスタによる単語標準パターン生成, 単語音声認識 音声研究資料 S80-61

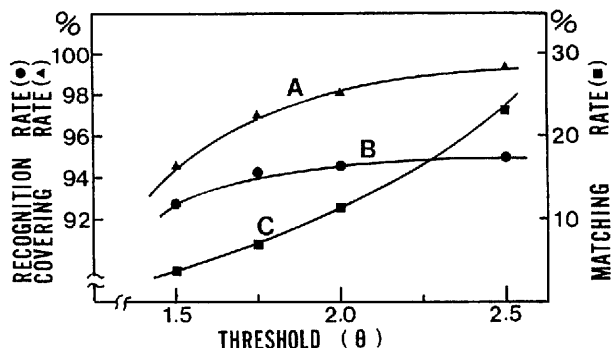


Fig.2 Relation between reduction rate of DP matching and recognition rate

Table.1 Comparison of recognition rate

A--SPLIT+pre-selection
B--FULL matching
C--SPLIT

θ	SPEAKER	MATCHING RATE (%)	RECOGNITION RATE (%)		
			A	B	C
2.0	K S	11.5	94.1	95.0	94.8
	K A	14.0	94.9	94.4	94.9
	N S	13.4	96.5	97.0	97.0
	K I	15.4	94.9	94.6	94.9
	MEAN	13.6	95.1	95.3	95.4
1.75	K S	7.1	93.7	95.0	94.8
	K A	9.0	94.9	94.4	94.9
	N S	8.8	95.7	97.0	97.0
	K I	9.6	94.9	94.6	94.9
	MEAN	8.6	94.8	95.3	95.4

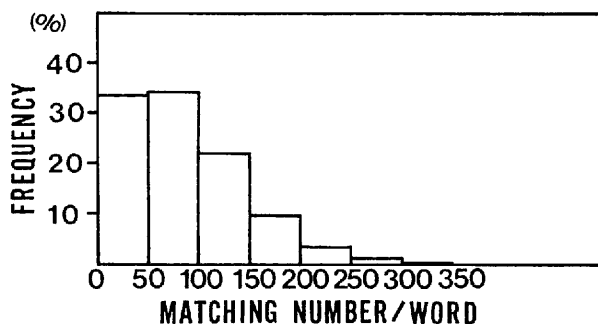


Fig.3 Histogram of matching number in the case of $\theta=2.0$