

論文 / 著書情報
Article / Book Information

論題(和文)	擬音韻と単語辞書のクラスタ化による不特定話者単語音声認識
Title(English)	
著者(和文)	古井 貞熙, 菅村 昇
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会1982年春季講演論文集, , , pp. 29-30
Citation(English)	, , , pp. 29-30
発行日 / Pub. date	1982, 3

○古井 貞熙 菅村 昇 (日本電信電話公社 武蔵野電気通信研究所)

I まえがき クラスタ化によって作成した擬音韻標準パターンと単語辞書を用いる SPLIT (String of Phoneme-Like Templates) 法による特定話者⁽¹⁾及び不特定話者⁽²⁾単語音声認識の原理と有効性についてはすでに報告した。ここでは不特定話者の場合について、その具体的方法を示し、その有効性に関する新たな実験結果について報告する。

II 不特定話者単語音声認識の手順
不特定話者の場合の擬音韻と単語辞書の作成を含む認識のブロック図を図1に示す。図中\$の処理は計算機、他は認識装置⁽³⁾によって行っている。

1) 擬音韻標準パターの作成法⁽⁴⁾ 作成用音声の総フレームを $S_1, S_2, \dots, S_n$ とする。
(S_i : 相関係数・クワストラムのベクトル)

① フレーム相互間の距離行列の計算。

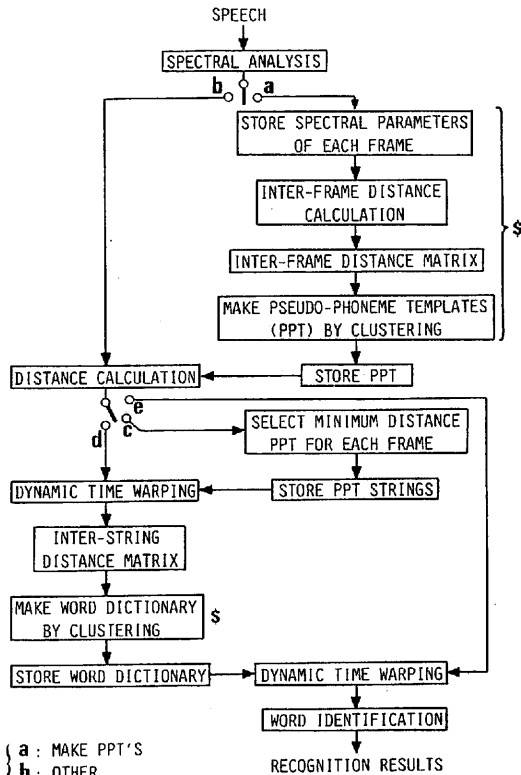
$$D = \{D_{ij}\} = \{\text{distance}(S_i, S_j)\}, 1 \leq i, j \leq n$$

② 各 i に対して $\{D_{ij} < \theta_s, (i, j)\}$ とする j の数 (m_i) の算出。 θ_s : しきい値。

③ 最大の m_i ($1 \leq i \leq n$) を有するフレーム S_i を求め、 S_i から θ_s 以内にある全フレーム $\{S_j\}$ を用いて、 S_j の各要素毎に平均値を計算し、擬音韻標準パターンとして着える: $\bar{S}_i$ 。

④ ③で用いた全フレームを初めのフレームセットから除去する。

②~④の手順を、初め a フレームセットがすべて除去されるか、定めた数 (l) の $\bar{S}_i$ が出力されるまで繰返す。実際には、 l を先に設定して、このとき $\bar{S}_1, \dots, \bar{S}_l$ で表現される全フレームのスペクトル歪が最小



a: MAKE PPT'S
b: OTHER
c: STORE PPT STRINGS
d: MAKE WORD DICTIONARY
e: RECOGNIZE INPUT SPEECH
\$: BY COMPUTER
OTHERS: BY HARDWARE

Fig.1-Principal operation of the speaker-independent word recognition system.

になるように θ_s を決め、このときの標準パターンを用いるのがよい。⁽⁴⁾

2) 単語辞書の作成法

擬音韻標準パターの作成と同じ手法を、各単語毎に多数話者の音声に適用して行う。ただし、①として単語(系列)間距離(DPマッチング後の距離)を用いる。③の処理において、系列間の平均化は行わず、クラスタの中心に位置する系列(擬音韻系列)を単語辞書に蓄える。 θ_s の値は、経験

* A Speaker-Independent Word Recognition System Using Pseudo-Phoneme Templates and A Word Dictionary. By Sadaoki FURUI & Noboru SUGAMURA (Musashino ECL, NTT)

的に次のように定めた。即ち、最初に作られるクラスに全話者の20%程度が含まれ、単独で孤立する系列のみを除いたときにクラス数の数が20程度となり、このとき全話者の60%程度が含まれるように定めた。

8) 未知入力認識

入力音声の各フレームと各擬音韻標準パターンとの距離をWLR尺度により計算し、この値と単語辞書を用いた両端点フリーDPマッチングにより、辞書中の各系列との距離を計算する。この距離が最も小さい3位までの候補単語の距離の値を用いて、多数決論理により単語を決定する。

III. 認識実験

1) これまでの結果⁽²⁾

① 擬音韻標準パターンは、同じ男性であれば誰から作っても認識結果に与える影響は少ない。

② 男女各1名の音声を総合して256種類程度の擬音韻標準パターンを作れば、男女多数の話者に共通に用いることができる。

③ 単語辞書のクラスタ化は、各単語毎にそれぞれ代表的な話者の系列を選ぶのが良い。

2) 31単語認識

研究所内の計算センター及び会議室利用に因する音声認識応答サービス⁽⁵⁾に用いられる31単語を対象として、交換機と内線を通った音声による認識実験を行った。未知音声の話者は男性31名で、擬音韻は男性1

名から作り、単語辞書は、条件1：男女264名の音声を用いてクラスタ化により作成した場合(平均17.6系列/単語)、条件2：本人を除く30名の系列をそのまま着えた場合、の実験を行った。その結果、認識率は条件1：98.0%、条件2：97.2%で、クラスタ化を行えば少ない系列で高い認識率が得られることがわかった。

3) 数字音声認識

男性229名、女性34名(計263名)の数字音声(イチ・サン・ヨン・ゴ・ロク・ナナ・ハチ・チュウ・レイ・ゼロ)の認識実験を行。た。擬音韻は男性1名から作成し、単語辞書は全員の音声のクラスタ化により作成したが(平均22系列/単語)、辞書中に登録された音声は、未知音声から除いた。認識結果は表1の誤認識行列に示す通りで、電話音声としては比較的良い認識率が得られている。

IV. むすび

SPLIT方式には、単語数が増加しても処理量やメモリーの増加が小さいという長所がある。この点を、実効的に単語数が非常に大きくなるマルチテンプレート方式の不特定話者単語音声認識に生かし、物理的尺度によって自動的にシステム構成ができる方式を実現し、高い認識性能が得られることを確認した。この方式は、認識装置の機能と若干のミニコンによる処理でシステムが作成でき、研究所内の音声認識応答サービスに用いられている。

Table 1-Confusion matrix for the digit recognition.

OUT IN	1	2	3	4	5	6	7	8	9	rel zero	RATE
1	228								5	1	97.4
2		235				1				3	98.3
3			243		1					2	98.8
4			6	227				1	2	1	95.0
5			1	1	240					1	98.8
6						244					100
7		1		2		1	242			1	98.0
8	1							240			99.6
9						1			233	4	97.5
rel zero		6		10				1	221	2	92.1
				1	1		1		4	234	97.1
											AVERAGE
											97.5

〔謝辞〕日頃御指導御鞭撻を頂く、当所 畔柳基雄部長、橋本統裕役、坂倉孝四研究室長、青柳工務部長、幸島信昭通信サービス研究室長、ならびに熱心に討論し実験に協力して頂く両研究室の諸氏に感謝する。

〔文献〕(1) 菅村他：音学音声研資 S80-61 (1980)

(2) 古井他：信学総全大 S7-7 (1982)

(3) 古井他：音学秋季議論 2-1-17 (1981)

(4) 菅村他：音学音声研資 S81-T1 (1982)

(5) 駄他：信学総全大 1403 (1982)