

論文 / 著書情報
Article / Book Information

論題(和文)	SPLIT法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)	昭和57年度電子通信学会総合全国大会, , , pp. 5-415-416
Citation(English)	, , , pp. 5-415-416
発行日 / Pub. date	1982,
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) 1982 Institute of Electronics, Information and Communication Engineers.

S7-7

SPLIT 法による不特定話者単語音声認識

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 不特定多数の話者が発声した単語音声の高い精度で認識したいという需要はかなり広く存在しているが、また技術的に困難な点が多い。これまでに検討された種々の方法の中で、比較的高い認識率が得られている方法には、識別関数法⁽¹⁾とヒルチテンプレート法⁽²⁾がある。前者は、多数話者の音声にもとづいて、異なる単語がうまく分離できるような識別関数を作成するものであり、後者は、多数話者の音声をいくつかのクラスに分類して、各クラスを代表する音声を標準パターン(テンプレート)として蓄えておき、入力音声とそれらのすべてのテンプレートとのマッチングを行うものである。前者の方法は、認識のための計算量は少ないが、識別関数の作成法に難しい点があり、後者の方法は、テンプレート数に比例して、認識のための計算量や記憶容量が膨大になる問題点がある。

ここでは、後者の方法に、筆者らがこれまでに検討してきたSPLIT法⁽³⁾による単語音声認識方式を適用することにより、大幅な計算量と記憶容量の削減を実現し、有効性を確認した結果について報告する。

Ⅱ SPLIT 法 この実験では、昨年試作を行った1000単語(1000テンプレート)が実時間で認識できる単語音声認識装置⁽³⁾を用いた。SPLIT法の原理を図1に示す⁽⁴⁾。(a)はスペクトル分析部、(b)は擬音語標準パターン、(c)は距離計算部、(d)は単語辞書、(e)はDP計算部、(f)は単語判定部である。単語辞書には、各単語に対して複数のテンプレート(記号列)が蓄えられている。

入力音声は高域強調ののち、16ms毎に相関分析され、10次のLPC分析によってパワー、自己相関係数、およびLPC係数ストラムの時系列に変換される。入力音声の各フレームと、あらかじめ蓄えてある256個程度の擬音語標準パターンとの距離をWLR尺度⁽⁵⁾によって計算し、距離行列として蓄える。次にこの距離行列と単語辞書を用いた両端点フリーDPマッチング⁽⁶⁾により、各テンプレートとの距離を計算し、この距離が最も小さい3位までの候補単語名(n_1, n_2, n_3)とそれぞれの距離の値(d_1, d_2, d_3)を出力する。各単語に対して複数のテンプレートが用意されているので、 $n_1 \sim n_3$ として同じ単語名が出力される場合が多いが、異なる単語名が含まれる場合もある。ここでは図2に示す多数決論理により、信頼度の高い単語決定を行っている。 θ はしきい値、 n_0 が最終的に出力される単語名である。

音声区間の検出に関しては、短時間パワーに対して、同様の雑音レベルに適合した2種類のしきい値を適用し、これらを越えた区間の継続時間を考慮して、始端と終端の決定を行う。

Ⅲ 擬音語標準パターンと単語辞書の作成法 擬音語標準パターンは、一人または複数の話者が30種程度の単語を発声した音声を用いて、クラス化手法によって自動的に決定し⁽⁴⁾、同じ標準パターンをすべての話者に共通に用いる。単語テンプレートは、各単語ごとに、多数話者の音声を用いて次のようにして作成する。まず単語音声を話者間でDPマッチングして相互間の距離行列を作成し、これをもとに、擬音語の作成と同様の手法により、自動的に話者のクラス化を行う。各クラス

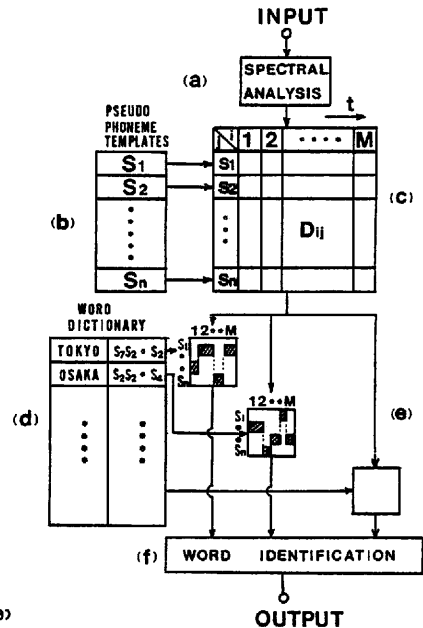


図1. SPLIT 法の原理

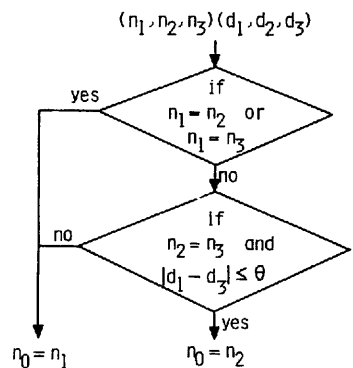


図2. 多数決論理による単語決定

* Speech Recognition Using Strings of Phoneme-Like Templates

IV 実験結果 ここでは具体的タスクとして、研究所内の訂算センターおよび会議室利用に關する音声認識応答サービス⁽⁷⁾をとり上げ、確認語(はい、いいえ)、案内名、ジョブ種別、システム名、週(今週、来週)、曜日、時間帯(午前、午後、1日)、会議室名等の31単語を認識語とした。このうち比較的認識が難しい次のような単語のセットをとり出し、種々の評価実験に用いた。

- Set 1: 曜日(月曜、火曜、…、土曜)
Set 2: システム名の一部(1号、2号)
Set 3: 時間帯の一部(午前、午後)

実験には、多数話者が通常の各種の電話機から、内線および交換機を通して入力した音声を用いた。

4.1. 実験1 男性39名、女性34名の1回ずつ発声したSet 1~3の音声を用い、擬音韻標準パターンを男性1名、女性1名、あるいは両者から作成した場合の認識実験を行った。擬音韻標準パターンは、同じ男性であればどの話者から作っても、認識結果にあまり大きな影響を与えないことがわかっている⁽⁸⁾ので、比較的早めと思われる話者を無作為に選び、その音声を用いた。

この実験では単語テンプレートのクラスタ化は行わず、本人を除く話者(同性)の系列を蓄えて実験を行った。平均認識率は表1に示す通りで、男性から作った擬音韻を用いると認識率が低下するが、男女両方の音声を用いて作成した擬音韻を用いれば、同性から作った場合と同程度の性能が得られることがわかる。

4.2. 実験2 男性39名のSet 1~3の音声を用い、擬音韻を男性1名または4名の音声から作った場合の実験を行った。前者の場合については、男性264名の音声のクラスタ化によって単語テンプレートを作った場合の実験も行った。結果は表2に示す通りで、擬音韻の作成に用いる話者の数が認識結果に与える影響は小さいこと、単語テンプレート数をクラスタ化によって作れば、テンプレート数が比較的小なくても高い認識率が得られることがわかる。

4.3. 実験3 男性31名の31単語を対象として実験を行った。擬音韻は男性1名から作成し、単語テンプレートのクラスタ化は比較的誤り率の高い18単語に対してのみ行い、残りの13単語のテンプレートは単語“はい”のクラスタ化で選ばれた話者の系列を用いた。結果は表3に示す通りで、クラスタ化を行えば、テンプレート数が17.6/30に低下しても、全体として同程度の認識率(97%)が得られることがわかる。ただし内訳は、クラスタ化を行った単語の誤り率が低下しているかわりに、クラスタ化を行わなかった単語の誤り率が増加して、両者の誤り率がほぼ同程度となっており、後者の単語に対してもクラスタ化を行えば、誤り率はさらに低下すると考えられる。

4.4. 実験4 男性258名のSet 1~3の音声を用いて実験を行った。単語テンプレートのクラスタ化を行わない場合は、話者を平均32人ずつのグループに分け、本人以外の系列を蓄えた。結果は表4に示す通りで、クラスタ化を行ったときはテンプレート数が22/31に低下しているにもかかわらず、誤り率も約3/5に低下しており、平均97.5%の良好な認識率が得られている。

V むすび SPLIT法による不特定話者単語音声認識方式を提案し、この有効性を確認した。

謝辞 日頃御指導御鞭撻を頂く当所研基盤研究部長、橋本統祐氏、ならびに板倉分四研究室長に感謝する。

文献 (1) 千葉他: 大アロ・パターン論文集, p.157 (1980) (2) L. R. Rabiner et al.: J. Acoust. Soc. Am. 66-3, p.663 (1979) (3) 古井他: 音学秋季論文 2-1-17 (1981) (4) 菅村他: 音学音声研資 S80-61 (1980) (5) 杉山他: 音学音声研資 S80-58 (1980) (6) 鹿野: 音学音声研資 S80-60 (1980) (7) 舩他: 本大会論文集 (1982) (8) 鹿野他: 音学秋季論文 2-1-18 (1981)

表1. 実験1の認識率(単語: Set 1~3)

		擬音韻標準パターン		
		男性1名	女性1名	男女各1名
入力音声	男性(39名)	2.8%	3.8%	2.8%
	女性(34名)	4.7%	2.9%	2.6%

表2. 実験2の認識率(単語: Set 1~3, 男性39名)

単語辞書	本人除く38名 (38テンプレート)		264名クラスタ化 (平均22テンプレート)
	男性1名	男性4名	男性1名
擬音韻標準パターン			
誤認識率	2.8%	2.6%	1.3%

表3. 実験3の認識率(31単語, 男性31名)

単語辞書	本人除く30名 (30テンプレート)	クラスタ化 (平均17.6テンプレート)	
		単語別クラスタ化 (平均22テンプレート)	共通クラスタ化 (平均14テンプレート)
18単語	3.9%	2.9%	
13単語	1.2%		2.7%
計	2.8%		2.8%

表4. 実験4の認識率
(単語: Set 1~3, 男性258名)

本人除く平均32名 (平均31テンプレート)	クラスタ化 (平均22テンプレート)
4.3%	2.5%