

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)	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. 553-554
Citation(English)	, , , pp. 553-554
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SPLIT法による特定話者大語い単語音声認識について報告してきたが⁽¹⁾、本稿では、マルチテンプレート法による不特定話者少数語いの認識について検討を行った。

2. 不特定話者単語音声認識

不特定話者単語音声認識については、これまでいくつかの方法が提案されて⁽²⁾⁽³⁾いるが、本稿では同一単語に対し複数の標準パターンを用意するマルチテンプレート法を採用し、パラメータレベルの標準パターンを用いる場合とSPLIT法との比較を行う。SPLIT法は、DPのためのスペクトル距離計算量が、マルチテンプレートも含めた認識対象語数に依存しないため、マルチテンプレート法には有効な方法といえる。

3. マルチテンプレートの作成法

話者 i と話者 j が発声した単語 X を $W_i^{(x)}$ と $W_j^{(x)}$ とし、 $W_i^{(x)}$ と $W_j^{(x)}$ の間でDPマッチングを行ったときのマッチング距離を $D_{ij}^{(x)}$ とする。スペクトル距離の閾値 Θ_s をあらかじめ定めておき、テンプレート作成に用いるすべての話者 j との比較において、 $D_{ij}^{(x)} \leq \Theta_s$ となる個数を各話者 i に対し計算する。これを m_i とする。最大の m_i をもつ単語 $W_i^{(x)}$ を中心単語パターンとよび、マルチテンプレートの一つの候補とする。同時に $W_i^{(x)}$ に Θ_s 以内のDPマッチング距離で含まれる単語を用いて後述するアルゴリズムで平均化操作を行い、平均単語標準パターン $\overline{W_i^{(x)}}$ を作成する。

つぎに $W_i^{(x)}$ とのDPマッチング距離が Θ_s 以下の単語と単語 $W_i^{(x)}$ をもとの単語セットから除く。以後このような処理をくり返し、 $W_i^{(x)}$ と $\overline{W_i^{(x)}}$ $i=1, 2, \dots$ を求める。これらの中から必要数を選びマルチテンプレートとする。つぎに平均化アルゴリズムについて説明する。中心単語パターン $W_i^{(x)}$ を基準にとり、同じクラスター(DPマッチング距離が Θ_s 以下)に属する単語パターン $W_j^{(x)}$ の間で、図1に示すような方法でDPマッチングによりフレームの対応をとる。 $W_i^{(x)}$ のフレーム間で相関係数を加算し平均をとることによって平均単語標準パターン $\overline{W_i^{(x)}}$ を作成する。このとき $W_i^{(x)}$ の1フレームが、 $W_j^{(x)}$ の複数のフレームに対応するときは、対応したフレームの一番始めのフレームの値のみを用いる。逆に $W_i^{(x)}$ の複数のフレームが $W_j^{(x)}$ の1フレームに対応づけられる場合は、 $W_i^{(x)}$ のそれぞれのフレームの値に $W_j^{(x)}$ の対応するフレームの値を加算する。

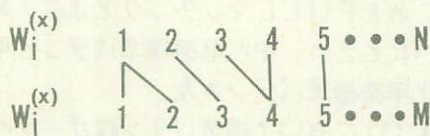


Fig. 1 Time registration of two words which belong to the same cluster

4. 実験方法 表1に示す条件で実験を行った。45名の発声者を15名ずつの3組に分け(これをA, B, Cとする) AとBを用いて前述した方法でマルチテンプレートを作成し、Cの認識実験を行

*Speaker-Independent Recognition of Isolated Words by SPLIT Method.
by Noboru SUGAMURA, Sadaoki FURUI (Musashino E.C.L., N.T.T.) and
Kazuo HAKODA (Yokosuka E.C.L., N.T.T.)

った。このような実験をA,B,Cの組み合わせを変えて3回行い、認識率の平均値を求めた。

Table.1 Experimental condition

発声者	男性45名(電話音声 601形および4号電話機+擬似加入者線 Q.5mmφ -4dB)
単語数	10数字を含む25単語
分析条件	8kHzサンプリング, プリエンファシス フレーム周期10msec
分析パラメータ	16次ケプストラム
DPマッチング	両端点フリー ⁽⁴⁾

5. 擬音韻標準パタンの作成と擬音韻による単語標準パタンの作成法

擬音韻標準パタンの作成には、前述した本人を含まない2組の話者の組み合わせによる30人分の10単語を用いて行った。⁽¹⁾ 3章で説明した方法によって作成した平均標準パタンの各フレームと擬音韻標準パターンとのマッチングを行い、擬音韻によるマルチテンプレートを作成する。実験で用いた擬音韻標準パターン数は128, 256, 400である。

6. 実験結果 実験は、つぎのような項目について行った。

(1) 擬音韻標準パターンを介さずに、各フレーム間の距離を直接計算する方法(以下これをFULLマッチングとよぶ)を用いたときの、中心単語標準パターンと平均化単語標準パタンの差。

(2) (1)において標準パターン数が一の場合と複数の場合(マルチテンプレート)との比較。

(3) マルチテンプレートでSPLIT法を用いる場合とFULLマッチングとの比較。

(4) 擬音韻標準パターン数の効果

項目(1)については、マルチテンプレートの数を30(異なる話者の組の全員のパタンを用いる)としたときの平均認

識率は94.1%であった。つぎに単語標準パターン数を1単語当たり1~10個の範囲で変えて実験を行ったところ、平均化パターンを用いる場合の方が、中心パターンを用いる場合よりも3~6%認識率が高く同じ認識率を得るための標準パターン数は後者の場合の半分程度に少なくてもことがわかった。項目(2)に関しては、5個の平均化標準パターンを用いたときの認識率は90.0%で、ただ1個の平均化標準パターンを用いる場合に比べると約6%の認識率の向上がみられたが、30パタンの場合と比べると約4%の認識率の低下がみられた。項目(3)に関しては、単語標準パターン数が5、擬音韻標準パターン数が256のとき、SPLIT法により約4%の認識率の低下がみられた。平均化された標準パターンに対するSPLIT法の適用法の改良が必要である。項目(4)に関しては、擬音韻パターン数が128と256の場合の認識率を比較すると、後者の方が2%程度高いが、256と400の間では差がなかった。誤認識単語の分析を行ったところ、SPLIT法でもFULLマッチングでも大差はなく、シチ↔シ、イチ↔シチの誤りが多いが、音声区間の検出誤りに起因しているとみられるものも含まれている。

7. あとがき SPLIT法を不特定少数単語音声認識に適用し、種々の検討を行った。今後、擬音韻標準パターン作成法等に改良を加え、認識率の向上につとめたい。

謝 辞 日比岡伸博氏、田中基礎研究部長、板倉第四研究室長に感謝します。

参考文献

- (1) 菅村, 箱田, 右井: 音声研究会資料 S80-61(1980)
- (2) L.R. Rabiner, S.E. Levinson, A.E. Rosenberg and J.G. Wilpon: IEEE Vol.1 ASSP-27, No.4 (1979)
- (3) 奥住, 藤田, 藤島: 音声研究会資料 S80-21(1980)
- (4) 鹿野: 音声研究会資料 S80-60(1980)