論文 / 著書情報
Article / Book Information

論題(和文)	擬音韻標準パタンによる大語い単語音声認識
Title(English)	
著者(和文)	管村昇, 古井貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	電子情報通信学会論文誌, Vol. J65-D, No. 8, pp. 1041-1048
Citation(English)	, Vol. J65-D, No. 8, pp. 1041-1048
発行日 / Pub. date	1982, 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) 1982 Institute of Electronics, Information and Communication Engineers.

電子通信学会論文誌　別刷　　Trans. IECE '82/8 Vol. J65-D No. 8

UDC 801.47:[159.937:534.784/.785]:681.322.05

擬音韻標準パタンによる大語い単語音声認識

正員　菅村　昇†　　正員　古井　貞熙†

Large Vocabulary Word Recognition Using Pseudo-Phoneme Templates

Noboru SUGAMURA† *and* Sadaoki FURUI†, *Regular Members*

あらまし　学習音声サンプルのクラスタリングによって擬似的な音韻標準パタンを作成し，この擬音韻標準パタンの系列で表現した単語辞書とDPによって認識を行う新しい単語音声認識法（SPLIT法とよぶ）を提案する．この方法はDPのための距離計算量，単語辞書のメモリ量が，単語のすべてのフレームのパラメータを単語辞書に蓄積しておく方法に比べ，少くて済むという特徴をもつ．この特徴は，認識すべき単語数が多くなればなるほど有効である．本論文では，この認識方法を特定話者，大語い（641都市名）単語音声認識に適用した結果について報告する．また誤認識単語の分析結果についても述べる．男性4名のデータを用いて実験を行った結果，擬音韻標準パタン数を64～256に選べば，単語の全フレームのパラメータを用いる場合との認識率の差は小さく，本論文で提案する認識方法が，特定話者の大語い単語音声認識に適用できる見通しが得られた．

1. まえがき

音声認識の中でも，単語ごとに区切って発声された音声を認識する単語音声認識は，セグメンテーションや音韻情報との対応づけなどむずかしい問題を回避することができ，特定の使用条件のもとでは実用化されはじめている．単語音声の認識方法としては，これまで主に単語全体を一つのパタンとみなして，単語単位のパラメータレベルでの辞書を用いて認識する方法[1]や音韻を単位とする標準パタンを作成し，この標準パタンを用いた単語辞書を使って認識する方法などがある[2]．入力単語音声と標準パタンとのマッチングには，両者ともダイナミックプログラミング（以下DPと略記する）の手法が用いられている．

ところで単語音声認識技術を音声入力ワードプロセッサなどへ適用しようとした場合，認識対象語い数が数千語程度となり，大語いに対する能率の良い単語音声認識手法を確立しておくことが重要となる．認識対象語い数が増加するにしたがって，前述した前者の認識方法では，単語辞書のメモリ量やDPマッチングの演算量が，ほぼ単語数に比例して増加する．後者はメモリ量や演算量の増加は，前者に比べて少く[3]，また話者変化に対する学習の容易さ[4]，単語辞書の自動作成[5]などの長所をもつ反面，音韻標準パタンの作成に人間による作業を必要とすること，複雑に変化するスペクトルパタンの時系列を，きわめて少数の音韻標準パタンで表現することにより，類似語の多い大語い単語認識においては，前者ほどの認識率が期待できないことなどの問題点がある．

本論文では，これら両認識方法の中間的なものとして，音韻との対応にとらわれない方法で標準パタンを作成し，このパタンを用いて単語認識を行う方法の提案と実験結果について報告する．

2. SPLIT法による単語音声認識

単語音声認識において，最終的な認識結果が正しければよいという立場に立てば，認識すべき単語が，どのような音韻から構成されているかは，必ずしも知る必要がない．また単語中のフレームごとのスペクトルパラメータのすべてが単語を認識するのに必要であるわけでもなく，スペクトルパラメータについても一種の量子化をした方が，単語辞書に対するメモリ量を減少させることができる．このような考え方を基本として，以下に説明するような新しい単語音声認識方法を

†電電公社武蔵野電気通信研究所，武蔵野市
Musashino Electrical Communication Laboratory, N. T. T., Musashino-shi, 180 Japan
論文番号：昭57-論 427[D-123]

提案する．

2.1 単語音声認識システム

前述した考え方をもとに，標準パタンの作成方法にクラスタリングの手法を用いる新しい単語認識法について述べる．この認識方法を，クラスタリングによって得られる擬音韻標準パタンの系列を単語辞書として用いることと，擬音韻標準パタンの構成にパラメータ分布空間を分割していくことを兼ねてSPLIT法(Word Recognition System Using Strings of Phoneme-Like Templates)とよぶことにする[6]．

認識システムの構成を図1に示し，認識方法について説明する．認識手順は，つぎの3つの段階に分けられる．

(1) 擬音韻標準パタンの作成
(2) 単語辞書（擬音韻パタン系列）の作成
(3) 単語音声認識

(1), (2), は，(3)の準備段階である．

2.1.1 擬音韻標準パタンの作成

未知音声入力の認識に先立って，あらかじめいくつかの単語を発声し，これを擬音韻標準パタン作成用のデータとして用いる．いまこのデータをＬＰＣ分析し

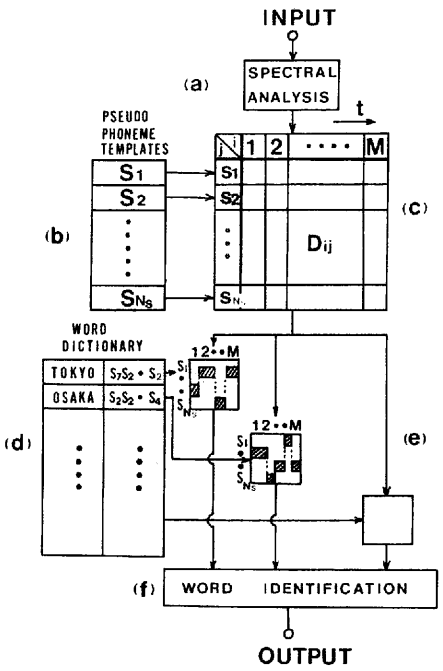

図1 ＳＰＬＩＴ方式による単語音声認識
Fig. 1 – Block diagram of the word recognition system using strings of phoneme-like templates.

P_1, P_2…, P_N のN個のフレームのパラメータセット（以後パタンとよぶ）が得られているとする．それぞれのパタンP_iは，n個のＬＰＣケプストラムパラメータで構成されているとし；これを$p_{i,k}$（$k=1, 2, \cdots, n$）で表わす．このような学習データを用いて，擬音韻標準パタンを作成する方法について以下に説明する[7]．

〔Step 1〕

N個のパタンP_iのそれぞれに対して，P_iとは異なるP_jとのスペクトル距離を次式で計算する．

$$D_{ij}=D(P_i, P_j)=\sum_{k=1}^{n}(p_{i,k}-p_{j,k})^2 \quad (1)$$

$i=1, 2, \cdots, N$
$j=1, 2, \cdots, N$, $i \neq j$

D_{ij} な各jについて計算し，$D_{ij} \leqq \theta_s$ 満たすパタン数を各iについて求める．このパタン数を$m_i=1, 2, \cdots, N$とする．ここでθ_sは，あらかじめ設定しておくスペクトル距離のしきい値である．

〔Step 2〕

最大のm_iをもつパタンP_iを見出し，P_iとP_jにθ_s以下の距離で含まれるすべてのパタンを用いて，これらのパタンのケプストラムパラメータに対応する自己相関係数を求め，この平均値を各要素ごとに計算し，擬音韻標準パタンS_1として蓄える．

〔Step 3〕

Step 2で用いたすべてのパタンを，はじめのパタンセットから除去し，Step 1，Step 2を，もとのパタンがなくなるまでくり返す．このとき(1)の距離計算は，はじめの1回だけ計算しておけばよい．2回目以後は，この距離行列を参照し，各パタンごとにθ_s以下のパタンを検索し，このパタンが，それ以前のどの擬音韻標準パタンにも含まれなかったかどうかだけを調べればよいのである．

以上の手順によって作成された擬音韻標準パタンセットを$S_1 \sim S_q$とする．qはθ_sの値によって1からNまで変化が可能である．θ_sを大きく設定することは粗い量子化を意味しており，θ_sを小さくとれば，より細かいスペクトル表現が可能となるかわりに，擬音韻標準パタン数が増加することになる．したがってθ_sの値は単語認識システムにおいて，どれくらいの擬音韻標準パタンを最終的に用いるかということから決定する必要がある．この擬音韻標準パタンの作成方法について，与えるべきパラメータはθ_sのみで，簡単に標準パタン

を構成できる．以上のような手順で決定された擬音韻標準パタンS_1～S_qのうち，設定されたパタン数だけを，図1の(b)に蓄積する．このとき必要なメモリ量は，選択する擬音韻標準パタン数をN_s，スペクトルパラメータの個数をn，1つのパラメータ精度をN_aビットとすると

$$N_s \times n \times N_a \quad \text{ビット} \tag{2}$$

となる．

2.1.2 単語辞書（擬音韻標準パタン系列）の作成

擬音韻標準パタンを用いて，認識対象単語をこれらの系列として表わす単語辞書を作成する．すなわち認識対象単語を発声し，スペクトル分析したパラメータと擬音韻標準パタンとのスペクトル距離計算をフレームごとに行い，スペクトル距離が最小となる擬音韻標準パタンを選択する．このようにして認識対象単語を擬音韻標準パタンの系列で表現し，図1(d)に蓄積しておく．このように単語辞書を擬音韻標準パタンの系列で表現することにより，単語辞書のメモリ容量が分析されたパラメータそのものを蓄積しておく方式に比べ大幅に少くて済む．たとえば，認識対象単語総数をL，単語lのフレーム数をM_l，擬音韻標準パタンを指定するのに必要なビット数をn_b（$=\log_2 N_s$）ビットとすると，単語辞書のメモリ容量は，

$$\left(\sum_{l=1}^{L} M_l \right) \times n_b \quad \text{ビット} \tag{3}$$

となる．

2.1.3 単語音声認識

あらかじめ準備した擬音韻標準パタンと単語辞書とを用いて，未知入力音声の認識を行う．入力音声は，図1(a)でスペクトル分析され，フレーム周期ごとに擬音韻標準パタンとのスペクトル距離計算が行われ距離行列が生成される．この距離行列と単語辞書を用いて，音声の時間伸縮を吸収するスペクトルマッチングをDPによって行い，マッチング距離が最小となる単語を認識結果として出力する．なおこの段階においてマッチング距離の値にしきい値を設けて，誤認識をさけるために，リジェクトすることも可能である．DPマッチングを行う場合の距離計算は，一つの未知入力単語に対して1回だけ計算しておけば，認識対象単語のすべてに利用できる．このためDPのための距離計算量が，単語単位のパラメータレベルの辞書を用いて認識する方法に比べ，大幅に減少できる．

2.2 SPLIT法の特徴

単語認識方法として，単語単位の標準パタンを用いて認識する方法（方法Aとする）と音韻を単位とする標準パタンを作成し，このパタンを用いた単語辞書を使って認識する方法（方法Bとする）をとりあげ，SPLIT法の特徴について，これらの方法と比較して説明する．

(1) 標準パタン作成に関して，方法Bでは音韻に1対1に対応するスペクトルの標準パタンを決定することがむずかしい．すなわちどれだけの学習データをもとに，どのような方法で各音韻の標準パタンを構成すればよいかということについて確立した手法はなく，ある程度熟練した人手にゆだねられることが多い．これに対して，SPLIT法では標準パタンの作成に，自動的なクラスタリング手法を用いており，つぎのような特徴をもっている．

(i) パラメータ分布空間に距離を導入することだけで標準パタンを作成することができ，人手が介入する過程がない．このことは大量のデータを処理するのに適しており，データベースが同じであれば，どこでも同じ標準パタンを作成することができ，あいまいさがない．

(ii) 国語に依存しない標準パタンを構成できるので，認識対象の言語に対する適応性がある．たとえば認識対象語が英語などであっても同様の処理で行える．

(iii) 認識対象単語の種類にかかわりなく標準パタンを構成することができ，単語を構成している音韻を意識する必要がない．

(2) SPLIT法では，DPマッチングのための距離計算量および標準パタン，単語辞書用のメモリ量が，方法Bに比べれば少し増加する可能性があるが，方法Aに比べれば大幅に減少できる．方法Aと比べ，計算量，メモリ量の比較を行うとつぎのようになる．

認識対象単語数をL，未知単語のフレーム数をM，DPの窓幅をN_w，擬音韻標準パタン数をN_sとすると，1単語を認識するためのDPに必要なスペクトル距離計算回数は，方法Aでは，約$M \times N_w \times L$であるのに対して，SPLIT法では，$M \times N_s$となる．したがって，SPLIT法の方法Aに対する1フレームあたりの距離計算量の比は

$$N_s / (N_w \times L) \tag{4}$$

となる．

つぎに単語辞書のメモリ量の比較を行う．各単

語のフレーム数をM_l，擬音韻標準パタン数をn_bビット，1フレームのスペクトルパラメータの個数をn個，1つのパラメータ精度をN_aビットとすると，方法Aでは

$$\left(\sum_{l=1}^{L} M_l\right) \times n \times N_a \quad \text{ビット} \qquad (5)$$

であるのに対し，SPLIT法では，(2)，(3)式より

$$N_s \times n \times N_a + \left(\sum_{l=1}^{L} M_l\right) \times n_b \quad \text{ビット} \qquad (6)$$

である．

SPLIT法の方法Aに対するDPのための距離計算量の比および標準パタン，単語辞書用のメモリ量の比を表1に整理して示す．表1には，具体的な数値を用いて，それぞれの比の概算値も示している．また表1の条件で単語数とこれらの比との関係を図2に示す．これらの結果から，メモリ量，DPのための距離計算量の点からは，SPLIT法は，認識対象語い数が多くなればなるほど，方法Aに比べて有利であることがわかる．

以上，SPLIT法の長所を方法AとBと比較して説明したが，問題点としては，つぎのようなことがあげられる．

(3) SPLIT法は，方法Aにおいて単語辞書のスペクトルパラメータをN_s種に量子化して表現したものを利用する方法と考えることができる．このことから量子化が粗い，すなわちN_sが小さい場合には，量子化誤差が大きくなり，スペクトルパラメータの時系列を精度よく表現できず，方法Aに比べ認識率が低下することが予想される．これは似かよった単語が多くなる大語いの単語認識の場合に特に注意を要する．また学習サンプルを用いてあらかじめ擬音韻標準パタンを作成する過程が必要であり，このためのメモリや演算が必要であることなどが，方法Aに比べての問題点である．

(4) SPLIT法や方法Aでは，スペクトルパタンと音韻との関係は，単語認識においては必ずしも必要としないという立場であり，認識対象語いについては，利用者が認識に先だってあらかじめ発声して登録しておくことを前提にしている．このため話者が変わった場合や単語辞書を追加する場合に利用者に大きな負担となる．これに対し方法Bでは，標準パタンと音韻の間で対応がついているため，辞書の自動作成ルールができれば，認識対象単語を発声しなくても単語辞書を作成することができる．また話者が変わった場合には，学習によってその話者の音韻標準パタンが作成できれば，話者への適応もきわめて容易に行える．

以上説明したように各認識方法とも長短所はあるが，本論文では，方法Aに比べ，メモリ量，DPのためのスペクトル距離計算量の減少効果の大きいSPLIT法を，特定話者，大語い単語音声認識に適用し認識率を評価尺度として検討する．ここでは認識対象単語はすべて発声して単語辞書を作成することを前提としており，発声しないで単語辞書を自動的に作成する方法や，話者が変わった場合に，単語辞書をその人に合わせてどのように変更するかについては今後の課題とする．

表1 SPLIT方式と方法Aのメモリ量，計算量の比較

項目 ＼ SPLIT法 / 方法A	例 L = 500単語 $\sum_{l=1}^{L} M_l$ = 25000 N_s = 128パタン n_b = 7ビット N_a = 16ビット n = 16パラメータ N_w = 15フレーム
1フレームあたりのDPのための距離計算量	$\frac{N_s}{N_w \times L}$	0.02
標準パタンおよび単語辞書用メモリ	$\{N_s \times n \times N_a + \sum_{l=1}^{L} M_l \times n_b\} / (\sum_{l=1}^{L} M_l) \times n \times N_a$	0.03

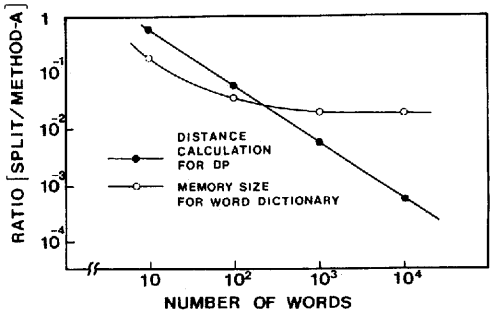

図2 DPのための距離計算量，単語辞書のメモリ量の比と単語数との関係

Fig. 2 – Relationship between the number of words and reduction rate from the viewpoint of memory size and the amount of calculation for DP matching.

3. 大語い単語音声認識

前章で説明したように，SPLIT法は，単語単位にスペクトルパラメータそのものの辞書を用いて認識す

る方法に比べ，認識率がそれほど低下しないなら，大語いになればなるほど有効性が発揮できる．本章では，SPLIT法を大語い単語音声認識に適用した実験結果について述べる．

3.1 認識実験条件

男性4名が，約2週間程度の期間をおいて2回発声した日本の都市名641単語を認識対象語として実験を行った．実験の諸条件を表2に示す．

この実験においてまず問題となるのは，641単語（16msのフレーム周期で約30000フレーム）の全部を用いて，前述したアルゴリズムで擬音韻標準パタンを構成しようとすれば，パタン相互間の距離計算量が膨大になる．そこで641単語の中から，その単語に含まれる音韻を意識せずに，ランダムに選んだ64個の単語（約3000フレーム）を用いて擬音韻標準パタンの作成を行い，256個を選択した実験に用いた．この場合，**2.3**で説明したSPLIT法の方法Aに対するDPのための距離計算量，および単語辞書のメモリ量の比は，それぞれ0.03，0.04となる．またDP漸化式の計算量まで含めた認識に要する総計算量の比は，約0.15となる．

1回目の発声データを用いて擬音韻標準パタンと単語辞書を作成し，2回目の発声データを未知サンプルとしたときの認識率と，その逆の場合の認識率とを各話者について求め，4名の話者で平均した結果を図3に示す．横軸は順位で，縦軸は，その順位までに正解が含まれる割合を示している．図3には，方法Aによる結果[8]も合わせて示している．距離尺度としては，ケプストラム距離（CEPと略記）と重みつき最尤スペクトル距離（Weighted Likelihood Ratio，WLRと略記）[9]の2種類を用いて行った．

この実験結果からつぎのことが明らかとなった．

(1) スペクトル距離尺度としてCEPを用いれば，SPLIT法と方法Aとの，第1位の認識率の差はない（SPLIT法が方法Aに比べ0.1％高い）．

表2 認識実験条件

実験対象単語	……	国内都市名641単語
発声者	……	男性4名，2回発声
分析条件	……	8kHz サンプリング32ms，ハミング窓フレーム周期16ms
特徴パラメータ	……	LPCケプストラム，自己相関係数 } 16次
スペクトル距離尺度	…	ケプストラム（CEP）または重みつき最大スペクトル距離（WER）

(2) スペクトル距離尺度としてWLRを用いた場合，CEPを用いた場合に比べ，SPLIT法では第1位の認識率は1％高くなるが，方法Aの場合の認識率の向上に比べれば，0.5％低くなる．これは方法Aの場合に比べ，擬音韻標準パタンを用いることにより，スペクトルが平均化され，スペクトルのピーク付近への重みつけ効果が小さくなるためと考えられる．

(3) 擬音韻標準パタンの作成用として64個の単語だけを用いたが，方法Aと比べ，第1位の認識率の低下は，WLR距離尺度の場合で0.4％であり，64個の単語から作成される擬音韻標準パタンが，それ以外の単語辞書に対しても有効であるということが実験的に明らかとなった．

(4) 話者による認識率の差は，最大1％であり，SPLIT法による認識率のみが極端に低下する話者はいない．

(5) 方法Aにおいて認識率の高い話者は，SPLIT法によっても高い認識率が得られた．

3.2 誤認識単語のエラー分析

前述した話者4名の認識実験におけるすべての誤認識単語について，誤認識のタイプをつぎのA～Eの5つに分類して図4に示す．Aは音声区間の検出エラー，Bは母音のエラー（UOZU→O：ZUなど），Cは語頭のみの子音のエラー（SAGA→KAGAなど），DはCに含まれない子音のエラー（O：GAKI→O：

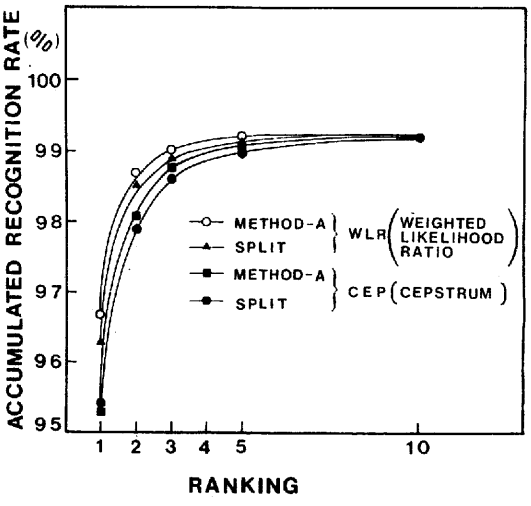

図3 順位と累積認識率との関係
Fig.3 – Comparison of accumulated recognition rate in relation to ranking.

MACHI など），Eはその他のエラー（MINOO→MINO など）である．この図からわかるように誤認識の中では，音声区間の検出エラーによるものが最も多い．これは騒音下（計算機室で騒音レベル70～85 dBA ）で収集した音声を用いたためである．ついで語頭の子音のエラーによるものが多いが，音韻的にきわめて近い単語は，DPマッチングのようにスペクトルマッチング量を総合的に評価する方法では，その認識能力にも限界があり，別のアプローチを考える必要がある[10]．誤認識単語の中で，母音の誤りに起因するものは，全体的に非常に少ない．またスペクトル距離尺度をWLRとした場合，Dのエラーが著しく減少することが実験的に明らかになり，WLR距離尺度の一つの有効性を示す結果が得られた．

3.3 擬音韻標準パタン数と認識率との関係

2.で述べたように，SPLIT法においては，未知入力音声は，フレーム周期毎に擬音韻標準パタンとの距離をただ一回計算しておけば，この距離計算結果を，すべての認識対象語いのDPマッチングに利用できる．この距離計算量は，擬音韻標準パタン数に比例する．ここでは，擬音韻標準パタン数を減少させた場合の認識率の変化を，先ほどと同じ4名の男性話者のデータを用いて検討を行った．ただしスペクトル距離尺度としてはCEPだけを用いた．また実験の結果，1回目と2回目の発声データのうち，どちらを未知データとして扱っても認識率が大差がないことから，2回目のみを未知データとして実験を行った．θ_1を0.20としたときのクラスタリングで得られた擬音韻標準パタンの中から，同一クラスタに属するパタン数が多い順に16，32，64，128，256個を選択し認識実験を行った．それぞれの擬音韻標準パタン数における認識率(第1位，2位以内，3位以内)を，4名の話者について平均した結果を図5に示す．この結果擬音韻標準パタン数を64にすれば256の場合に比べ1%低下し，32にすれば1.5%程度低下することが明らかになった．

2.で述べた擬音韻標準パタン構成法からわかるように，擬音韻標準パタンはパラメータ分布空間において，分布密度の高い部分から生成されていく．したがって生成される順位が上位の標準パタンと下位の擬音韻標準パタンとでは単語辞書で使用される頻度が異なり，上位の標準パタンほど重要な役割を果たすと考えられる．このため擬音韻標準パタン数を32程度にしても認識率の低下はわずかである．下位の擬音韻標準パタンを追加していくことによりわずかずつ認識率が向上していくことから，これらのパタンは出現頻度は高くはないが，スペクトルパタンの差異を細かく表現するのに寄与していると考えられる．

以上の実験結果から，擬音韻標準パタン数を減少させても64～256程度に選べば，認識率の低下は方法Aに比べ1%程度に押えられることが明らかになった．また正解が3位以内に入る割合は，擬音韻標準パタン数を16個にしても，98%以上であることが明らかになった．

以上は1回だけのクラスタリングで得られた擬音韻標準パタンの中から標準パタンを選択した場合であるが，2.で述べたように，最終的に用いる擬音韻標準パタン数が決定している場合，それに応じたクラスタリングのしきい値を設定し擬音韻標準パタンを作成する

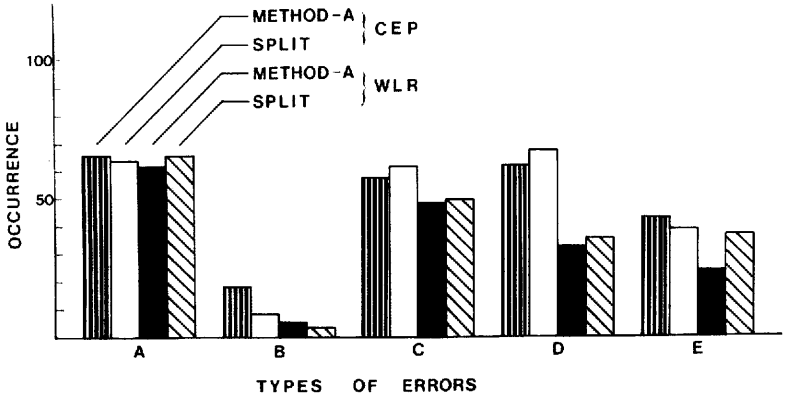

図4 誤認識単語のエラーの種類による分類
Fig.4 - Classification of recognition errors.

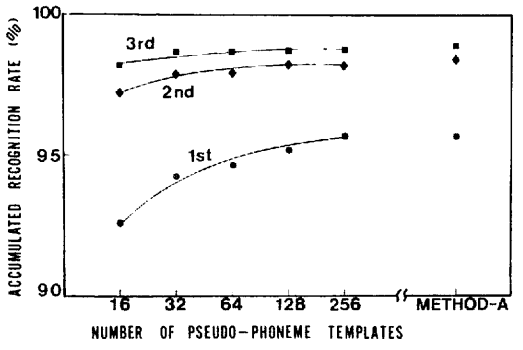

図5　擬音韻標準パタン数と正解が第1位，2位以内，3位以内に含まれる割合との関係

Fig.5 - Relationship between accumurated recognition rate and the number of pseudo - phoneme templates(number in figure means ranking).

ことが望ましい．クラスタリングのしきい値を適切に設定することにより，パターン数16個で93%，32個で96%の認識率が得られる（話者1名の実験結果）ことが明らかになっている[11]．

3.4　擬音韻標準パタンと音韻性との関係

ここでは，擬音韻標準パタン数を32個とした場合の，擬音韻標準パタンの位置関係と音韻性との関係について述べる．擬音韻標準パタンの位置関係は，多次元尺度構成法（Kruskal の方法）[12]を用いて2次元で表現した．一方発声された単語の各フレームについて，あらかじめ人間が音韻の種類を付与しておき，これとマッチングされた擬音韻標準パタンとの対応をとり，各音韻ごとに度数分布をとる．これらの結果を5母音について整理したものを図6に示す．擬音韻標準パタン2，7，11，15，27，28が1つのクラスタ（/a/）に含められているのは，これらのパタンは5母音のうち，/a/に最もよく対応しているという意味である．さらに/a/に対応しているパタンの中で，最も頻度の高いものについてスペクトルの形状を図示している．これらの結果，32個の擬音韻標準パタンの多くは，図6に示すように5母音のどれかに対応しており，しかもよく分離されていることがわかる．これが擬音韻標準パタン数を32個程度に減少しても認識率がそれほど低下しない一つの理由と考えられる．スペクトルパタンの高域にピークが見られるのは，前述したように音声を計算機室で収集したため室内騒音が重畳したことによる．

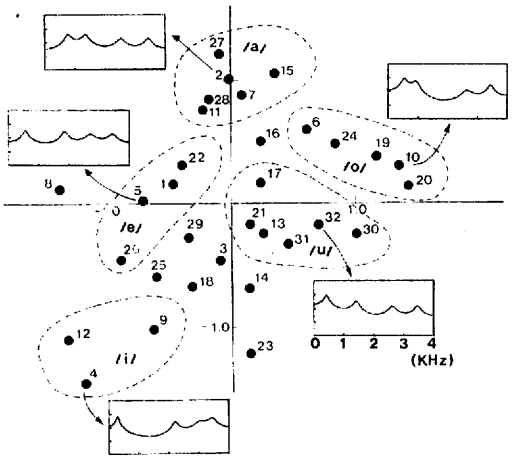

図6　32個の擬音韻標準パタンの位置関係と音韻性との関係

Fig.6 - Locations of 32 pseudo - phoneme templates by multidimensional scaling.

4. むすび

本論文では，新しい単語音声認識方法について，その手法と特徴を従来の代表的な認識方法と比較して述べ，大語いの単語音声認識実験を通じてパラメータレベルの単語辞書を用いる方法との認識率の差を明らかにした．この結果，標準パタン数を64～256個程度に選べば，パラメータレベルの単語辞書を用いる場合と比べ，DPのための距離計算量と単語辞書用の記憶容量が大幅に少くて済み，しかも認識率にはほとんど差がないことが明らかとなり，ここで提案した認識方法が，特定話者大語い単語音声認識に利用できる見通しが得られた．また本論文で提案した認識方法は，DPのための距離計算量が認識対象語い数に依存せず，マルメテンプレート法による少数語い不特定話者の単語認識にも適用できる可能性がある[13]．今後の問題としては，DPそのものの計算量を減少させるためのマッチングすべき単語の能率よい予備選択法[14]，話者変更に伴う単語辞書のより能率のよい書き換え法，新しい単語辞書の追加法などがある．

謝辞　日頃御指導頂く畔柳基礎研究部長，橋本基礎研究部統括役，板倉第四研究室長に深謝します．音声データの収集等に御尽力された横須賀電気通信研究所宅内機器研究部音声入出力方式研究室の箱田調査員に感謝します．また認識実験に関していろいろ御協力，御助言を頂いた第四研究室鹿野調査員，杉山研究主任

に感謝します．最後に，熱心に御討論頂く第四研究室の諸氏に御礼申し上げます．

文　献

(1) 迫江，千葉：“動的計画法を利用した時間正規化に基づく連続音声認識”，音学誌，27，9，p.483(1971)．
(2) 好田，橋本，斎藤：“数字音声の機械認識系”，信学論(D)，55-D,3,pp.186-193(昭47-03)．
(3) 長島，中津：“音韻単位の標準パタンを用いた実時間単語音声認識装置”，音声研究会資料，S78-22(1978-06)．
(4) 好田，橋本，斎藤：不完備型の学習サンプルによる音声の認識について”，信学会インホメーション研資，IT71-61(1971)．
(5) 箱田：“音韻を単位とする単語音声認識における辞書の自動作成法”，日本音響学会講論集，1－1－19(1980-10)．
(6) 管村，箱田，古井：“クラスタ化による音韻標準パタンを用いた単語音声認識”，音声研究会資料，S80-61(1980)．
(7) 管村，板倉：“フレーム単位のスペクトルマッチングによる音声情報圧縮”，日本音響学会講論集，2-2-7(1979-06)．
(8) 鹿野：“大語い単語音声認識におけるＬＰＣスペクトルマッチング尺度の評価”，音声研究会資料，S30-60(1980)．
(9) 杉山，鹿野：“ピークに重みをおいたＬＰＣスペクトルマッチング尺度の提案”，音声研究会資料，S30-03(1980)．
(10) 古井：“スペクトル時系列の大局的・局所的特徴の組合せによる単語音声認識”，日本音響会講論集，1-1-18(1980-10)．
(11) 管村，古井：“ＳＰＬＩＴ法における標準パタン数の効果”，昭57信学総全大，1367．
(12) ニャナデシカン著，丘本，磯貝訳：“統計的多変量データ解析”(昭54)．
(13) 管村，古井，箱田：“SPLIT法による不特定話者単語認識”，日本音響学会講論集，3-1-17(1981-05)．
(14) 管村，古井，箱田：“予備選択による大語い単語音声認識”，日本音響学会講演集，3-1-16(1981-05)．

(昭和56年12月21日受付，57年3月18日再受付)