論文 / 著書情報
Article / Book Information

論題(和文)	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
Title	
著者(和文)	菅村 昇, 箱田 和雄, 古井 貞熙
Author	SADAOKI FURUI
出典(和文)	日本音響学会 音声研究会資料 S80-61, , S80-61, pp. 479-486
Journal/Book name	, , S80-61, pp. 479-486
発行日 / Issue date	1980, 12

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

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

昭和55年12月15日　　於 国際電電研究所

内容梗概　単語音声認識の新しい手法として，クラスタリングによって標準パタンを作成し，この標準パタンを用いて認識対象単語を表現する単語辞書を作成し，これらとDP手法を用いて単語認識を行う方法を提案する。この方法は，DPのための計算量，および単語データを蓄積しておくメモリ量が，単語のすべてのデータを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。本稿では，この認識方法を，少数語い（空港名67単語）および大語い（都市名641単語）に適用した結果について報告する。実験の結果，標準パタン数を128～256に選べば，全フレームデータを用いる場合との認識率の差は小さく，本方法が特定話者の大語い単語音声認識に利用できる見通しを得た。

Word Recognition Using Pseudo-Phoneme Templates

Noboru SUGAMURA, Kazuo HAKODA and Sadaoki FULUI
(Musashino E.C.L., N.T.T.)

Dec. 15, 1980　　at K.D.D. Research Laboratory

Abstract　This paper discribes a new word recognition system which uses a set of pseudo-phoneme templates and a word dictionary. The pseudo-phoneme templates are automatically extracted from training words using a clustering techniques without linguistic information. The word dictionary contains symbolic spellings of these templates. A spectral distance between each frame of input speech and each pseudo-phoneme templates is calculated and stored as an element of a distance matrix. Word identification is based on the total distance obtained by summing up the distance over the input word length using distance matrix. Experimental results for 67 airport names and 641 city names recognition show the efficiency of this system in large vocabulary word recognition from the viewpoint of memory size and the amount of calculation.

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言語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標準パタンを用いて認識する方法や[1]，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[2]などがある。認識対象の単語数が増加するにしたがって，前者は標準パタンを蓄えておくメモリや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，音韻標準パタンの構成に人手をわずらわせることや，単語辞書の追加にやや問題点がある。

本稿では，両方式の中間的なものとして，言語レベルにとらわれない方法で，擬似的な音韻標準パタンを作成し，このパタンを用いて単語認識実験を試みた結果について報告する。ここで提案する単語認識方法は，大語いになるほど有効性が大きい。

2. 新しい単語音声認識法

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，どのような音素から構成されているかは，必ずしも知る必要がない。したがって標準パタンそのものも，あえて言語レベルとの対応をとる必要がない。一方単語全体のすべての情報を用いれば，それだけ標準パタンを蓄積しておくメモリが増加する。また必ずしもすべてのフレームのデータが認識するのに必要であるわけではなく，標準パタンそのものも一種の量子化をした方が，メモリを有効に利用できることになる。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音声認識法（Word Recognition System Using Strings of Phoneme-Like Templates 略して以後 SPLIT法とよぶ）について説明する。

認識システムのブロック図を図1に示し，各部について説明する。

(1) 学習サンプルによって，擬音韻標準パタン $S_1 \sim S_n$ を作成する。（図1－(b)）$S_1 \sim S_n$ はそれぞれ10～16個のスペクトルパラメータによって表現される。学習サンプルとしては，認識対象単語のすべてを用いる方法と，一部の単語のみを用いる方法が考えられる。擬音韻標準パタンの作成方法は後述する。

(2) 各単語の学習サンプルと標準パタン $S_1 \sim S_n$ とのマッチングをフレーム単位に行い，各単語を $S_1 \sim S_n$ の系列で表現する単語辞書を作成する。（図1－(d)）

(3) 未知サンプルをスペクトル分析し（図1－(a)）各フレームごとに $S_1 \sim S_n$ の標準パタンとの距離を計算して距離マトリクスを作成する。（図1－(c)）この距離マトリクスの計算は，

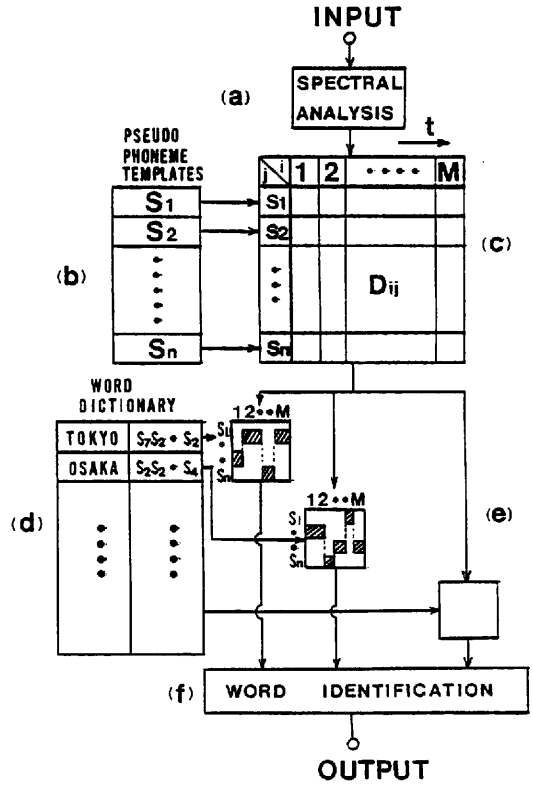


Fig.1 Block diagram of the word recognition system using strings of phoneme-like templates

認識対象語い数に関係なく未知サンプルに対して，1度行えばよい。
(4) 未知サンプルと各擬音韻系列とのマッチングをDPによって行い（図1-(e)），最小距離パタンを有する単語を認識結果として出力する。（図1-(f)）

2-2. SPLIT法の特徴

前述したようにSPLIT法によれば，DPマッチングを行う場合の距離としては，(2)で得られた各単語の標準パタン系列を参照して，(3)で計算した距離マトリクスの要素を検索して用いる。この距離マトリクスは，一つの未知サンプルに対して1回だけ計算しておけば，認識対象単語のすべてのDPマッチングに利用できる。たとえば認識対象単語数をL，未知サンプルのフレーム数をM，DPの窓幅をNw，標準パタン数をnとすると，単語全体を一つの標準パタンとして認識する方法では，1単語を認識するのに必要な距離計算回数は，約M×Nw×Lであるのに対し，SPLIT法では，M×nとなる。したがって1フレームあたりの計算量の比は，R=n/(Nw×L)となる。

つぎに標準パタンを蓄積しておくためのメモリ量の比較を行う。各単語の標準パタンの平均フレーム数を$\bar{M}$，標準パタン数をn，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，一つのパラメータ精度をN_aビットとすると，単語全体の情報を用いる方法では，$\bar{M}\times L\times N\times N_a$ビット，SPLIT法では，$n\times N\times N_a+L\times\bar{M}\times n_b$ビットのメモリを必要とする。第1項が，擬音韻標準パタン用，第2項が，単語辞書用のメモリである。

Table.1 Comparison of the two word recognition methods from the viewpoint of memory size and the amount of calculation

	B/A ↗SPLIT ↘FULL	L=500 words $\bar{M}$=50 frames n=128 templates n_b=7 bits Na=16 bits N=16 parameters Nw=15 frames
Amount of distance calculation per frame	$\frac{n}{Nw*L}$	0.017
Memory size for templates and word dictionary	$(n*N*Na+L*\bar{M}*n_b)/(L*\bar{M}*N*Na)$	0.03

以上の計算量の比，メモリ量の比の比較を表1に整理して示し，具体的な数値を用いて，それぞれの比の概算値を示している。表中の"FULL"は，単語全体を一つの標準パタンとして認識する方法を表わしている。このことから明らかなように，SPLIT法は，単語数が多くなればなるほど有効であることがわかる。

2-3 標準パタンの作成法

標準パタンの構成法としては，言語レベルにとらわれないクラスタリング手法を用いる。この方法は，従来の音韻ごとの標準パタンを用いる方法に比べ，つぎのような特徴をもっている。
(1) パタン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，人手をわずらわせる作業がない。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。
(2) 言語（国語）に依存しないで標準パタンを構成できるので，認識対象の言語（国語）に対する適応性がある。
(3) 認識対象単語（単語の種類）にかかわりなく標準パタンを構成できる。

標準パタンの作成方法は，筆者らが先に報告したパタンマッチングボコーダで用いた方法と全く同じ方法である。[3] 以下，その手順について簡単に述べる。

(i) 認識すべき単語の一部あるいは全部を学習サンプルとして，標準パタン作成用のデータとして用いる。いま，この総フレーム数をnとし各フレームのパラメータセット（パタンとよぶ）をS_1～S_nで表わす。S_1～S_nは，一般に10～16個のスペクトルパラメータから構成されるが，いま仮にこの個数を10とし，S_iを

$$S_i=(P_{1i}, P_{2i}, \cdots, P_{10i})$$

で表わす。
(ii) 各パタンS_iとすべての異なるパタンS_jとのスペクトル距離を計算し，この値があらかじめ定めておいた閾値θ_sよりも小さくなるパタン数を，各パタンごとに求める。これをm_i $i=1,2,\cdots,n$とする。
(iii) 最大のm_iを有するパタンS_iを求め，S_iからθ_s以内に含まれるすべてのパタンを用いて，各要素ごとに平均値を計算し標準パタンと

して蓄える。

(iv)(iii)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，(i)～(iii)の手順をもとのパタンがすべてなくなるまでくり返す。

以上の手順によって構成された標準パタンセットを$\overline{S_1}$～$\overline{S_\ell}$とする。ℓはθ_sの値によって1からℓ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の表現が可能となるかわりに，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，単語認識で最終的にどれくらいの標準パタンを用いるかということから決定する必要がある。

上述のクラスタリングにおいては，パタンの分布密度の高い部分から順に標準パタンを発生させることができる。またこの方法において，与えるべきパラメータはθ_sのみで，非常に簡単に標準パタンを構成できる。

3. 認識実験結果

認識実験としては，つぎの3つの条件について行った。

(1)少数語い（67単語），個人ごとの標準パタンによる実験

発声者ごとに標準パタンと各単語の標準パタン系列を作成する。標準パタン作成には全単語を用いる。

(2)少数語い（67単語），共通の標準パタンによる実験

発声者間で共通に用いる標準パタンを作成し，各単語の標準パタン系列は，発声者ごとに作成したものを用いる。

(3)大語い（641単語），個人ごとの標準パタンによる実験

発声した単語の一部を用いて標準パタンと各単語の標準パタン系列を発声者ごとに作成したものを用いる。

以上3つの実験について，以下に説明する。

3-1 少数語い，個人ごとの標準パタンによる実験

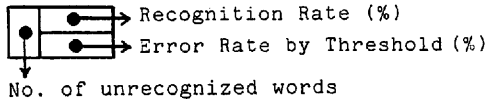
男性10名の発声した国内空港名67単語を認識対象とし，話者ごとに標準パタンと単語辞書を作成する方法で，認識実験を行った。実験条件を表2，実験結果を表3に示す。実験で用いた標準パタン数は，64，96，128の3種類であり，これらのパタンを作成するために，個人ごとに2200～3100フレーム（表3右欄）データを用いている。表3の中の誤り率というのは，マッチング距離の値にある閾値を設けて，正しい単語を棄却する誤りと，異なる単語を受理する誤りが等しくなるようにし

Table.2 Experimental condition in the case of 67 airport names of 10 male speakers

実験対象単語	空港名67単語
発声者	男性10名（各単語学習用テスト用各1回発声）
分析条件	8kHzサンプリング 30msec ハミング窓 フレーム周期 15msec
特徴パラメータ	LPCケプストラム（10次）
DP手法	両端点フリーDP[4]

Table.3 Experimental results by SPLIT method in the case of making pseudo phoneme templates and word dicionary for each speaker

Speaker	FULL		SPLIT 128		SPLIT 96		SPLIT 64		No.of frames
1	1	98.5 1.49	1	98.5 1.49	1	98.5 1.49	1	97.0 2.31	3106
2	1	98.5 1.49	1	98.5 1.49	2	97.0 1.49	2	97.0 1.49	2876
3	2	97.0 2.99	2	97.0 2.99	2	97.0 2.99	3	95.5 2.99	2722
4	2	97.0 1.56	3	95.5 1.49	3	95.5 1.49	3	95.5 1.49	2833
5	2	97.0 2.99	2	97.0 2.99	2	97.0 2.99	2	97.0 2.99	2250
6	0	100.0 1.13	0	100.0 1.49	0	100.0 1.49	0	100.0 1.49	2655
7	1	98.5 1.49	1	98.5 1.49	1	98.5 1.49	1	98.5 1.49	2182
8	2	97.0 1.49	3	95.5 1.49	3	95.5 1.49	3	95.5 1.49	2383
9	1	98.5 1.49	1	98.5 1.49	1	98.5 1.49	2	97.0 1.49	2789
10	2	97.0 2.78	3	95.5 2.96	3	95.5 2.99	3	95.5 5.97	2990
SUM. or MEAN	14	97.9 1.89	17	97.4 1.94	18	97.3 1.94	21	96.9 2.17	26786



たときの誤り率である。[5]

この結果，標準パタン数が64になると認識率は急に低下するが，128では，全フレームのパタンを用いた場合（表中"FULL"）とほぼ同等の結果が得られた。128個の標準パタンは，個人ごとの総フレーム数の約5％（話者による平均値）であり，この128個の標準パタンにθs以内の距離で含まれるパタン数の割合は，平均的に約80％である。

3-2 少数語い，共通の標準パタンによる実験

前節では，個人ごとに標準パタンと単語辞書を作成し認識実験を行った結果について述べたが，本節では話者間に共通の標準パタンを作成し，単語辞書のみを個人ごとに作成し認識実験を行った結果について述べる。

まず話者間に共通に使用する標準パタンの作成方法について説明する。

(i) 1人目の発声者のデータをもとに，2-3で説明した方法により，標準パタン候補 $\overline{S}_1 \sim \overline{S}_\ell$ を作成する。このときそれぞれのパタンに含まれるパタン数を $m_1 \sim m_\ell$ とする。

(ii) 2人目の発声者の各フレームのデータを用いて $\overline{S}_1 \sim \overline{S}_\ell$ とのスペクトルマッチングを行い，θs以内の距離で $\overline{S}_1 \sim \overline{S}_\ell$ のどれかに含まれるパタンは除去する。このパタン総数をNcとする。またどのパタンにも含まれなかったパタン数をNnとする。このとき新しく含んだパタン数だけ $m_1 \sim m_\ell$ を変化させる。これを $m'_1 \sim m'_\ell$ とする。

(iii) Nn個のパタンに対し，クラスタリング手法を適用し，標準パタン候補を作成する。これを $\overline{S}_{\ell+1} \sim \overline{S}_{\ell'}$ とする。またそれぞれのパタンに含まれるパタン数を $m_{\ell+1} \sim m_{\ell'}$ とする。

(iv) 3人目の話者に対しては，$\overline{S}_1 \sim \overline{S}_{\ell'}$ を標準パタン候補として，(ii)(iii)をくり返し，以下4人目から10人目の話者に対しても，標準パタン候補を更新しながら同様の処理を行う。

(v) 最終的には，含まれるパタン数が多いパタンから順に，認識に必要な数だけ選択する。

以上の手順をわかりやすくするために，図2に2人の話者，2次元データを例にとり示す。話者Aのデータをもとに，A1～A6の6つのクラスターに分け，それぞれのクラスターに含まれるパタンから，標準パタン候補を求める。このときパラメータの平均化を行うので，それぞれの標準パタン候補にθs以内の距離で含まれる領域は，実線で描いた円から，破線で描いた円に変化する。つぎに2人目の話者のデータを用いて，新しい標準パタン候補を作成する。まず破線で描いた円内に含まれるパタンを除去する。このとき，もし2人目の話者のデータが，ある破線円内に含まれれば，そのクラスタに含まれるパタン数を，増加させる。たとえば，A1のクラスターは，6から7にパタン数を変化させる。つぎに含まれなかった2人目の話者データに対し，クラスタリングを行いB1～B4のクラスターを作成し，標準パタン候補をつくる。最終的に，4つの標準パタンを選択するとすれば A1, A2, A4, B1が選択される。

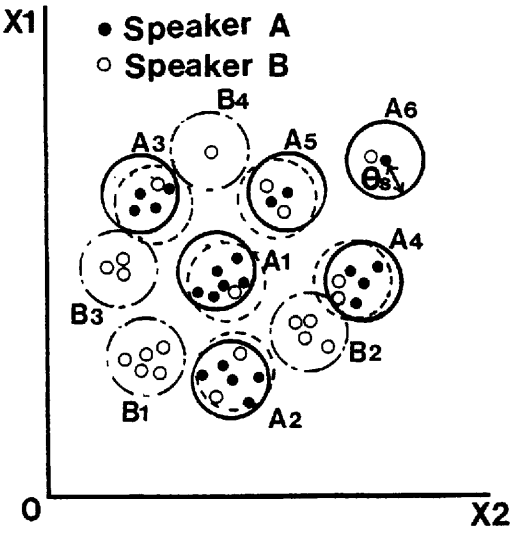


Fig.2 Clustering example in the case of two speakers and two dimentinal space

つぎにこのように話者が増えるごとに標準パタンを追加していく場合に，個人ごとに新しい標準パタン数がいくらになるか，あるいは，標準パタン候補に含まれるパタン数Ncが，全パタン数N（=Nc+Nn）に対してどれくらいの割合であるかということを図3に示す。

直線Cの高さは，個人ごとに新しく生じた標準パタン候補を表わしている。直線Bは，マッチングすべき標準パタン候補数を示しており，直線Aは，この標準パタン候補のどれかにθs以内の距離で含まれるパタン数Ncの全パタン

数に対する割合を個人ごとに示したものである。たとえば，話者1に対しては，マッチングすべき標準パタン候補数は0で，約400程度の話者1の標準パタン候補が作成される。このとき話者1のパタンは，マッチングすべき標準パタン候補がないからNc/Nは0%となる。つぎに話者2については話者1で生じた400程度の標準パタン候補とマッチングを行う。このときこれらのパタンに含まれるパタン数の割合は，40%弱となっている。含まれなかったパタンに対しては，クラスタリングの結果，約300程度の新しい標準パタン候補が生じている。同様にして3人目～10人目までの結果を示している。ただし図にも表示しているように，4人目からのデータに対するマッチングすべき標準パタン候補数としては1024で上限を押えており，各段階で発生する標準パタン候補に対して含まれるパタン数でソーティングし，1025位以下のパタンは切り捨てる方法をとっている。

このような方法で，標準パタンとして128個を選択し，個人ごとにこのパタンを用いて単語辞書を作成し実験を行った結果を，前節の結果と比較して表4に示す。この結果，話者間に共通のパタンを用いることにより，全フレームデータを用いる場合に比べ，約1%認識率の低下がみられた。この場合の認識率は，個人ごとに標準パタンを作成し，64個のパタンを選択して認識実験を行った場合にほぼ対応する。標準パタンを共通に用いて128個とした場合，この値は，10人の総フレーム数の0.5%であり，これにθs以内のスペクトル距離で含まれるパタン数の割合は，話者間の平均で約25%である。

3-3 大語い，個人ごとの標準パタンによる実験

2-2で述べたように，SPLIT法は，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。本節では，この方法が，大語い単語認識に適用できるか，あるいは適用した場合に，全フレームのデータを用いる場合に比べ，どの程度認識率が低下するのかなどを検討することを目的として，男性1名の発声者による，日本の都市名641単語の認識実験を行った。

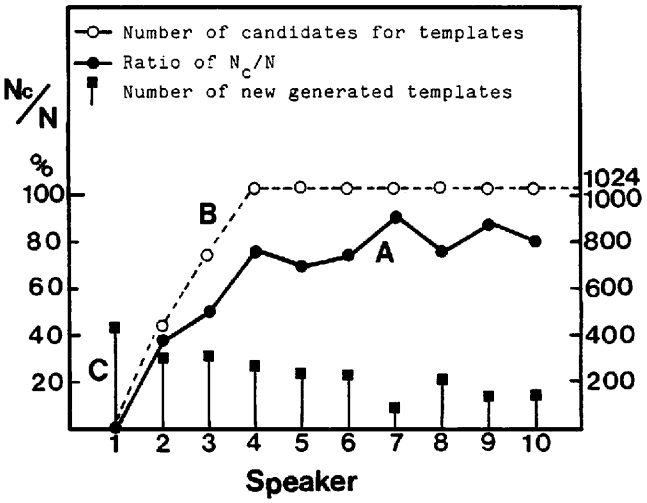


Fig.3 Number of new generated templates and a ratio Nc/N

Table.4 Experimental results by SPLIT method in the case of using common pseudo-phoneme templates

Speaker	all frames		Pseudo-Phoneme Templates common 128		each speaker 128		each speaker 64	
1	1	98.5 1.49	3	95.5 1.49	1	98.5 1.49	2	97.0 2.31
2	1	98.5 1.49	1	98.5 1.49	1	98.5 1.49	2	97.0 1.49
3	2	97.0 2.99	1	98.5 2.99	2	97.0 2.99	3	95.5 2.99
4	2	97.0 1.56	3	95.5 1.56	3	95.5 1.49	3	95.5 1.49
5	2	97.0 2.99	2	97.0 2.99	2	97.0 2.99	2	97.0 2.99
6	0	100.0 1.13	1	98.5 1.49	0	100.0 1.49	0	100.0 1.49
7	1	98.5 1.49	1	98.5 1.49	1	98.5 1.49	1	98.5 1.49
8	2	97.0 1.49	3	95.5 1.49	3	95.5 1.49	3	95.5 1.49
9	1	98.5 1.49	2	97.0 1.49	1	98.5 1.49	2	97.0 1.49
10	2	97.0 2.78	3	95.5 3.01	3	95.5 2.96	3	95.5 5.97
sum. mean	14	97.9 1.89	20	97.0 1.95	17	97.4 1.94	21	96.9 2.17

この実験においてまず問題となるのは，641単語（16msec分析で約30,000フレーム）の全部を用いて標準パタンを構成しようとすれば，膨大な演算量となることである。そこでここでは，641単語の中から任意に選んだ64の単語を用いて行った。表5に示す実験条件で，1回目の発声データをもとに標準パタンと単語辞書をつくり認識実験を行った結果（表中1→2）と，2回目の発声データをもとに同様のことを行った結果（表中2→1）を表6に示す。また全フレームデータを用いた場合（表中"FULL"）の結果も示す。距離計算としては，ケプストラム（CEPと略す）とWLR[6]の2種類を用い，標準パタンとしては128個と256個を用いた。これらの標準パタンの作成には，約3,000フレームのデータを使用した。2-2で説明した方法によって計算したメモリ量の比は，パタン数128の場合0.03，256の場合は0.04となる。

表6の結果から，標準パタン数を256にした場合，ケプストラム距離を用いれば，全フレームのデータを用いた場合との差は，ほとんど生じないことが明らかとなった。またパタン数128にした場合との差も小さい。スペクトル距離をWLRにした場合は，CEPを用いた場合に比べ，認識率は，わずかに高くなるが，全フレームのデータを用いた場合ほどの改善度は得られなかった。

標準パタン作成のための単語数を64個にしても，認識率の低下は，ほとんどないことも明らかとなった。また標準パタン作成に用いた単語の認識率が，他に比べ特に高いということもなく，標準パタンの作成に，すべてのデータを用いる必要がないことも明らかとなった。

つぎに誤認識の単語を，CEPで標準パタン256の場合と全フレームデータを用いた場合（いずれも1→2）との比較を表7に示す。"C"は正しく認識されたこと，"*"はDP窓の範囲に入らないこと，"?"は正解が5位以内に入っていないことを示している。また誤認識の理由を右欄にA～Eの5つに分類して示す。Aは切り出しエラーによるもの，B母音のエラー，C語頭の子音のエラー，DはCに含まれない子音のエラー，Eはその他のエラーである。

この表からわかるように，誤認識の単語は，両方法で必ずしも一致していない。騒音下で録音した音声を用いたため，誤りの中では，切り出しミスが最も多く，ついで文頭の子音のエラーによるものが多い。このような音韻的にきわめて近いものは，DPマッチングのように，ス

Table.5 Experimental condition in the case of 641 city names of one speaker

実験対象単語	都市名 641単語
発声者	男性1名，2回発声
分析条件	8kHzサンプリング 32msec ハミング窓 フレーム周期 16msec
特徴パラメータ	LPCケプストラム WLR } 16次
DP手法	両端点フリーDP[7]

Table.6 Experimental results by SPLIT method in 641 city names recognition

Distance Measure		CEP					WLR				
Matching Technique		FULL	SPLIT				FULL	SPLIT			
No.of Templates		all frames	128		256		all frames	128		256	
Matching Direction		mean	1→2	2→1	1→2	2→1	mean	1→2	2→1	1→2	2→1
Ranking	=1	95.0	94.4	94.5	95.3	94.2	95.9	95.0	94.5	95.5	94.5
			94.5		94.8			94.8		95.0	
	≦2	97.9	97.4	97.5	97.5	97.0	97.9	97.5	97.2	97.5	97.4
			97.5		97.3			97.4		97.5	
	≦5	98.6	98.3	98.4	98.4	98.4	98.6	98.3	98.6	98.3	98.6
			98.4		98.4			98.5		98.5	
	≦10	98.7	98.4	98.6	98.6	98.9	98.7	98.4	98.8	98.6	98.8
			98.5		98.8			98.6		98.7	
Error Rate %		1.13	1.51	1.36	1.33	1.39	1.11	1.65	1.24	1.55	1.24
			1.44		1.36			1.44		1.40	

ペクトルマッチング量を総合的に評価する方法では，その認識にも限界があり別のアプローチを考える必要がある[8]。

4 あとがき

本稿で説明したSPLIT法は，標準パタンの作成には何ら言語的な情報を必要とせず，人手をわずらわせることもない。このような簡単な方法で作成した比較的少数の標準パタンを用いて認識を行っても，全フレームデータを用いて認識を行う場合に比べて，認識率の低下はわずかであり，SPLIT法が，特定話者の大語い単語音声認識に利用できる見通しが得られた。

Table.7 Error list for 641 city names recognition in the case of 256 pseudo-phoneme templates and cepstrum distance measure

Distance Measure	C E P				Reason for unrecognized				
No. of Templates	all frames		256		A	B	C	D	E
W O R D	rank		rank						
MON BETSU	*		*		●				
MUTSU	*		*		●				
ESASHI	*		*		●				
GYO:DA	SHIMODA	2	SHIMODA	2					●
AGEO	TAKEO	2	C	1				●	
KUKI	1-FUJI 2-TOKI	3	C	1				●	
KIMIZU	SHIMIZU	2	C	1			●		
O:ME	KO:BE	2	KO:BE	2				●	
SHIRONE	HIKONE	2	1-HIKONE 2-HIROSAKI	3				●	
UOZU	O:ZU	2	1-O:ZU 2-O:TSU	3		●			
KOMATSU	?		?		●				
MATSUTO	C	1	MATSUDO	2				●	
UEDA	1-IKEDA 2-UENO	3	C	1	●				
SUZAKA	C	1	SHIBATA	2		●			
CHINO	HINO	2	HINO	2			●		
SAKU	*		*		●				
O:GAKI	C	1	O:MACHI	2				●	
SHIMIZU	C	1	KIMIZU	2			●		
FUJI	C	1	UJI	2			●		
AN-JO	SAN-JO	2	SAN-JO	2			●		
O:BU	?		?		●				
CHITA	?		?		●				
CHIRYU:	1-KIRYU: 2-CHICHIBU	3	1-KIRYU: 2-CHICHIBU	3	●				
TSU	SUZU	2	SUZU	2					●
MINOO	MINO	2	MINO	2					●
KATANO	*		*		●				
NISHINOMIYA	FUJINOMIYA	2	C	1		●			
SANDA	HANDA	2	C	1			●		
KURE	UBE	2	UBE	2				●	
HO:FU	1-KO:FU 2-O:BU	3	1-KO:FU 2-O:BU 3-CHO:FU 4-KO:NOSU	5			●		
TOKUSHIMA	FUKUSHIMA	2	FUKUSHIMA	2		●			
SUKUMO	*		*		●				
SAGA	KAGA	2	1-KAGA 2-TAMA	3			●		
TOSU	*		*		●				
TAKU	C	1	SAKU	2			●		
HITOYOSHI	KURAYOSHI	2	1-MIYOSHI 2-YU:KI 3-KURAYOSHI	4	●				
ARAO	NANAO	2	C	1					●
KUSHIMA	TSUKUSHIMA	2	C	1			●		

C-correct　*-out of adjustment window　?-more than 5th ranking
A-detection error of speech period　**B**-vowel error in the first mora
C-beginning consonant error　**D**-consonant error exc.C　**E**-others

謝辞

日頃御指導頂く野田基礎研究部長，小池第四研究室長に深謝します。また大語い認識に関して，いろいろ御協力，助言を頂いた第四研究室鹿野調査員，杉山社員に感謝します。

参考文献

(1) 進工，千葉：音学誌 27，9，P.483 (1971)
(2) 好田，楠木，斎藤：信学論 55-D 3 P.186 (1972)
(3) 管村，板倉：音声研究会資料 S79-08 (1979)
(4) S.Furui and A.E.Rosenberg：ICASSP'80 P.1060
(5) 古井，板倉，斎藤：信学論 55-A 10 P.549 (1972)
(6) 杉山，鹿野：音声研究会資料 S80-03 (1980)
(7) 鹿野：音響学会講論集 1-1-10 (1980-10)
(8) 古井：音響学会講論集 1-1-18 (1980-10)