論文 / 著書情報
Article / Book Information

論題(和文)	大語彙単語音声認識のためのスペクトル動特性を用いた予備選択法
Title(English)	
著者(和文)	古井貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会1986年秋季講演論文集, , , pp. 33-34
Citation(English)	, , , pp. 33-34
発行日 / Pub. date	1986, 10

1—3—17　大語彙単語音声認識のための スペクトル動特性を用いた予備選択法*

古井 貞熙（NTT電気通信研究所）

1．まえがき

不特定話者の単語音声を認識する方法に、マルチテンプレート法、即ち各単語について複数の標準パターンを用意する方法がある。この際標準パターンの記憶容量と認識のための演算処理量が増大する問題があるが、これを大幅に減らす有効な方法にＳＰＬＩＴ法がある[1]。ＳＰＬＩＴ法では距離計算量が語彙数に依存しないため、語彙が大きくなるほどそのメリットが発揮されるが、極めて大語彙になるとＤＴＷ（ＤＰ演算）も無視できなくなるので、予備選択によってＤＴＷを行なう単語候補を絞る必要が生ずる。予備選択の方法にはすでに種々の試みがあるが、最近試みられている方法に、ベクトル量子化の歪量を用いるものがある[2]。これは各単語について、話者による変動を考慮したスペクトル包絡の符号帳を用意しておき、各符号帳で量子化したときの歪が比較的小さい単語を候補として選択する方法である。ただしこの符号帳は時間情報を持たないため不適切な対応づけが起り、語彙数が大きくなると性能が低下する。符号帳に時間情報を持たせる方法として、単語中の相対位置情報を確率的に各符号に与える方法[3]もあるが、音声区間の終端が検出されないと処理が開始できず、しかも音声区間検出の精度の影響を受けやすい問題がある。また単語別に符号帳を用意するため、距離計算量が語彙数に比例して増大する。ここでは、これらの問題点を解決する方法を検討した。

2．予備選択の方法

図１に示すように音声波をケプストラム分析したのち、7～11フレーム(56～88 ms)の区間を8ms毎にシフトし、動的情報として各ケプストラム係数及び対数エネルギー時系列の線形回帰係数を抽出する[4]。ケプストラムと回帰係数をセットにして、各単語毎に複数話者の全フレームデータをクラスタ化し、各単語のWB(Word-based)－符号帳を作成する（ＷＢ－符号帳の大きさは32ベクトルでほぼ十分である）。次に、全単語の符号帳をまとめて再度クラスタ化し、U(Universal)－符号帳を作る。各単語のＷＢ－符号帳の各要素に最も近いＵ－符号帳の要素を選び、これを各単語の符号帳とする。入力音声に対しては、各単語の符号帳でベクトル量子化を行ない、歪の比較的小さい単語候補を選択して、ＤＴＷ処理へ送る。

この方法では、各単語の符号帳がＵ－符号帳の部分集合になっているので、ＳＰＬＩＴ法と同様に、距離計算は入力音声とＵ－符号帳の間でのみ行なって、距離行列の形で蓄え、後の処理（予備選択及びＤＴＷ）ではこれを検索すればよい。またＤＴＷは各単語について複数の標準パターンと行なう必要があるが、予備

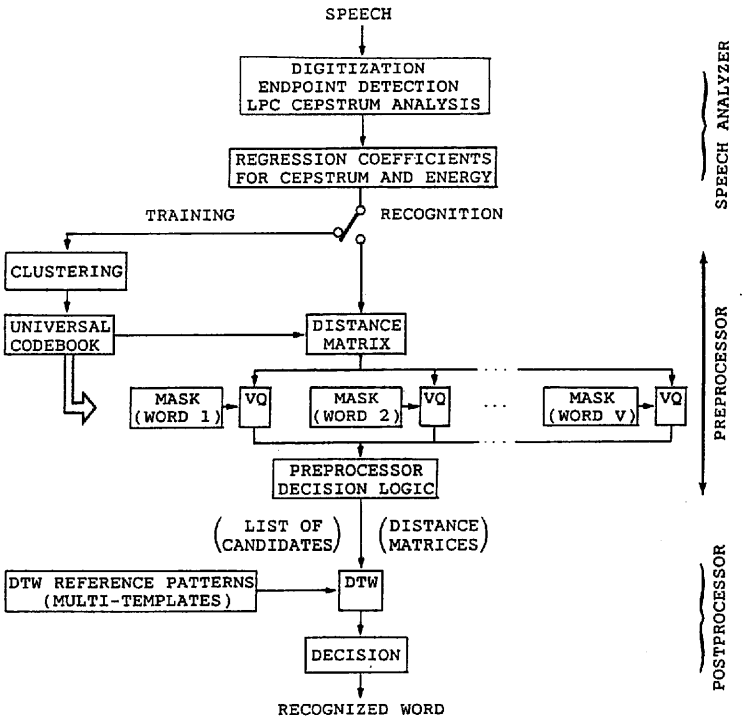

Fig. 1 - Block diagram of isolated word recognizer incorporating a VQ preprocessor and a DTW-based postprocessor.

*Word Pre-selection Method Based on Dynamic Spectral Features for Large Vocabulary Word Recognition. By Sadaoki Furui (NTT Electrical Communications Laboratories)

選択用の符号帳は各単語に1つですむので、全体としての処理量は1/10程度に小さくすることができる。

3. 実験結果

都市名100単語を対象とし、男性話者の声質を比較的代表していると推定される4名の話者を学習サンプルとして、これと異なる男性20名が各単語を1回ずつ発声した2000サンプルについて認識実験を行なった。DTWはSPLIT法で行ない、各単語について4標準パターンを用いた。

回帰係数を抽出する長さを11フレーム（最適値）としたときの、特徴パラメータによる予備選択効果の違いを図2に示す。図中，は回帰係数を示し、横軸は量子化歪の小さい方から第何位まで選択するかを示す。この実験では共通化する前のVB－符号帳を用いている。ケプストラムのみの場合に比べて、エネルギーの回帰係数、さらにケプストラムの回帰係数を組み合わせて用いることにより、大幅に誤り率が低下することがわかる。

符号帳の共通化を行なう前と後の比較を図3に示す。この実験では、DTWでの最適条件[4]との整合性を考慮して、回帰係数を抽出する長さは7フレームとした（11フレームの場合との結果の違いは小さい）。U－符号帳の大きさは、512と1024の2種類に変えた。歪量に対するしきい値を$\theta = \theta_0 + \mathrm{Min}(d)$で与え、これによって選択を行なった。ここでMin(d)は各入力音声についての全候補単語に関する最小の歪量で、図には定数θ_0を変えたときの結果が示してある。1024の要素を用いれば、共通化を行なう前に近い性能を得ることができることがわかる。

DTWのみの場合、予備選択のみの場合、及び両者の組み合わせの場合の結果を表1に示す。組み合わせの場合は、第10位まで又は適切なしきい値で予備選択を行なった。全体としての誤り率をほとんど増加させることなく、DTWを行なう候補数を1/10以下に絞ることができる。しきい値で予備選択を行なった時、2000サンプル中496サンプルは、一意に正しい単語に候補が絞られる。

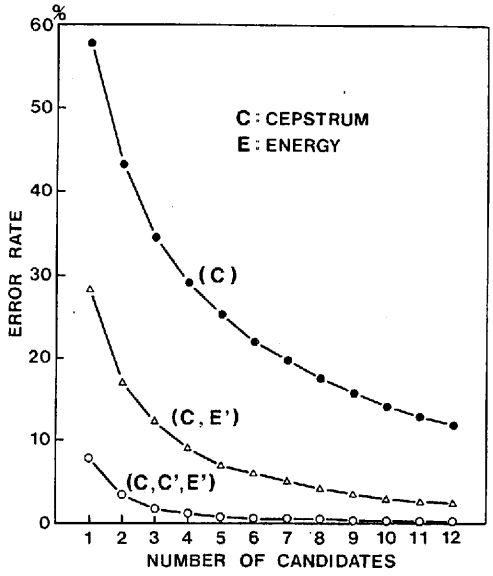

Fig. 2 - Comparison of the preprocessor performances for three feature parameter conditions.

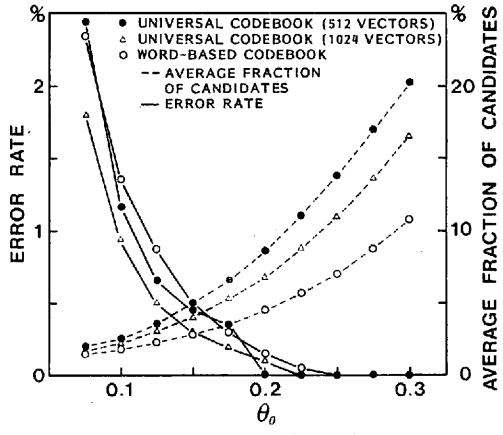

Fig. 3 - Comparison between the performances using universal codebook and word-based codebook, as a function of the preprocessor threshold.

4. むすび

不特定話者大語彙単語音声認識のための、スペクトル動特性を用いた候補単語予備選択方法を提案し、高い有効性を確認した。今後実際に大語彙を対象とした評価実験を進めて行きたい。

Table 1 - Experimental results for the word recognizer incorporating a VQ preprocessor and a DTW-based postprocessor.

		OVERALL ERROR RATE (%) PRESELECTION ALONE	OVERALL ERROR RATE (%) PRESELECTION & DTW	NUMBER OF CANDIDATES FOR DTW
DTW ALONE		-	2.0	100
WORD-BASED CODEBOOK	PRESELECTION BY TOP	7.9	-	-
	TOP TEN	0.2	2.1	10
	$<\theta$ (θ_0=.2)	0.2	2.1	4.5
UNIVERSAL CODEBOOK	PRESELECTION BY TOP	10.9	-	-
	TOP TEN	0.4	2.3	10
	$<\theta$ (θ_0=.2)	0.1	2.0	6.8

[文献] [1]菅村他：信学論，J65-D，pp.1041-1048(1982) [2]K-C.Pan,et al.：IEEE Trans.ASSP-33,pp.546-560(1985) [3]A.F.Berch,et al.：AT&T Tech.Journ.,64,pp.1047-1063(1985) [4]S.Furui：IEEE Trans.ASSP-34,pp.52-59(1986)