

論文 / 著書情報
Article / Book Information

論題(和文)	ケプストラムの動特性を用いた単語音声認識
Title(English)	
著者(和文)	古井貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	昭和60年度電子通信学会総合全国大会, , , pp. 6-191
Citation(English)	, , , pp. 6-191
発行日 / Pub. date	1985,
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) 1985 Institute of Electronics, Information and Communication Engineers.

1607

ケプストラムの動特性を用いた単語音声認識

Isolated Word Recognition Using Dynamic Features of Cepstrum Coefficients

古井 貞熙

Sadaoki FURUI

日本電信電話公社武蔵野電気通信研究所

Musashino Electrical Communication Laboratory, N.T.T.

1. はしがき 音声知覚において重要な役割を果たしているスペクトルの動的特徴を音声認識に用いる試みとして、筆者らは以前に話者認識において、ケプストラム時系列の小区間毎の1~2次の多項式展開係数を元の時系列と組み合わせて用いる方法の有効性を示した[2]。ここではこの方法を改良して、不特定話者の単語音声認識に適用した試みについて述べる。

2. 単語音声認識系の構成

2.1. 分析部 認識系の構成を図1に示す。音声波をまず8kHzで標本化し、パワーをもとに雑音区間と音声区間の分離を行う。後に端点フリーのDPマッチングを行うので、音声区間の切り出しは必ずしも厳密に行う必要はなく、むしろ両端に雑音区間をつけて切り出しておく。次にフレーム周期8ms、フレーム長32msでLPC分析を行い、1~10次のケプストラムと対数パワーの時系列を得る。この時系列から、フレーム周期毎に7フレームの区間を切り出し、この区間における各時系列の1次の多項式展開係数、即ち線形回帰係数 $a_i(k)$ を計算する。これとケプストラムの時系列 $x_i(k)$ をあわせて、特徴パラメータ時系列とする。認識部での処理量を減らすため、フレームの平滑化と間引きを行って16ms毎の時系列に変換する。

2.3. 認識部 標準パターン(R)のkフレームと入力音声(I)のlフレームとの距離 $D(k, l)$ を次のように定義して、Staggered Array法[3]によるDPマッチングを行い、単語を決定する。重み係数 w_j の値は、各特徴パラメータの有効性に従って、予備実験によりあらかじめ決定しておく。

$$D(k, l) = \left\{ w_1 \sum_{i=1}^{10} (x_i^R(k) - x_i^I(l))^2 + w_2 (a_0^R(k) - a_0^I(l))^2 + w_3 \sum_{i=1}^{10} (a_i^R(k) - a_i^I(l))^2 \right\} / \left(\sum_{j=1}^3 w_j \right) \quad (1)$$

3. 認識実験 計算機室・室内騒音約70dB(A)で発声された都市名100単語音声を用いて、男性4名の全単語音声をマルチテンプレートとして蓄え、これ以外の男性話者20名の音声(2000サンプル)の認識を行った。用いるパラメータの組み合わせを変えたときの誤認識率の変化を図2に示す。ケプストラムの回帰係数はケプストラムと同程度の認識能力を有し、両者を組み合わせると誤認識率が1/2に低下する。パワーの回帰係数よりもケプストラムの回帰係数の方が効果が大きい。ケプストラムと、ケプストラムとパワーの回帰係数をすべて組み合わせたとときの誤認識率は、ケプストラムのみの場合の2/5であり、誤りの極端に多い話者がなくなり、聴覚的類似性の少ない単語間の誤りが大幅に減少している。

4. おわりに スペクトルの静的特徴と動的特徴を組み合わせた単語音声認識系を提案し、高い有効性を確認した。動的特徴には、伝送系の影響を受けにくい長所もある[2]。今後は本方法をSPLIT法[4,5]に適用し、多数話者の音声を用いた評価を行う計画である。

[謝辞] 種々のアドバイスを頂いた第四研究室・室長、好田調査役、DPのプログラムを提供して頂いた鹿野調査役他、研究室の諸氏に感謝する。

◆ [1] 古井 貞熙: 音響学会秋季講論, 1-1-12 (昭59-10)

文 [2] Furui, S.: IEEE Trans. ASSP-29, 2, pp. 254-272 (1981)

献 [3] 鹿野、相川: 音響学会音声研資, S82-15 (昭57)

◆ [4] 管村、古井: 信学論, J65-D, 8, pp. 1041-1048 (昭57) [5] 管村、鹿野、好田: 信学論, J67-D, 10, pp. 1210-1217 (昭59)

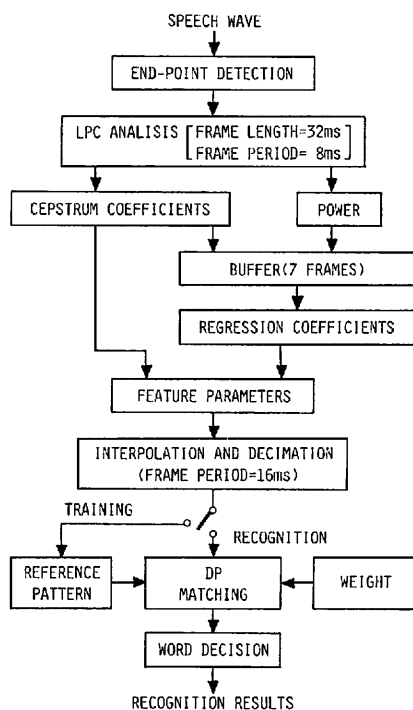


Fig. 1-Block diagram of the isolated word recognition system using dynamic features of spectrum.

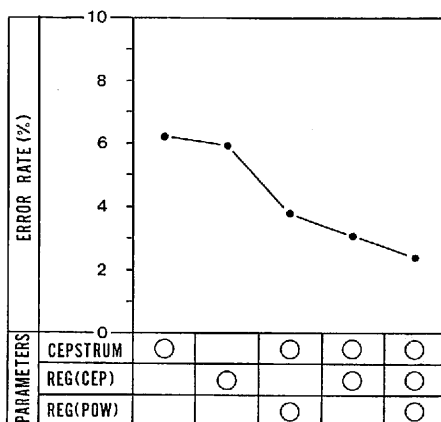


Fig. 2-Speaker independent word recognition results for five conditions of feature parameter combinations.

○ indicates the parameter set used in the recognition.