

擬似零点ケプストラム係数を音素認識の特徴量として併用する場合の効果

6N-1

西本 修 渡辺 淳二 深林 太計志

静岡大学工学部

1. はじめに

音声認識の特徴量としてLPCケプストラム係数がよく用いられている。極零型分析により求まる零点ケプストラム係数もLPCケプストラム等と併用すると有用である[1]。しかしながら、極零型分析により零点ケプストラム係数を求めるための計算量は非常に多い。計算量が少なく、容易に求めることができる擬似零点ケプストラム係数の音素認識における効果について報告する。

2. 擬似零点ケプストラム係数

12kHz サンプリングの音声波形を36程度の大きい分析次数(p')で線形予測分析すると、スペクトルのエネルギーの大きい山の部分だけでなく、近似の程度はよくないが、エネルギーの小さい谷の部分も近似するようになる。このような大きい分析次数(p')と18程度の分析次数(p)より求まる線形予測係数からそれぞれのケプストラム係数を求め、次式で表される $C_z(p', p, B)$ をここでは擬似零点ケプストラム係数と呼ぶことにする。

$$C_z(p', p, B) = C(p') \cdot C(p, B), \quad (1)$$

$$i=1, 2, \dots, p \cdots p'$$

ここで、 $C(p')$ は大きい分析次数 p' で求まる線形予測係数からのケプストラム係数。 $C(p, B)$ は分析次数 p で求まる線形予測係数 $a_i(p)$ を次式で表されるバンド幅拡大操作を行ったあとの係数からのケプストラム係数である。

Effectiveness of Combined Use of Pseudo-zero Cepstrum Coefficients in the Phoneme Recognition.
Osamu Nishimoto, Junji Watanabe and Takeshi Fukabayashi
Faculty of Engineering, Shizuoka University,
Hamamatsu, 432-8561, Japan

$$a_i(p, B) = a_i(p) \exp(-\pi B T), \quad (2)$$

$$i=1, 2, \dots, p$$

ここで T はサンプリング周期である。

3. 音声資料と特徴ベクトル、認識実験

音声資料として東北大松下単語音声データベースより男性話者30名が発声した、212単語集合の音声データベース(24kHz サンプリング) [2]を用いた。

認識対象音素は/w, y, r, z, b, d, g, m, n, η/である。これらの音声資料を後続母音へのわたりを基準として(有声破裂音は破裂時点を基準とした)前後2500点を切り出し、12kHzにダウンサンプリングを行い、実験用の資料とした。音声資料の総数は9797である。

特徴ベクトルはシフト幅10msの連続した7フレーム分の基本特徴ベクトルと第1フレームで正規化した第2~7フレームの対数パワーから構成されたものを用いた。ここで基本特徴ベクトルとはパワー成分を除いた1フレームから得られる特徴ベクトルのことをいう。認識実験に用いた基本特徴ベクトルの成分と次元は次のとおりである。

全極：分析次数 p が18のときのLPCケプストラム係数18次元を成分とする。

極零・零：極分析次数18、零点分析次数18のときの極零型分析からの極ケプストラム係数と零点ケプストラム係数のそれぞれの次数の和から得られる極零点ケプストラム係数18次元と零点ケプストラム係数18次元を成分とする。

零：極零型分析からの零点ケプストラム係数18次元を成分とする。

擬極零・擬零(B): 大きい分析次数 p が 36 のときの LPC ケプストラム係数 18 次元と式(1)の $C_z(36, 18, B)$ 18 次元を成分とする。

擬零(B, q): 式(1)の $C_z(36, 18, B)$ q 次元を成分とする。

なお、零と擬零(B, q)を基本特徴ベクトルとする特徴ベクトルはパワー成分を含まない特徴ベクトルを用いて実験した。

実験方法は 25 人を学習用として、それから半別分析に基づく変換のための変換行列とテンプレートを求め、それを基に他の 5 人分を認識する実験を 6 回繰り返す方法、5 人取って置き法をとった。認識結果はこれらの平均の認識率で示す。

4. 認識結果

基本特徴ベクトルを変えた場合の認識結果を図-1 に示す。図-1 から擬極零・擬零(0)は従来の極零・零より 3.4%ほど認識率が落ちるものの全極より 0.8%良い事が

わかる。

5. むすび

求めるのに計算量の少ない擬似零点ケプストラム係数の有効性を検討した。認識率は従来の極零・零よりおちるものの全極より良い結果が得た。さらに認識率を上げるための今後の課題として、 B の値の適当な設定の検討、擬零(0,18)より、擬零(0,36)の方が認識率が良いことを考慮して、擬零(0,36)の利用方法の検討等がある。

<参考文献>

- [1] 深林, 柴田, 杉浦(聡), 渡辺 : 音響学会誌, 53, pp.675-682(1997).
- [2] 牧野, 二矢田, 真船, 城戸 : 音響学会誌, 48, pp.899-905(1992).

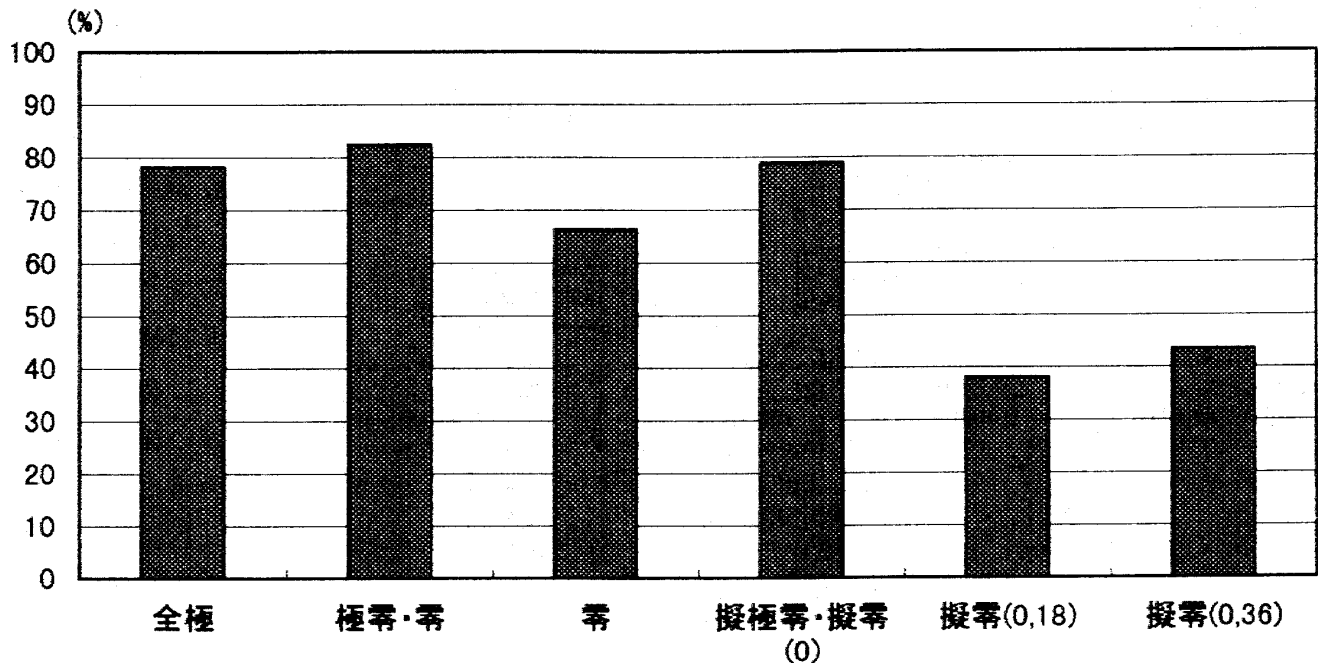


図-1 基本特徴ベクトルの構成法による認識結果