ケプストラム分布を考慮した耐雑音音声認識*

岩丸俊彦, 秋田昌憲　緑川洋一　(大分大)

1. はじめに

雑音環境下では、音声スペクトルに雑音が付加されることで、低レベル部のレベル上昇が起こりスペクトル包絡の谷の部分が埋もれてしまう。これにより、認識率の低下が起こってしまう。この場合しきい値を決め、谷付けを行う方法もあるが(1)間違った谷付けを行い誤認識する場合がある。また作業が煩雑になる。谷が埋もれることによりスペクトルが平坦になると、データ全体でのケプストラム係数の分布も小さくなることが推察される。本報告ではこのことを利用してケプストラム係数への重み付けという方法で認識実験を行う。ここでは、まずケプストラム係数の分布を雑音環境別に求め、ケプストラムの最適重み関数を推定し音声認識に利用する。

2. 雑音環境音声のケプストラム分布

ここでは、男性話者 8 名が 10 数字音声を 3 回発声した音声とそれに 2 種類の雑音(pink, automobile)を SN 比 0dB、10dB で加えたデータを用いる。これらを改良ケプストラム法(2)により、繰り返し 4 回、加速係数 1.0、次数 25、フレーム周期 5ms で分析したケプストラム分布を次数ごとに示す。単語区間については Fig1~5、データ全区間については Fig6~10 にそれぞれ次数 1 のケプストラム分布の例を示す。各グラフで横軸はケプストラム値、縦軸はその値のフレーム数を表す。ただし、ここで単語区間は無雑音の音声波形の視察で求められている。

これらの図より無雑音の時と各ノイズ別でのケプストラム値の分布を比較すると無雑音のケプストラム分布より分布の幅が狭くなっていることがわかる。

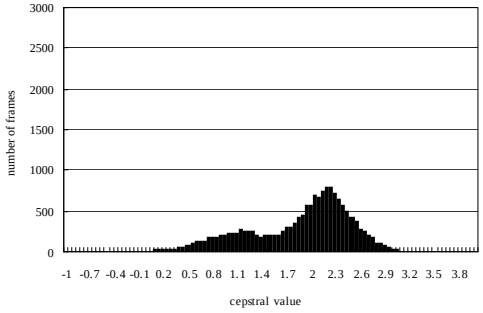

Fig1 Distribution of cepstral values(clean)
Cepstral order1　The word section

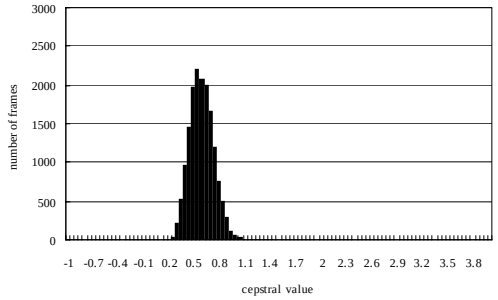

Fig2 Distribution of cepstral values(pink0dB)
Cepstral order1　The word section

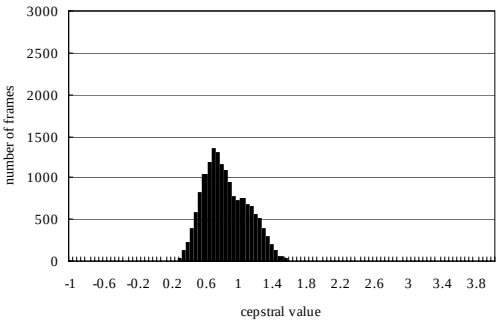

Fig3 Distribution of cepstral values(pink10dB)
Cepstral order1　The word section

---

* Speech Recognition under the noisy Enviromment considering the distribution of the cepstral values.
by IWAMARU toshihiko and AKITA masanori MIDORIKAWA yoichi(oita university)

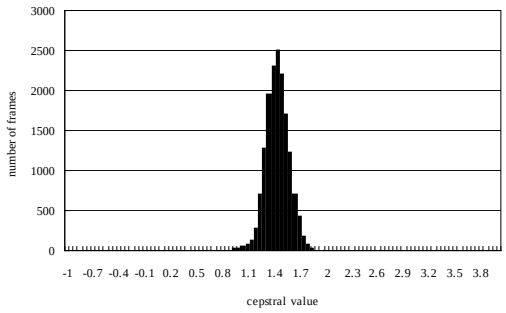

Fig4 Distribution of cepstral values(auto0dB)
Cepstral order1　The word section

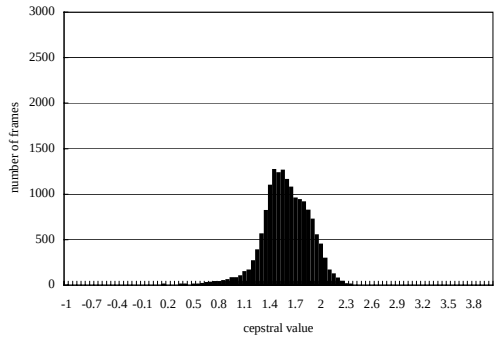

Fig5 Distribution of cepstral values(auto10dB)
Cepstral order1　The word section

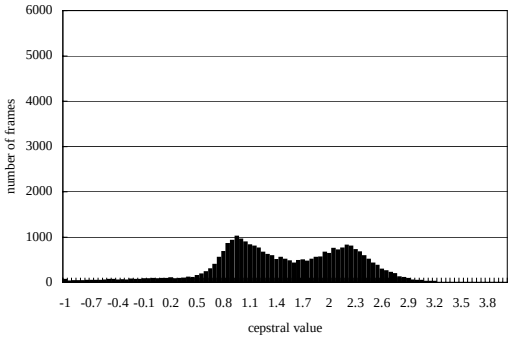

Fig6 Distribution of cepstral values(clean)
Cepstral order1　The entire section

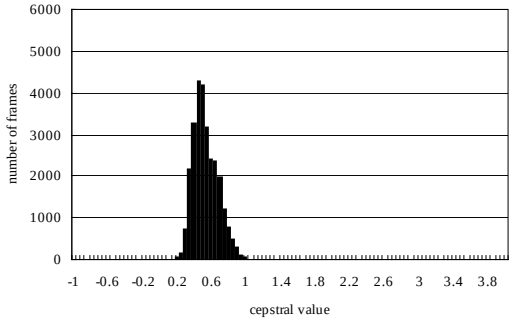

Fig7 Distribution of cepstral values(pink0dB)
Cepstral order1　The entire section

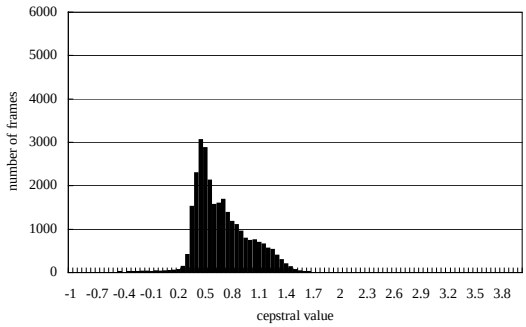

Fig8 Distribution of cepstral values(pink10dB)
Cepstral order1　The entire section

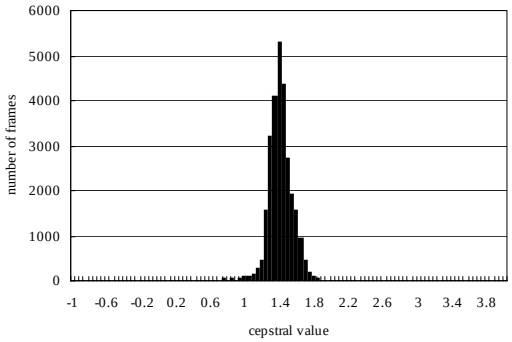

Fig9 distribution of cepstral values(auto0dB)
Cepstral order1　The entire section

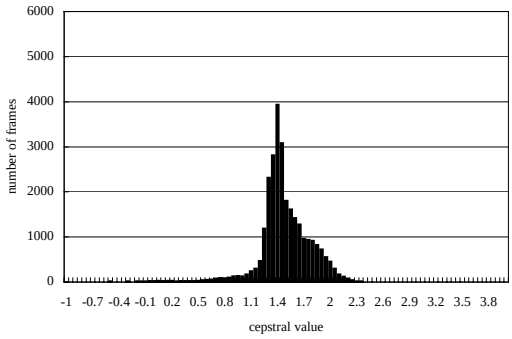

Fig10 distribution of cepstral values(auto10dB)
Cepstral order1　The entire section

全区間と単語区間でのケプストラム分布を比較しても単語区間の方が狭くなっていることが分かる。またケプストラムの平均値はノイズの SN 比が大きくなるほど小さくなり、また二種類のノイズ間では pink の方が小さくなっている。これは背景ノイズの性質そのものを表していると考えられる。

2 次以降のケプストラム値の分布についてはここでは示していないが、次数の値によって異なることもあるが分布の幅は 1 次の分布と同様の傾向で小さくなっている。また SN

比が低い時、分布の幅が小さくなり標準偏差の値も小さくなっている。

これらのケプストラム分布の平均値を SN 比別に比較してみると、無雑音のケプストラムに対して大体 pink noise0 dB では 0.23、pink noise10 dB では 0.44、automobile0 dB では 0.36、automobile10 dB では 0.57 倍となっている。

そこで、無雑音の時のデータを標準パターンとし、雑音環境音声を照合する場合、無雑音のケプストラム分布に近づけるようにするために各ノイズのケプストラム値全体に上記の係数の逆数を重み係数として掛けることを考える。

3. 認識実験

認識実験は、無雑音データのケプストラムを標準パタンとして、各雑音データを照合する。マッチング次数 15 次、マッチング時のフレーム周期は 10ms として、不特定話者認識を行っている。時間軸整合は、5 フレーム端点フリーDP マッチングを用いている。マッチング時に雑音データのケプストラムの全次数に一様に重み係数をかけるものとする。

前章に示したように、全区間でのケプストラム分布より各ノイズに適した重み係数を考え、無雑音と平均値を比較してみる。Fig11、には無雑音と重み付けしていない pink noise0, dB におけるケプストラム値の全フレーム平均値を次数ごとに示す。このように雑音データでは全体的にケプストラム値が小さくなっていることがわかる。

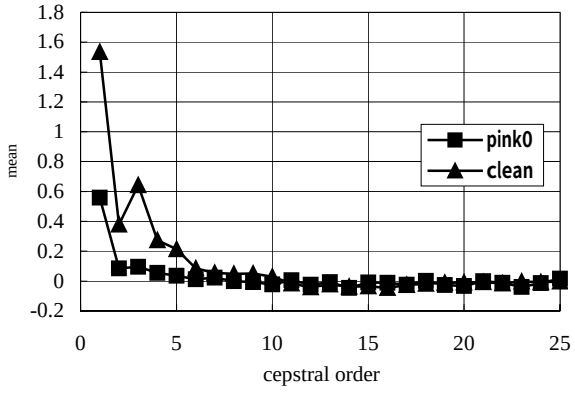

Fig11 Average value of the cepstral coefficients
（Clean vs Pink 0dB）

Fig12 には無雑音と雑音環境データのケプストラムに前章を参考にした係数で重み付けした場合のケプストラム分布例を示す。

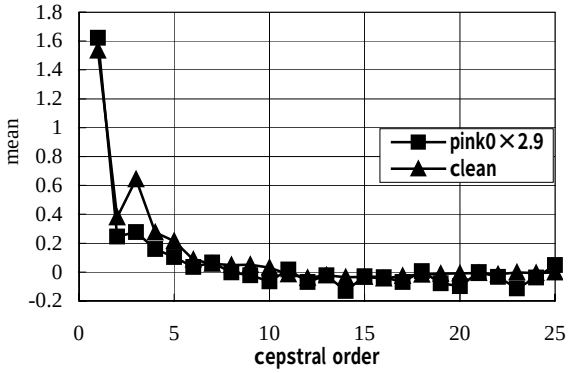

Fig12 Average value of the cepstral coefficients
（Clean vs Pink 0dB）

Fig12 の重み付けしたものと Fig.11 の重み付けしてないものを noise 別で比較すると重み付けすることによりケプストラム値が無雑音に近づいている。Automobile では重み付けしてもしなくても pink noise 程の変化は見られなかった。

次に、雑音環境データのケプストラムに重み付けして認識した場合の認識率について、重み係数による変化の様子を Fig13 に示す。

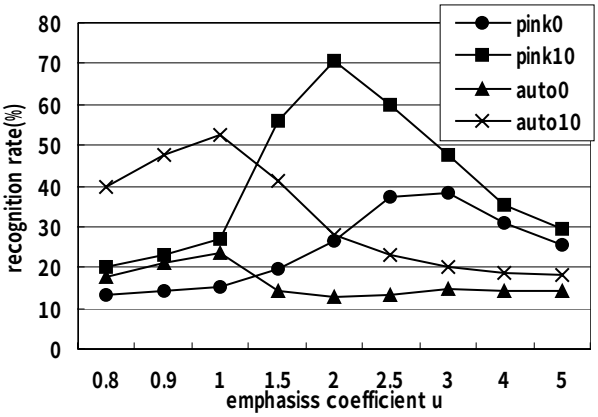

Fig13 Recognition results of noisy digit using emphasized cepstral coefficients

Fig13 より pink noise 0dB では重み係数 u=2.5～3.0、pink noise 10dB では u=1.5～2.5、automobile 0,10dB では u=0.9～1.5 の間で最も認識率が向上している。

更に、それぞれの区間について詳しく分析すると pink noise 0dB の場合 u=2.9、pink noise 10dB の場合 u=1.9、automobile 0,10dB の場合 u=1.1 の時が最も認識率が向上した。これら

の結果は前章で求めた無雑音ケプストラムと雑音環境データケプストラムの平均値比較の結果と大体一致していることがわかる。

ここで得られた最適重み係数をかけた雑音環境音声ケプストラムによるスペクトル包絡を無雑音ケプストラムによるスペクトル包絡と比較してみる。

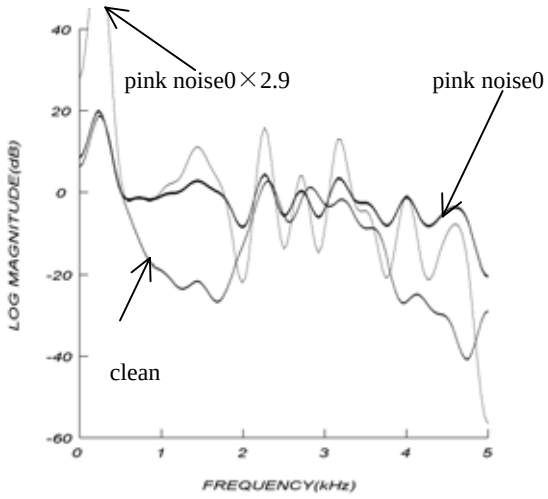

Fig14 spectral envelope of Japanese vowel /i/ using emphasized cepstral coefficients (pink 0dB)

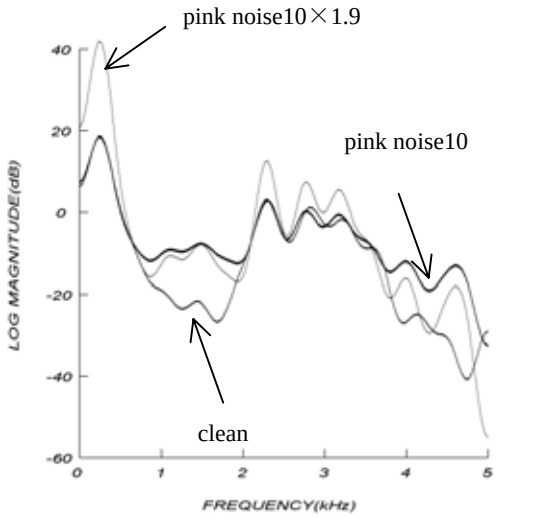

Fig15 spectral envelope of Japanese vowel /i/ using emphasized cepstral coefficients (pink 10dB)

Fig14、Fig15 の pink noise0,10 dB について重み付けしていない場合とした場合の音声スペクトルについて考察すると、pink noise0,10 dB の両方ともが、重み付けした場合の方が無雑音との音声スペクトルと似た特徴が得られた。これにより重み付けした方が良い認識結果が得られることが分かった。

automobile0,10 dB については、重み付けしない場合と重み付けした場合の音声スペクトルについて考察すると重み付けしていない方が無雑音と音声スペクトルの特徴が似ているが、重み付けした方は、谷の部分に着目し、しきい値を決めてやれば認識率の向上が見られると判断できる。

4. まとめ

今回の実験では、pink noise0 dB の場合 u=2.9、pink noise10 dB の場合 u=1.9、automobile 0,10 dB の場合 u=1.1 の時がそれぞれのノイズ別での重み係数の最適値であり、認識率が一番向上された事が分かった。

今後の課題として、雑音環境による最適重み係数の自動算出や次数別の最適重み係数の検討などが必要である。

参考文献

(1) 秋田,吉田,緑川 “非直線関数を用いたスペクトル規則変形による耐雑音音声認識“, 信学技法 EA2003-102, pp.1-6 (2003)

(2) 阿部,今井 “改良ケプストラム法によるスペクトル包絡の抽出 ”, 信学論(A) J62-A No.4, pp.217-223 (1979)