

重み付けケプストラムによる雑音環境音声認識*

衛藤 雄大 秋田 昌憲 緑川 洋一
大分大学

1. まえがき

雑音環境下での音声認識では、分析された音声に雑音が混入することは避けられない。雑音付加により、音声データを表すスペクトル包絡は、谷の部分はその雑音で埋もれレベル上昇を起こす。低域部分で特徴の大きな音声は、高域部分のスペクトル包絡に大きな変形が生じてしまうため、認識は困難なものとなり認識率が低下する。

音声認識では、雑音を差し引いたり^[1]、雑音によって失われた谷を付加する^[2]などの方法が用いられるが、雑音によるスペクトル包絡への影響は低周波数領域で小さいため、Oppenheim の再帰式^[3]による周波数軸変換を行い、低域を強調することで認識率の向上を試みる。また母音と子音とでは音素的な重要度が異なることを考慮し、有声・無声判断をし、それぞれの区間で周波数変換係数を変える時変周波数変換を行った。

最後に、ケプストラムにおいては次数が上がった場合、その値が小さくなることに着目し、高次のケプストラムに重みを付け、特徴付けをして認識実験を行う。周波数軸変換、重み付けを併用した場合の認識実験も行い、その影響について検討する。

2. 周波数軸変換とケプストラムの重み付け

周波数軸変換は、Oppenheim の再帰式法を用いる。ここでは u を周波数軸変換係数とする。

ケプストラムへの重み付けは、入力・標準の音声データのケプストラムを $c(m)$ とするとき、(1)式のような操作を加え、これを重み付けケプストラム $c_w(m)$ とする。

$$c_w(m) = 1 + \frac{m-1}{u} c(m) \quad (m = 1, 2, \dots, N) \quad (1)$$

ここで m はケプストラム次数で u は強調の度合いを決めるものである。ケプストラムでスペクトル包絡を表すと(2)式となる。

$$S(k) = \sum_{m=0}^{N-1} c(m) \cos \frac{2\pi km}{N} \quad (2)$$

これを周波数成分 k で微分すると(3)式となるので、これを群遅延スペクトルという。

$$\frac{d}{dk} S(k) = -\frac{2\pi}{N} \sum_{m=0}^{N-1} mc(m) \sin \frac{2\pi km}{N} \quad (3)$$

これから(1)式で $u = 1$ とした場合は群遅延スペクトルのマッチングを行っていることになる。

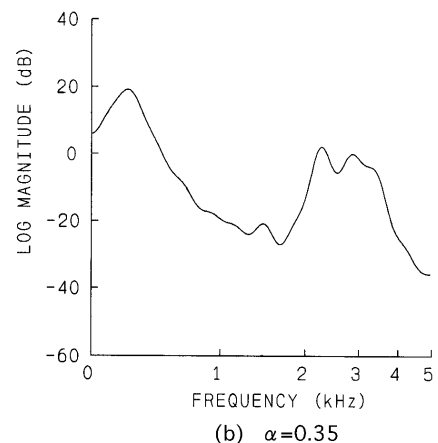
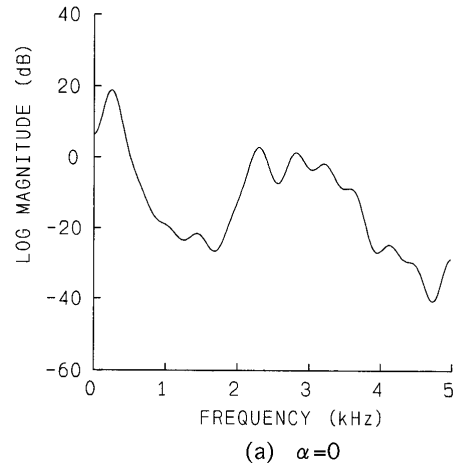


図1 周波数軸変換でのスペクトル包絡変化例
(a) $\alpha=0$ 周波数軸変換は行わない
(b) $\alpha=0.35$ で周波数軸変換

* Speech Recognition Under the Noisy Environment Using Weighted Cepstral Coefficients.
Takehiro ETO, Masanori AKITA and Yohichi MIDORIKAWA (Oita University)

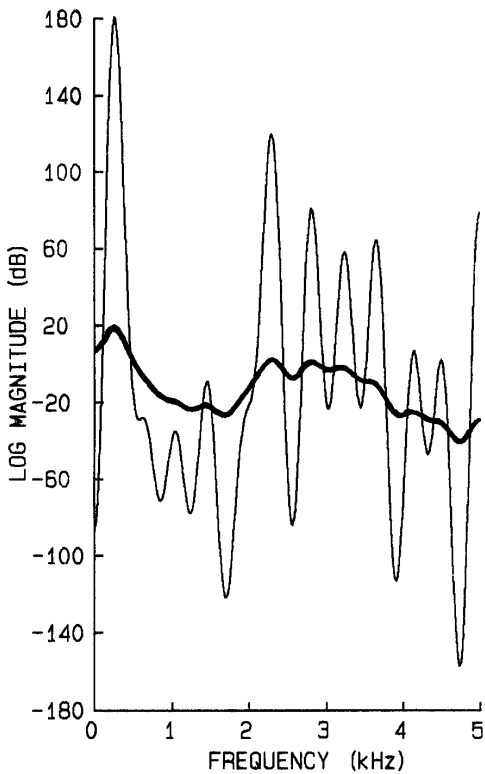


図2 重み付けケプストラムでの
スペクトル包絡変化例
(太線 - 重みなし 細線 - 群遅延スペクトル)

男性話者が /i/ と発音したときのスペクトル包絡を用い、図1に周波数軸変換を施した場合の変化例を、図2にケプストラムに重みを付けた場合の変化例を示す。

3. 認識実験

使用する音声データは、男性話者8名が各3回発音した10数字音声を標準パターンとし、S/N比0dB,10dBのピンクノイズ、自動車ノイズを付加したものの、計5パターンのデータを用いる。

図3に有声・無声の区別をせずに全区間において周波数軸変換した場合の認識率の変化を示す。

最も雑音影響の大きな0dB ピンクノイズ付加音声は変化させても認識率の向上はあまりみられないが、10dB ピンクノイズ付加音声は $\alpha=0.70$ で最も認識率が良くなっている。これはピンクノイズが周波数の低・中・高域の全区間で雑音レベルが高いため、音声特徴の最も大きな低周波数領域を強調することによりその影響を小さくすることができたと考えられる。

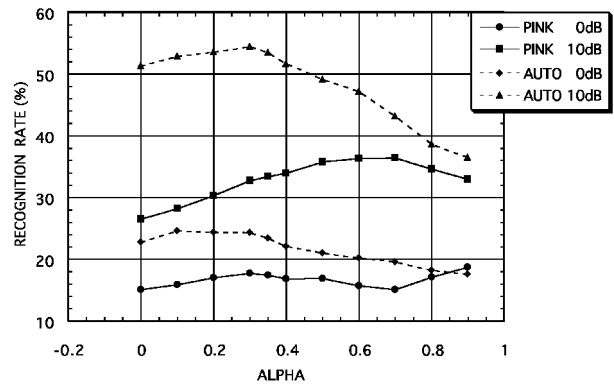


図3 全区間周波数軸変換した認識率

自動車ノイズを付加した音声については $\alpha=0.30$ で認識率の向上がみられる。しかしその向上率はわずかで、 $\alpha=0.40$ を超えると認識率は低下する。この傾向は雑音レベルの小さい10dBの自動車ノイズ付加音声に強く現れている。自動車ノイズは高周波数領域でわずかに雑音レベルが低くなっているため、音声にこのノイズが付加しても、高域にはもとのスペクトル包絡の特徴が残っている可能性が高い。このため極端に低域を強調することで、高域に残っている特徴が消されてしまい、逆に認識率の低下を招いてしまったと考えられる。

以上の結果を踏まえると、全ての雑音付加音声に対して $\alpha=0.35$ 付近での認識率がよく、人間の聴覚特性を考慮したメルケプストラムの有効性が実証された。

次に音声を有声・無声判断した後に、無声部のスペクトル包絡を利用しない方法と、有声・無声各々の区間で異なった周波数変換係数を用いて変換する方法で実験を行う。

有声・無声判断は、無雑音データの最小位相ケプストラムを $c(i)$ とするとき、

$$UV = \sum_{i=0}^3 c(i) \quad (4)$$

とし、(4)式で UV が -1.0 以上の場合を有声部、 -1.0 より小さければ無声部と判断する。

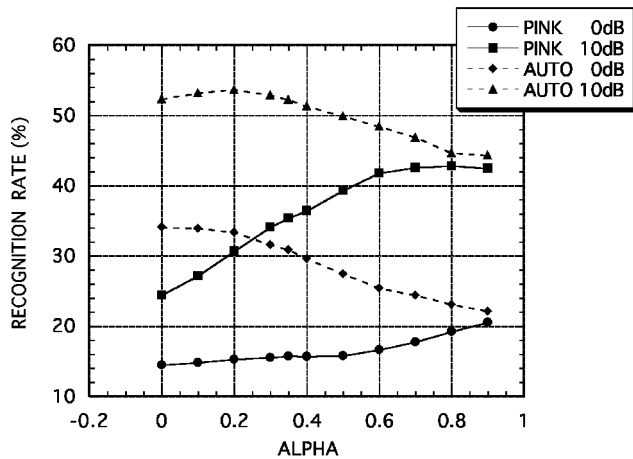


図4 無声部を一定値に規格化した場合の認識率

まず、無声部のスペクトルを一定値に規格化した場合、つまり有声部のみを比較対象としたときの実験の結果を図4に示す。

無声部の値を規格化しても全区間周波数軸変換した結果とあまり大きな変化はみられないが、0dB 自動車ノイズを付加した音声に限っては大きく認識率が上がっている。この結果から0dB 自動車ノイズは無声部分のスペクトル包絡に大きな変形を生じさせていることがわかる。ピンクノイズ付加音声では α の値を大きくし低周波数領域を強調することで、また、自動車ノイズ付加音声では極端な強調は避けることで良い結果が顕著に出ていることから、低域から高域まで雑音レベルの高い環境下では低域強調が、高域で雑音レベルの低い環境下ではメルケプストラムが音声認識に有効であることがわかる。

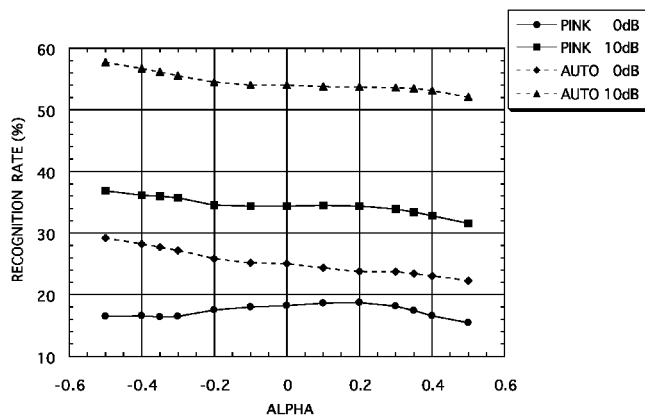


図5 時変周波数軸変換した認識率

図5は有声部の $\alpha = -0.35$ 一定とし、無声部の α を変化させた場合の実験結果を示す。この時変周波数軸変換による認識実験では、ピンクノイズ、自動車ノイズのどちらの雑音が付加されても、無声部の低域を強調すると認識率は落ちてしまう。逆に無声部の高周波数領域を強調すると認識率が上がる。

無声部の $\alpha = -0.50$ 一定とし有声部の α を変化させた場合の認識率は図6のようになる。

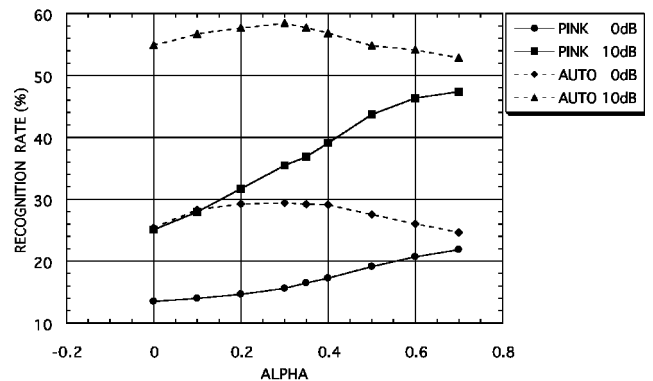


図6 無声部の α を一定とした認識率

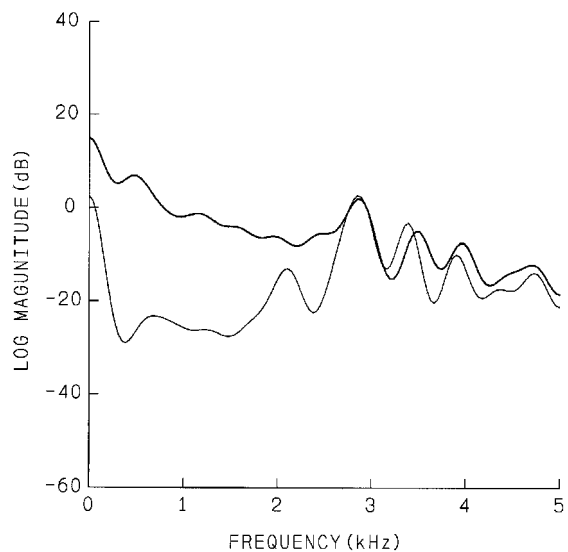


図7 子音部のスペクトル包絡
(太線 - 雑音付加 細線 - 無雑音)

図7は子音部の雑音付加音声と無雑音のスペクトルを比較したものであるが、これを見てわかるように子音部においては高周波数領域でスペクトルに特徴が残っている。このため無声部のスペクトルは高域を強調する周波数軸変換を行うことで認識率が良くなるのがわかる。

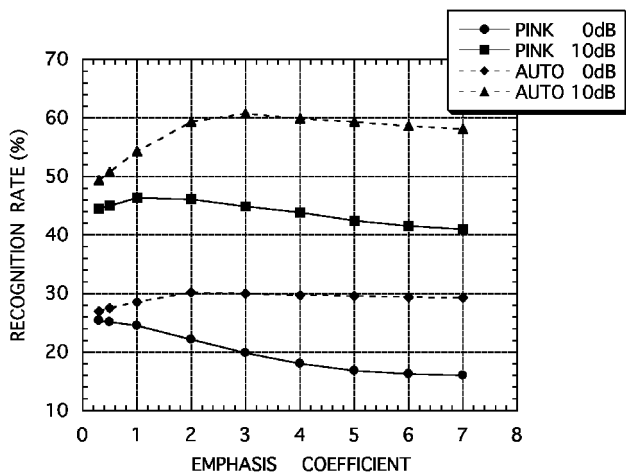


図8 重み付けケプストラムによる認識率
(無声部は一定値に規格化)

最後に、ケプストラムに重みを付け、特徴付けを施し認識実験を行った。まず、音声特徴が大きく現れていると思われる有声部のみ $\alpha=0.35$ で周波数軸変換を施し、無声部は一定値に規格化する。そのスペクトル比較の認識結果を図8に示す。

どの雑音付加音声においても認識率の向上が見られ、 u の値を大きくし、ケプストラム係数が小さくなるにつれ、認識率も下がっていくので、ケプストラムへの重み付けは音声認識に有効であることがわかる。しかし、ケプストラムに大きな係数を付けるすぎるとは逆に認識率の低下を招いている。

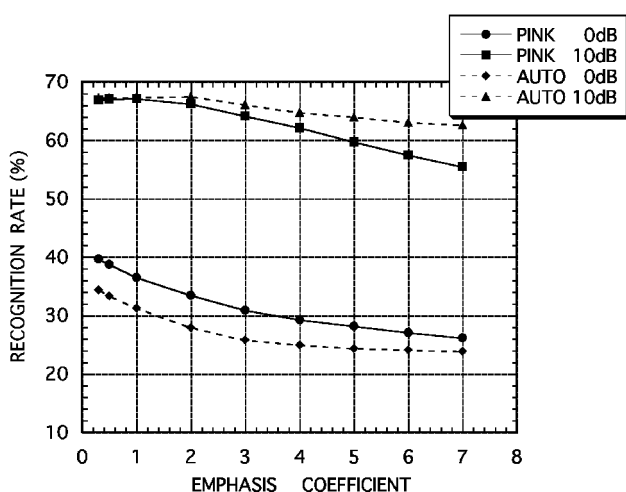


図9 重み付けケプストラムによる認識率
(全区間 $\alpha=0.35$ で周波数軸変換)

無声部も比較対象とし、様々に時変周波数変換して認識実験を行った。そのうち最も認識率に向上の見られた全区間 $\alpha=0.35$ で周波数軸変換した結果を図9に示す。

無声部を比較することは誤差を増し、認識率低下をもたらすのではと思われたが、ピンクノイズ付加音声で認識率は大きく向上し、どの雑音付加音声も u の値が小さいと効果が高い。 u の値を小さくすればするほど高次のケプストラムに、より大きな係数がかかることを考えると、雑音が重畳しても無声部の高次ケプストラムには音声情報が残っていると思われる。

5.まとめ

以上の結果から、周波数軸変換による低域の強調は音声認識に有効であるが、音声により効果のある変換係数は違うので、その自動判断などが今後の課題となる。また高次ケプストラムへの重み付けは、大きな係数、つまり u の値が小さいほど認識率は良くなり、特に高周波でも雑音レベルが高い場合の音声認識で、大きな効果が得られた。今後は他の方法との併用による音声認識の検討などが挙げられる。

6.参考文献

- [1]S.F.Boll:IEEE Trans. ASSP ASSP-27,2,pp.113-120(1979)
- [2]秋田、大倉：信学技法 EA95-57(1995)
- [3]A.V.Oppenheim and D.H.Johnson:Proc.IEEE60 pp.681-691(1972)