

雑音環境音声認識における スペクトル強調と時間平滑化パラメータの評価*

金親慎也 秋田昌憲 緑川洋一(大分大学)

1. まえがき

現在、音声認識を利用したシステムは多いが、完全な音声認識というのは未だに実現されていない。その大きな問題として雑音がある。今回の実験ではそういった雑音を数字音声に擬似的に付加した音声データを用いて、何も雑音のついていない無雑音と比べることで、認識率の高い変形方法を知ることが大きな目的である。これまで、雑音が付加されることで、図1に示すように、低レベル部でのレベル上昇を起し谷が埋もれ、認識率の低下につながることに着目し、規則変形法を用いて谷を擬似的に付加する^[1]実験を試みてきた。

本稿では、重み関数により時間方向の動的特徴を考慮して谷の誤付加解消を試みるスペクトルの平滑化・区間差分化^[2]を行うとともに、雑音の影響を受けにくいスペクトルの低域周波数領域のピークを強調するスペクトル包絡強度軸の非線形変換法^[3]と併用してその認識率の変化を検討した。

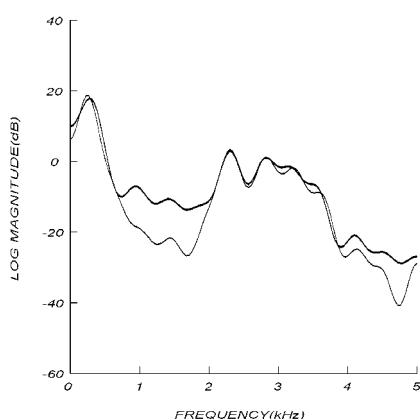


図 1. 男性話者/i/のスペクトル
包絡の雑音による影響

2. スペクトルの平滑化・区間差分化

これまでの実験の有声部のスペクトル包絡に谷を付加する方法については、原スペクトル包絡としきい値関数との交点が不規則で谷の誤付加などの問題が見られ、またフレームごとの静的特徴しか考慮していない為、時間方向や周波数方向への不安定性や不連続性などが生じる。この改善のため、時間方向重み関数による平滑化・区間差分化によって、フレーム間の動的特徴を考慮することによって、谷の誤付加解消を試みた。

次のような、ケプストラム係数、重み関数を定義し、平滑化区間差分化を行う。

kフレームの、次数iのケプストラム係数を $c_i^{(k)}$ とするとkフレームのケプストラムは

$$C^{(k)} = (c_0^{(k)}, c_1^{(k)}, \dots, c_i^{(k)})^T \quad (1)$$

と表すことができる。

次に、平滑化フレーム数をLフレームとし、そのLフレームを1ブロックとする。そのブロックの真ん中のフレームをkフレームとすると、区間Lのケプストラムは、

$$B^{(k)} = (c^{(k-\frac{L-1}{2})}, \dots, c^{(k-1)}, c^{(k)}, c^{(k+1)}, \dots, c^{(k+\frac{L-1}{2})}) \quad (2)$$

となる。

このケプストラムを時間方向重み関数によって、平滑化・区間差分化する。このとき重み関数を $w(j)$ 、平滑化フレーム数をLフレームとすると重みつき強調ケプストラム係数は次

*Evaluation of the Parameters for the emphasis and smoothing of spectrum in Speech Recognition under the noisy environment. Shinya Kaneoya, Masanori Akita, and Yoichi Midorikawa (Oita University)

式によって定義される。

$$V_i^{(k)} = \sum_{j=0}^{\frac{L-1}{2}} \left[w(j + \frac{L-1}{2} + 1) \cdot c_j^{(k+j)} \right] \quad (3)$$

$$w(j) = \begin{cases} 2j/L & (0 \leq j < L/2) \\ 2-2j/L & (L/2 \leq j < L) \end{cases} \quad (4)$$

次に平滑化した場合のスペクトル包絡の時間変化図を図2に示す。

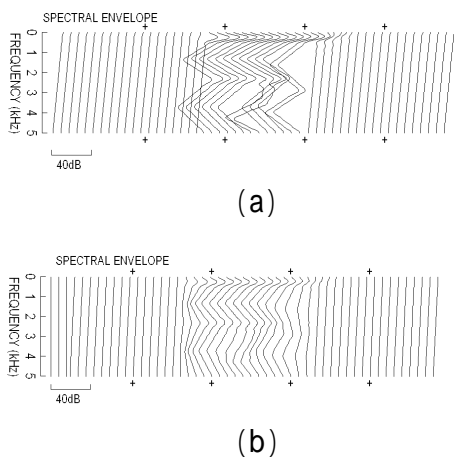


図2. (a)谷を付加した場合のスペクトル包絡
(b)重み関数 $w(j)$ 、 $L=7$ フレームで平滑化した場合のスペクトル包絡

3. スペクトル強度軸の非線形変換法

雑音によるスペクトル包絡全体の平滑化を補正する意味で、スペクトルのピーク域の強度レベルを強調する式(6)の非線形変換を行う。また γ を強調係数とし、元のスペクトルを $S(k)$ 、変換後のスペクトルを $S'(k)$ とし式(5)に示す。

また、図3にスペクトル包絡のピーク域の強度レベルを強調した例を示す。

$$S'(k) = \begin{cases} S(k) & (S(k) \leq TH1) \\ TH1 + \gamma(S(k) - TH1) & (S(k) > TH1) \end{cases} \quad (5)$$

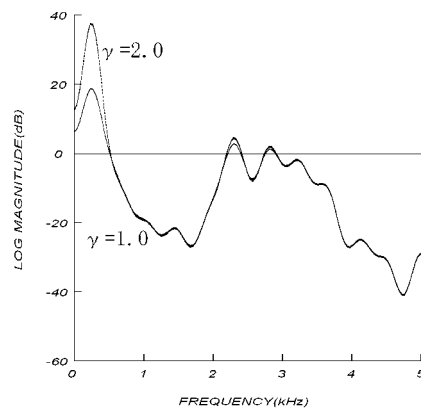


図3. スペクトル強度軸変換例
($TH1=0\text{dB}$, $\gamma=2$)

4. 認識実験

使用音声は男性話者8名が各3回発声した10数字音声を用い、S/N比がそれぞれ0(dB)、10(dB)のピンクノイズ、自動車ノイズを波形上で付加し、またフレーム周期10(msec)フレーム長25.6(msec)、26次で改良ケプストラム分析^[4]されている。標準パターンに入力と同様に変形を加えた、無雑音を用い、入りに雑音を付加したものを用いて、ケプストラム次数15次で5フレーム端点フリーDPマッチングによる不特定話者認識を行った。

まずはじめに、しきい値を $TH1, TH2, TH3, TH4 = 0, -10, -30, -40$ (dB) で谷を付加したものに、 $w(j)$ の重み関数を用いてフレームLを変えて平滑化をおこなった場合の認識結果を図4に示す。

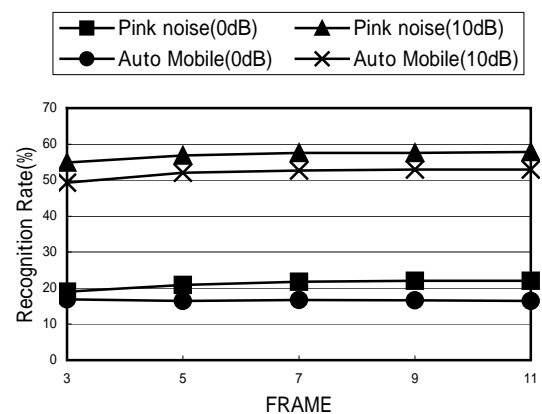


図4. フレームLを変えた場合の認識結果

実験結果をみると、どのノイズにおいても谷の規則変形のみの場合と比べると認識率の向上がみられた。平均的には $L=7$ フレームがよく、平滑化のかけ過ぎは、認識率の低下につながると考えられる。またスペクトル包絡をみてもフレーム数が、9フレーム以上で平滑化を行うと、スペクトル包絡が平坦化しているのがわかる。また、自動車ノイズ 0dB では、平滑化をかけると、認識率の低下が見られるが、これは谷を付加するしきい値によって、うまく谷が付加されていなかったり、谷がもともと浅いため、平滑化を行うとかえって谷を平坦化させてしまい、特徴が薄くなるためだと考えられる。

次に同様に谷を付加した後に強度軸強調を行い、強調係数を変えた場合の認識結果を図5に示す。

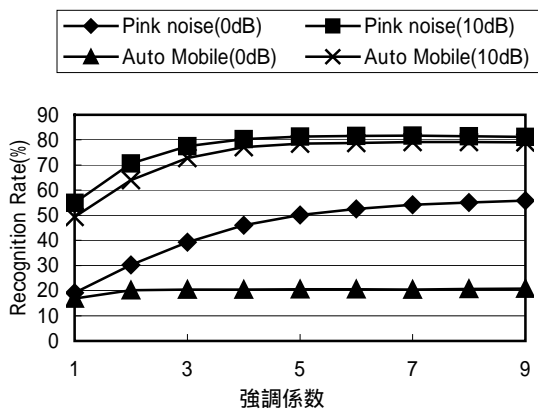


図5. 強調係数を変えた場合の認識結果

この結果から、ピーク強調を用いた場合自動車ノイズよりもピンクノイズの方が顕著に認識率の向上がみられた。これは、ピーク強調では主に低域周波数領域のピークを強調しているため自動車ノイズのように高域周波数領域よりも低域周波数領域のレベルが高いような雑音は、雑音の成分までも強調してしまい、認識率が低下すると考えられる。

次に同様に谷を付加した後に平滑化、ピーク強調の順で変形を行い、フレーム L と強

調係数を変えた場合の認識率の結果をノイズ別に図6～9に示す。

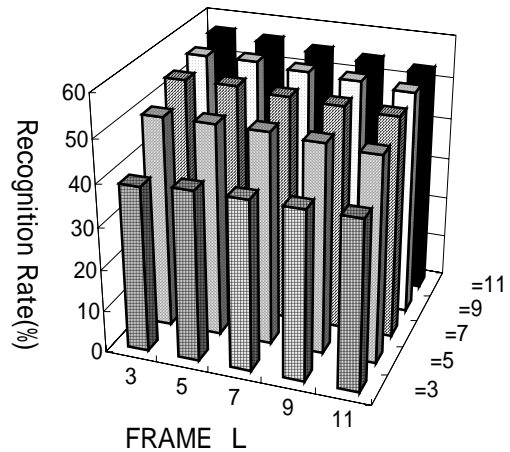


図6. 平滑化とピーク強調を併用した場合の認識率 Pink noise(0dB)

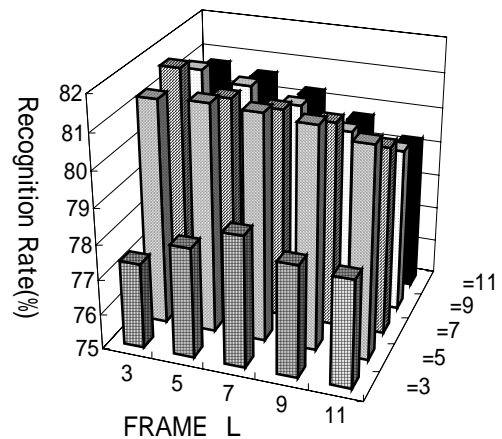


図7. 平滑化とピーク強調を併用した場合の認識率 Pink noise(10dB)

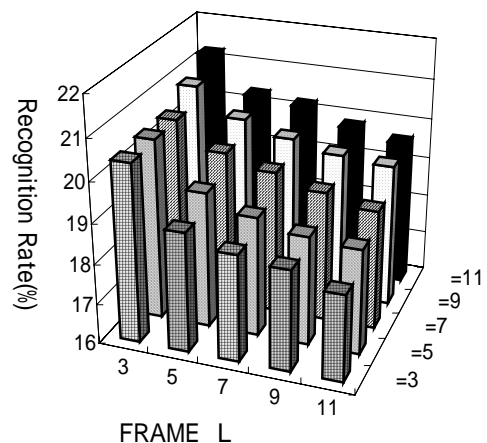


図8. 平滑化とピーク強調を併用した場合の認識率 Auto Mobile(0dB)

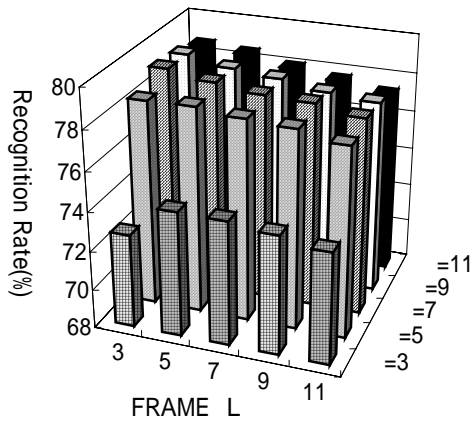


図9. 平滑化とピーク強調を併用した場合の認識率 Auto Mobile(10dB)

以上の結果から平滑化とピーク強調を併用した場合の認識結果から最も認識率の高いパラメータの組み合わせを表1に示す。

表1. 平滑化とピーク強調を併用した場合でノイズ別での適したパラメータ

	フレームL	強調係数
Pink noise(0dB)	3	11
Pink noise(10dB)	3	7
Auto Mobile(0dB)	3	11
Auto Mobile(10dB)	3	7

この結果から、平滑化では、L=3 フレームで、すべてのノイズにおいて、最も認識率が高く、またピーク強調では、ノイズレベルの高いものほど、強調を大きくかけることで認識率が向上すると考えられる。

最後に今回の認識実験を段階別にかけて、図10に示す。

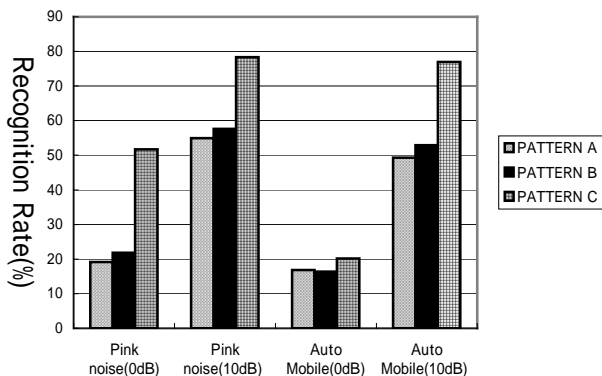


図10. 各変形法組み合わせた場合の認識結果

PATTERN A:規則変形

PATTERN B:規則変形+平滑化

PATTERN C:規則変形+平滑化+ピーク強調

5. まとめ

スペクトルの平滑化・区間差分化では、どのノイズにおいても谷の規則変形の場合と比べると認識率の向上が見られた。また非線形変換法と組み合わせることによってどのノイズにおいても大幅な認識率の向上が見られた。今後は、新たな重み関数の提案や、変形を行った音声を合成しプレファラテストと認識結果の関係を考慮しながらノイズ別の適切なしきい値の自動設定などが検討される。

6. 参考文献

[1]秋田、大倉:信学技法 EA98-57(1995)

[2]坪江、秋田、緑川:秋音講論 1-Q-10 (2001)

[3]秋田、有川:音響学会春季講演集 3-Q-4 pp125-126(1999)

[4]安部、今井:“改良ケプストラム法によるスペクトル包絡の抽出“J62-A,4,pp217-223 (1979)