

# 多段しきい値によるスペクトルピーク強調を用いた耐雑音音声認識\*

松本晃紘, 秋田昌憲, 緑川洋一 (大分大)

## 1 はじめに

音声認識においての基本的認識方法は音声スペクトル包絡の比較である。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埋もれてしまう。そのことが認識率の低下につながる。そこで、雑音によって埋まった谷を回復するために、スペクトル包絡規則変形法<sup>[1]</sup>によりしきい値を設定し低周波数領域から谷を復元する実験が行われ、有効性が示されている。

しかし、スペクトル包絡規則変形法による谷の復元も十分ではない。そこで、雑音環境下においてスペクトル包絡の谷部では、レベル上昇が見られるが低周波領域における高レベル部分では特徴が無雑音とほぼ一致しているということに着目し、スペクトルのピーク域の強度軸レベルを強調し、形状変形に対応することを考える。これまでにしきい値関数より上の部分を係数により強調するというスペクトル強度軸非線形法の有効性を試みてきた。だが、そのままでは雑音成分まで強調してしまうおそれがあるため、本実験ではしきい値関数を変化させることで問題解消を、またスペクトル包絡規則変形法とスペクトル強度軸非線形法を併用することによる認識率の向上を試みた。

## 2 実験方法

### 2.1 スペクトル包絡規則変形法

音声のあるフレームのスペクトル包絡を  $S(f)$  とする。  $P(f)$  を定常有色雑音環境を擬似する低レベル制限直線関数としレベル制限スペクトル  $S'(f)$  は式(1)となる。  $TH1, TH2$  はこの場合のしきい値であり、別途定められる。

$$S'(f) = \begin{cases} S(f) & (S(f) \leq P(f)) \\ P(f) & (S(f) \geq P(f)) \end{cases} \quad (1)$$

$$P(f) = TH1 + \frac{f}{5000}(TH2 - TH1) \quad (2)$$

次に、  $P(f)$  と  $S(f)$  の交点間で埋もれている谷を復活させることを考える。

$W(f)$  は谷の深さを擬似する直線関数であり、しきい値を  $TH3, TH4$  とする。

$$W(f) = TH3 + \frac{f}{5000}(TH4 - TH3) \quad (3)$$

$P(f)$  と包絡の最初の交点を  $f_1$ ,  $f_1$  の 1.2(kHz) 以降の周波数で原スペクトル包絡と  $P(f)$  との交点を  $f_2$ ,  $f_3$  は  $f_1$  と  $f_2$  の中点と  $W(f)$  の交点として次のような関数で谷を付加する。ここで、  $THL(f)$  は谷の左、  $THR(f)$  は谷の右を形成する関数である。

$$THL(f) = P(f_1) + \{W(f_3) - P(f_1)\} \frac{f - f_1}{f_3 - f_1} \quad (4)$$

$$THR(f) = P(f_2) + \{W(f_3) - P(f_2)\} \frac{f - f_2}{f_3 - f_2} \quad (5)$$

第二の谷についてはしきい値を新たに  $TH5$  とおき、  $f_2$  以降の周波数で原スペクトル包絡と低レベル制限直線関数

$$Q(f) = TH1 + \frac{f}{5000}(TH5 - TH1) \quad (6)$$

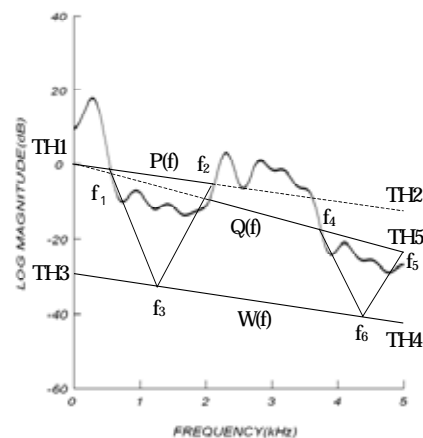


Fig. 1 Example illustrating of the Spectral Emphasis by rules

\* Speech Recognition under the Noisy Environment using Emphasis of the Spectral Envelopes by Multistage threshold function. Akihiro Matsumoto, Masanori Akita, Yoichi Midorikawa (Oita University)

を上から下に横切る周波数を  $f_4$ ,  $f_4$  の 1.2(kHz) 以降の周波数で原スペクトル包絡と  $P(f)$  との交点を  $f_5$ ,  $f_1$  と  $f_2$  の中点と  $W(f)$  の交点を  $f_6$  として第一の谷と同様な方法によってつける。ただし、 $f_5$  に相当する周波数がない場合、ナイキスト周波数をもってこの周波数の代わりとするものである。

## 2.2 スペクトル強度軸非線形法

雑音環境下においてスペクトル包絡の谷において、レベル上昇が見られるが低周波領域における特徴は一致している。そこで、スペクトルのピーク域の強度軸レベルを強調し、形状変形に対応することを考える。

従来のスペクトル強度軸非線形法では、直線しきい値の上部に変化係数をかけるというものだったが、そのままでは雑音成分のピークまで強調してしまうおそれがある。そのため、新たに3つのしきい値関数を提案した。元のスペクトルを  $S(f)$ 、変換係数  $\gamma$  とすると、しきい値  $TH(f)$ 、変換後のスペクトル  $S'(f)$  は下式のように表すことができる。

### ・パターン 1

$$\begin{cases} th(f) = TH1 + \frac{TH2 - TH1}{N} f \quad (0 \leq f \leq 1500) \\ th(f) = TH1 + \frac{TH3 - TH1}{N} f \quad (1500 \leq f \leq 3000) \end{cases} \quad (9)$$

### ・パターン 2

$$th(f) = TH1 + \frac{TH2 - TH1}{N^2} f^2 \quad (10)$$

### ・パターン 3

$$th(f) = TH1 + (TH2 - TH1) \sqrt{f/N} \quad (11)$$

### 変換後のスペクトル

$$S'(f) = \begin{cases} S(f) & [S(f) \leq th(f)] \\ th(f) + \gamma(S(f) - th(f)) & [S(f) > th(f)] \end{cases} \quad (12)$$

$$(0 \leq f \leq 3000)$$

$$S'(f) = S(f) \quad (3000 < f \leq 5000) \quad (13)$$

パターン 1 は直線しきい値を二つ用いたものであり、第二ホルマント以降の雑音部の強調を抑えるためであり、パターン 2 は各ピー

ク部分のレベルの高さが右下がりなためしきい値関数をこれと似た二次曲線としたものである。パターン 3 はパターン 2 の逆関数である。これはパターン 2 で有効性が得られなかった場合、逆関数ではどうかとういことを確認するためである。パターン 2,3 の曲線を Fig.2 に示す。また、強度軸強調を行った時のスペクトル包絡を Fig. 3 に示す。

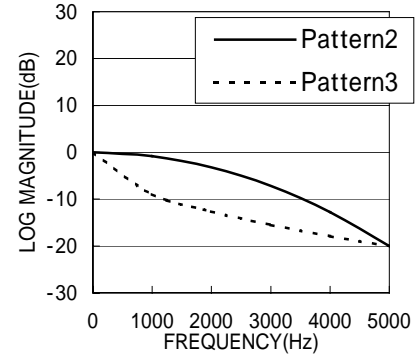
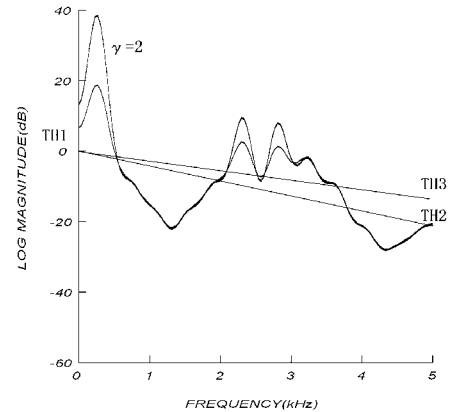
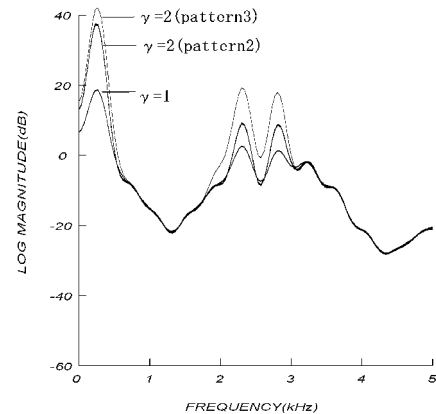


Fig.2 Threshold functions used in the experiments (Pattern 2,3)



(a) Pattern 1



(b) Pattern 2,3

Fig.3 Spectral Envelopes emphasized using the threshold functions

### 3 実験結果

#### 3.1 スペクトル包絡規則変形法

標準パターンに雑音のないデータに谷を付加したものとし、雑音を重畳したデータに谷を付加したデータとの認識実験を行う。実験を行った際に、Table 1のようにしきい値を変化させる。このときの認識率を Table 2 に示し、認識率の平均値を Fig.3 に示す。

Table 2 The thresholds used in the transform of spectral envelopes by rules

PATTERN	Threshold function(dB)				
	TH1	TH2	TH3	TH4	TH5
A	0	-10	-30	-40	-20
B	0	-10	-30	-40	-30
C	0	-10	-20	-30	-20
D	0	-10	-20	-30	-30
E	0	-20	-30	-40	-30
F	-10	-20	-30	-40	-30
G	5	-20	-30	-40	-30
H	5	-10	-20	-40	-30
I	0	-10	-30	-40	-15
J	-5	-10	-30	-30	-20

Table 3 Recognition rates using spectral transform (%)

PATTERN	Pink		Mobile	
	0dB	10dB	0dB	10dB
A	22.26	59.56	21.61	57.72
B	16.23	33.57	29.42	70.75
C	23.59	70.50	28.29	70.40
D	14.70	34.09	38.39	80.71
E	15.16	32.56	23.61	67.68
F	15.42	49.62	27.50	54.50
G	14.70	30.65	26.94	68.95
H	12.56	24.68	39.92	71.29
I	20.79	59.23	19.17	53.35
J	23.12	66.23	20.67	57.42

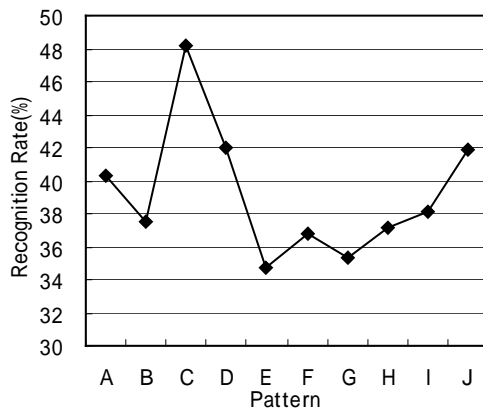


Fig.4 Recognition Results using the transform of spectral envelopes by rules

この結果、最も良い認識率を得たのがパターンCである。良い認識率を得られたパターンC,D,Jより、二つ目の谷はあまり深くないほうが良いということが分かる。

#### 3.2 スペクトル強度軸非線形法

標準パターンを雑音のないデータに谷を付加後ピーク強調したものと、雑音を重畳したデータに同様変形したものととの認識実験を行った、谷の付加は Table 2 にあるパターンCで付加する。ピーク強調は Table 4,5 のようにしきい値を設定し、強調係数は2から9まで変化させる。強調を行わないときは =1 である。各しきい値パターンでの認識率の平均値を Fig.5 から Fig.7 に示す。

Table 4 The thresholds pattern used in the emphasis of the local peaks (pattern1)

PATTERN	Threshold function(dB)			
	TH1	TH2	TH3	TH4
A	0	-20	0	-5
B	0	-10	0	-5
C	0	-20	0	-10
D	0	-20	-5	-10

Table 5 The thresholds pattern used in the emphasis of the local peaks (pattern2,3)

PATTERN	Threshold function(dB)	
	TH 1	TH 2
A	0	-5
B	0	-10
C	0	-20

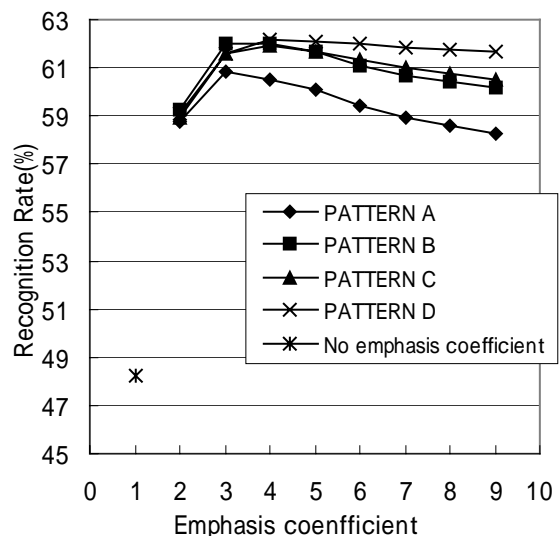


Fig.5 Recognition Results by pattern1

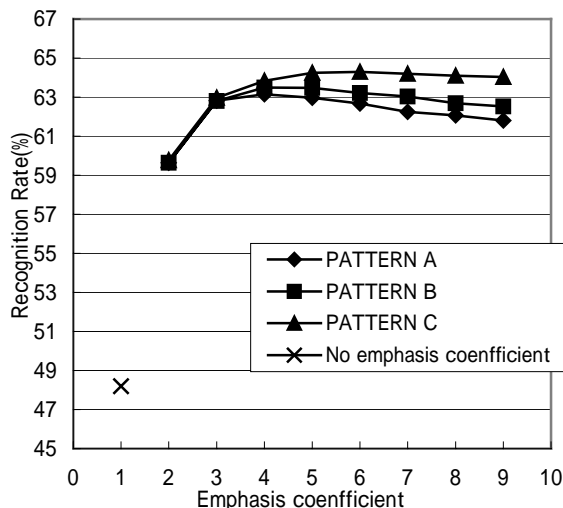


Fig.6 Recognition Results by pattern2

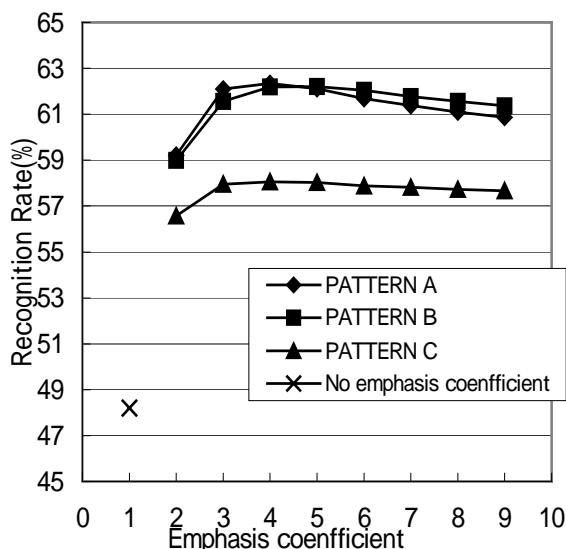


Fig.7 Recognition Results by pattern2

この結果、すべてのしきい値パターンにおいて認識率の向上が見られた。最も良い認識率を得ることができたのがパターン 2-C の =6 のときである。

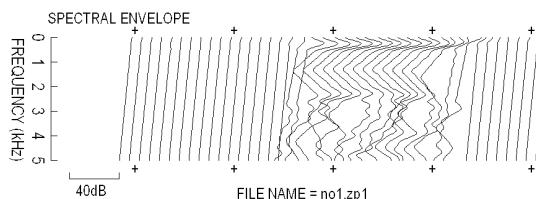


Fig.8 Time series of the transformed spectral envelopes of Japanese word /ichi/ (expert) (pink noise 10dB, TH1=0dB, TH2=-10dB, TH3=-20dB, TH4=-30dB, TH5=-20dB, =1)

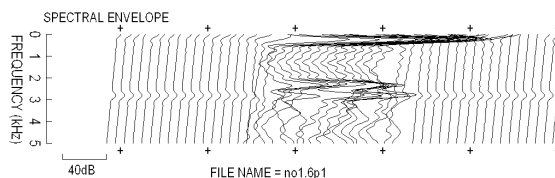


Fig.9 Time series of the transformed spectral envelopes of Japanese word /ichi/ (expert) (pink noise 10dB, TH1=0dB, TH2=-10dB, TH3=-20dB, TH4=-30dB, TH5=-20, =6)

ここで Fig.8 に谷付けのみのスペクトル時間変化の図、Fig.9 に最適強度軸強調パターン 2-A( =6)のときのスペクトル時間変化の図を示す。Fig.8,9 より強調したとき雑音部分への強調が抑えられていることが分かる。

#### 4 まとめ

本実験では雑音環境下においてのスペクトル包絡規則変形法とスペクトル強度軸非線形法による認識率の向上を試みた。

その結果、スペクトル強度軸非線形法による認識率の向上は二次曲線が有効であることが分かった。また、 の上げすぎによる認識率の低下は全体の概形が変わりすぎることが原因だと考えられる。

今後の課題をしては、さらに認識率を向上させることのできるしきい値関数の確立である。

#### 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” 信学法, EA2003-102, pp.1-6, (2003).
- [2] 秋田,川野,緑川: “改良型スペクトル強度軸強調による雑音環境音声認識” 信学技法 EA2002-91, pp.19-24, (2002).