

# スペクトル強度軸の非線形変換による 雑音環境音声認識\*

有川 尚志 秋田昌憲  
(大分大学工学部)

## 1. はじめに

雑音環境下における音声認識では、音声の特徴を示す谷間の部分の音声特徴に変化が現れてくる。その結果認識率は低下してしまう。これを補正する目的でこれまで、雑音環境下での音声認識におけるスペクトル包絡レベルの対処法として、直線しきい値関数を用いた簡易スペクトルサブストラクション<sup>[1]</sup>によるスペクトルへの疑似雑音条件付加補正、また逆に、欠落したスペクトル包絡の谷の部分の規則による復活<sup>[2]</sup>、及び無雑音のスペクトル包絡の平滑化などを行い認識実験を行ってきた<sup>[3][4]</sup>。ここでは、今まで挙げた方法より簡易的で同一の効果をもたらすと考えられるスペクトル強度軸変形法を提案している。これはスペクトルの低レベルの強度軸を圧縮するもので谷の部分のレベル差を減少するものである。この方法を雑音環境音声認識に適用して効果を検討する。今までの実験から、認識は主にデータの有声部で行われているので、音声の特徴が現れる有声部を重視し、無声部を0とする方式も実施している。

## 2. スペクトル包絡強度軸の非線形変換法

雑音を重畳した音声では母音性音素のスペクトル包絡の谷の部分の部分が埋まってしまう包絡全体が平滑化されるのでこれを補正する意味で次のような非線形変換を行う。

元のスペクトルを  $S(k)$ 、強調係数を  $\beta$  とすると変形後のスペクトル  $S'(k)$  は

$$S'(k) = \begin{cases} S(k) & (S(k) \geq TH1) \\ TH1 + \beta(TH1 - S(k)) & (S(k) < TH1) \end{cases} \quad (1)$$

と表すことができる。ここで非線形変換する前とした後のスペクトル包絡図を図1に示す。図1の(a)では音声の特徴を表す谷の部分の部分が埋まっているのが分かる。

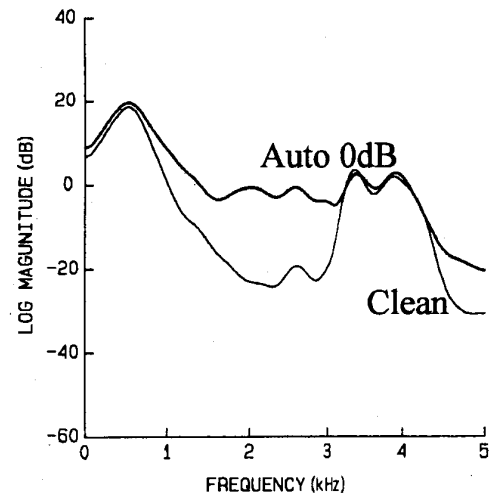


図1. (a) 非線形変換前の  
スペクトル包絡図

\*Speech Recognition under the Noisy Environment by Warping the Spectral Magnitude  
,Naoshi Arikawa and Masanori Akita (Oita University)

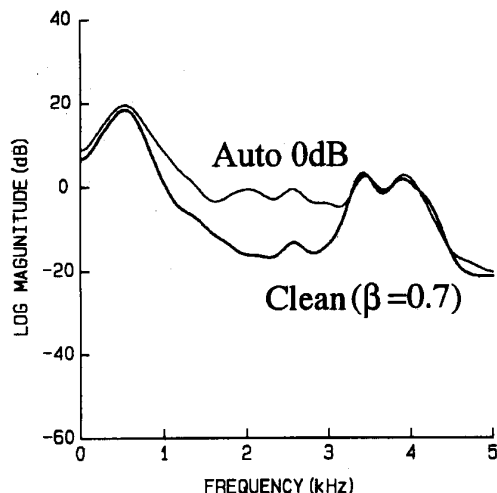


図1. (b) 非線形変換後のスペクトル包絡図 ( $\beta = 0.7$ )

図1 (a) は自動車ノイズ0dBを重畳した場合で (b) は無雑音を非線形変換 ( $\beta = 0.7$ ) したものである。TH1は0dBとしている。図1において非線形変換することで、波形が近くなることが分かる。

### 3. 有声無声判断法

ここで言う有声・無声判断は無雑音最小位相ケプストラムを  $\alpha(i)$  するとき

$$UV = \sum_{i=0}^3 \alpha(i) \quad (2)$$

とし (2) 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。ここで音声の特徴が現れる有声部を重視し、無声部の影響を無くすため、無声部を0とする方法を述べる。無声部を0とするには、周波数変換された後のケプストラムを  $\tilde{\alpha}(i)$  とすると

$$\tilde{\alpha}(i) = \begin{cases} k\tilde{\alpha}(i) & (uv < -1) \\ \tilde{\alpha}(i) & (uv \geq -1) \end{cases} \quad (3)$$

(3) 式において  $k=0$  とし無声部を0とする。

### 4. 認識結果

認識実験は男性話者8名が各3回ずつ発声した10数字に2種類の有色雑音を重畳したものをを用いている。ここでピンクノイズは零周波数に対しナイキスト周波数でノイズレベルが10dB低い有色雑音であり、自動車ノイズは同様にナイキスト周波数でノイズレベルが30dB低い有色雑音である。そのため、前者の方が白色雑音に近い性質を持っている。これらのデータは、フレーム周期10msで改良ケプストラム法でケプストラム分析し、再帰式法によってメル周波数軸に周波数変換される。包絡変換の際の有声・無声判断は無雑音データのケプストラムの低次部の和で自動判断している。認識実験は5フレーム端点フリーDPマッチングによる不特定話者によって行われる。標準パターンに雑音付加したものを入力パターンに無雑音を用いている。まず前章で述べたTH1を0, -10, -20dBとした場合の有声部無声部とも式(1)で変形した場合の結果を図2、図3、図4に示す。なお強調係数1の時は従来法である。

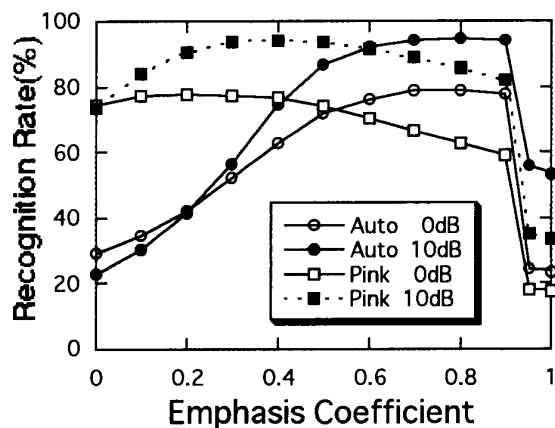


図2 TH1=0dBにおいて非線形変換した場合の認識結果

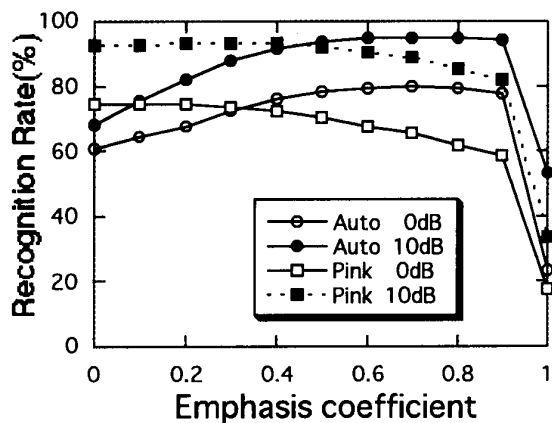


図3. TH1=-10dBにおいて非線形変換した場合の認識結果

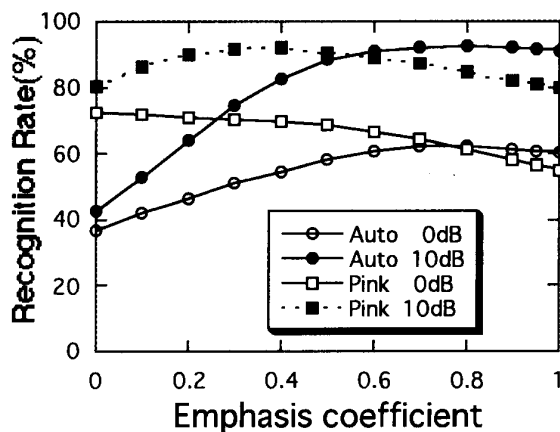


図5. 有声無声判断後 TH1=0dBにおいて非線形変換した場合の認識結果

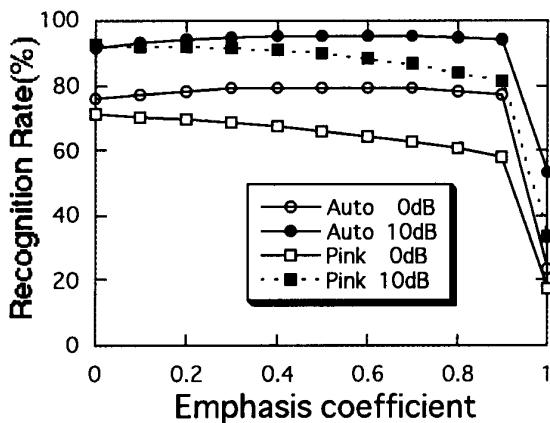


図4. TH1=-20dBにおいて非線形変換した場合の認識結果

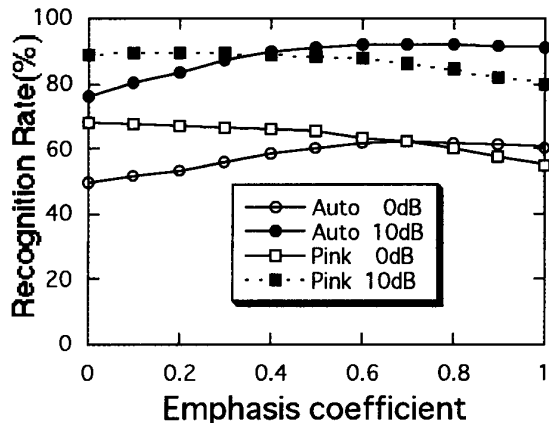


図6. 有声無声判断後 TH1=-10dBにおいて非線形変換した場合の認識結果

ここで、ピンクノイズ0dB、ピンクノイズ10dBではTH1=0dBの場合が最高認識率を示しており、自動車ノイズ0dBの場合はTH1=-10dB自動車ノイズ10dBにおいてはTH1=-20dBの場合に最高認識率を示している。次に有声無声判断して有声部のみ非線形変換無声部0の場合の結果を図5、図6、図7に示す。

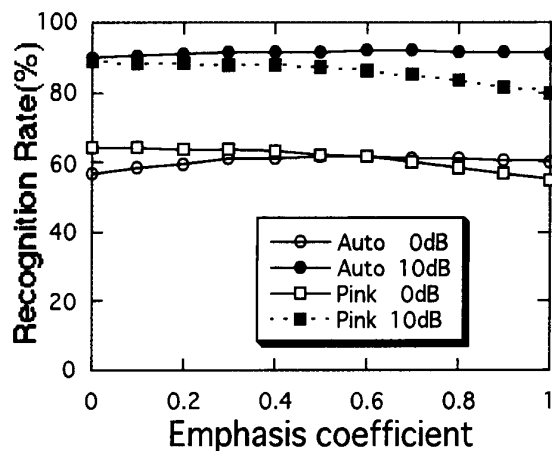


図7. 有声無声判断後 TH1=-20dBにおいて非線形変換した場合の認識結果

有声無声判断した後では、全ての雑音に対してTH1=0dBの場合が最高認識率を示している。この場合TH1を下げるとスペクトル包絡の変形があまり行われなことが考えられる。今回の認識実験での最高認識率とそのときのTH1と強調係数 $\beta$ の値を表1、表2に示す。

表1. 非線形変換した場合の最高認識率とその時のTH1と強調係数 $\beta$ の値

	Auto 0dB	Auto10dB	Pink 0dB	Pink 10dB
従来法 (%)	23.41	53.47	17.42	33.43
TH1	-10	-20	0	0
強調係数 $\beta$	0.7	0.4	0.2	0.4
最高認識率 (%)	80.04	95.16	77.54	94.33

表2. 有声無声判断後非線形変換した場合の最高認識率とその時のTH1と強調係数 $\beta$ の値

	Auto 0dB	Auto10dB	Pink 0dB	Pink 10dB
従来法 (%)	60.12	90.83	54.88	79.74
TH1	0	0	0	0
強調係数 $\beta$	0.8	0.8	0	0.4
最高認識率 (%)	62.44	90.83	72.20	91.79

## 5. まとめ

4

本報告では、雑音により音声汚染された場合のスペクトル包絡において無雑音のスペクトル包絡を非線形変換することとあわせ有声部だけに着目し認識実験を行ってきた。認識率はスペクトル包絡

を非線形変換することで強調係数が適切な場合において全ての雑音に対して有効性を示した。最高の認識率は全ての雑音において非線形変換のみ施した場合がよい結果が得られた。今後の課題としてTH1の適切な値の検討、周波数変換係数変化の判断の規範および各々の条件における最適なスペクトル包絡の最適変型法についての詳細な検討、及び距離尺度を含めた本方式によるによるスペクトル変形を用いた場合のマッチング方式の検討等が課題となる。

## 参考文献

- [1] S.F. Boll; "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Trans. ASSP Assp-27, 2, pp. 113-120 (1979)
- [2] 秋田, 凶師: "スペクトル変形と音声合成を利用した雑音環境音声から雑音除去の一方法", 信学春全大 A-299 (1994)
- [3] 木村, 秋田: "スペクトルのレベル制限と帯域制限を併用した雑音環境音声認識", 信学技法 EA92-81 (1992)
- [4] 秋田, 大倉: "雑音環境におけるスペクトル変形回復の一方法", 信学技法 EA95-57 (1995)