時間方向平滑化による雑音環境音声認識

江上　茂　　秋田　昌憲　　緑川　洋一
大分大学工学部

１．　まえがき

時代はとうとう２１世紀となり、音声認識に関する技術も大変発達してきたが、雑音環境下における完全な音声認識というのは未だ実現していない。しかし、すべてが簡略化していく現代において、音声による機械への入力方法は大変有効であり、これからさらに進歩していく分野であるといえる。

今回は、雑音環境下における音声認識の向上のため、雑音を数字音声に擬似的に重畳した音声データを使用し、それと雑音を付加する前の数字音声との認識率を見ることよって、音声認識の向上を目的とする実験を行った。

音声認識に用いるケプストラムにより抽出されるスペクトル包絡は、無雑音環境下では、はっきりと見られる谷が、雑音環境下においては低レベル部のレベル上昇により谷が見られない。本実験では、その雑音が重畳されたスペクトル包絡に谷を付加する実験をまず行い、続いて重み関数の設定によるスペクトル包絡の平滑化によって、谷の付加の実験では不規則に付き、乱れているスペクトル包絡を、谷の誤負荷の修正による認識率の向上を図った。以上の実験を、認識率の結果からもっとも適したスペクトル変形法と、しきい値を探り出した。

２．谷の付加によるスペクトル包絡変形法

雑音環境下では、特に有声部の場合、図１のようにスペクトル包絡の低レベル部が上昇してしまい谷の部分がなくなってしまう。そこで、しきい値を TH1～TH4 まで設定し図２のような規則にしたがって谷を付加し、スペクトル包絡の特長の回復を図る。

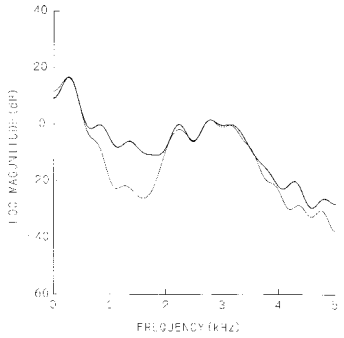


図１　自動車ノイズ S/N 比 10dB が重畳されたスペクトル包絡

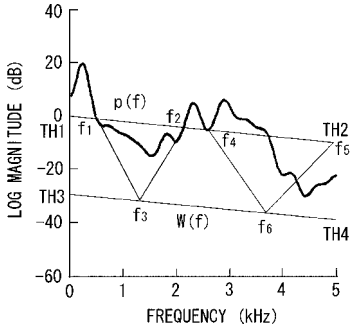


図２　規則による谷の付加の説明

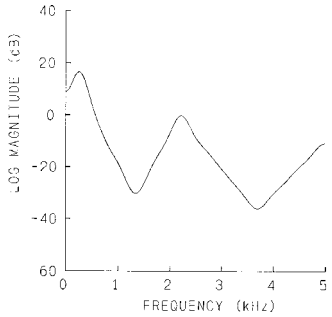


図３　図１に谷を付加したスペクトル包絡

Speech Recognition under the Noisy Environment Using the Smoothing Function to Cepstral Coefficients.

Shigeru Egami ,Masanori Akita ,and Yoichi Midorikawa　(Oita University)

３．　平滑化による谷の誤付加解消

図２において、しきい値ともとのスペクトル包絡との交点を探し、これらの周波数のところに谷を付加しているが、これらの交点は不規則で、得られる包絡は図４のように乱れている。また、ノイズ付加によってスペクトルの谷が変化することにより静的特徴は変化するが、となりのフレームとの特徴差など動的特徴の変化は静的特徴より少ないと考えられる。そのため、動的特徴を認識パラメータとすると認識率が向上する可能性があると考えられる。ゆえに、次のようなケプストラム係数、重み関数を定義し乱れた谷の修正を行う。

$C_i^{(k)}$　：次数ｉのケプストラム係数

（ｋ：フレーム数）

○ｋフレームのケプストラム

$$C^{(k)} = (c_0^{(k)}, c_1^{(k)}, \cdots c_i^{(k)})^T$$

○平滑後のｋフレームのケプストラム

（Ｌ：平滑化フレーム数）

$$B^{(k)} = \left(C^{(k-\frac{L-1}{2})}, \cdots; C^{(k-1)}, C^{(k)}, C^{(k+1)}, \cdots C^{(k+\frac{L-1}{2})} \right)$$

○重み付ケプストラム係数

$$V_i^{(k)} = \sum_{\frac{L-1}{2}}^{\frac{L+1}{2}} [w(j + \frac{L-1}{2} + 1) \cdot c_i^{(k+j)}]$$

○重み関数

$$w(j) = \frac{1}{L} \quad (0 \le j \le L-1)$$

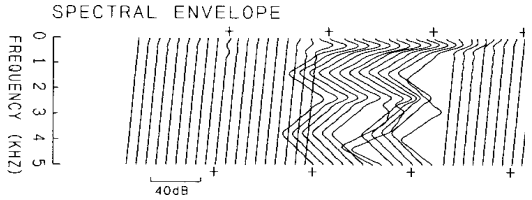


図４　谷を付加したスペクトル包絡の時間変化

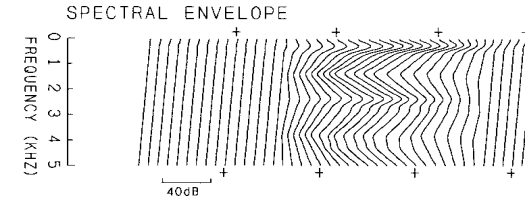


図５　平滑化フレーム L=9 で平滑化されたスペクトル包絡の時間変化

４．　認識実験

本実験で使用する音声データは８人の男性がそれぞれ１０数字を３回発声し、それに S/N 比 0dB、10dB の自動車ノイズ、ピンクノイズをそれぞれ重畳したもので、フレーム周期 10msec で改良ケプストラム分析されている。また、標準パターンに無雑音の音声データケプストラムを、入力データに雑音を重畳し変形を行った音声データケプストラムを使用した。認識方法として、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おこなった。

まず、表１の A から F までのパターンでしきい値を変化し、それによる埋れた谷の回復によって認識率の向上を図った。表２には変形を行わない場合の認識率と、谷をしきい値パターン A～F で谷を付加した場合の認識率の変化である。このことにより、谷の回復による認識率の変化と、適切なしきい値の値を調べる。

表１　しきい値の変化

（単位はｄＢ）

Pattern	TH1	TH2	TH3	TH4
A	0	-10	-20	-30
B	0	-10	-30	-40
C	0	-20	-20	-30
D	0	-20	-30	-30
E	-10	-20	-20	-30
F	-10	-20	-20	-40

表２　谷を付加した場合の認識率の変化

（単位は%）

Pattern	Pink		Mobile	
	0dB	10dB	0dB	10dB
-	15.4	27.1	23.41	52.68
A	24.66	66.07	19.27	59.15
B	19.88	56.13	17.26	50.46
C	20.46	64.68	26.01	68.69
D	19.54	61.17	21.81	63.55
E	20.5	67.22	25.1	61.47
F	21.17	55.02	24.37	52.26

変形を加えない場合に比べて、谷を付加した場合のほうが全体的に認識率が上昇した。特に、ピンクノイズにおいては、しきい値パターンＡが、自動車ノイズにおいてはＣパターンで最も認識率の上昇が見られる。

次に重み関数を用いた平滑化により、乱れた谷の修正による認識実験を行った。谷を付加する際のしきい値のパターンは同様に Ａ～Ｆ の６パターンで行った。また、平滑化を行う際の平滑化フレーム数Ｌは９フレームで実験を行った。

表３　平滑化における認識結果

（単位は%）

Pattern	Pink		Mobile	
	0dB	10dB	0dB	10dB
A	27.1	69.21	17.92	62.3
B	23.13	58.97	16.47	53.81
C	18.57	68.08	28.12	71.63
D	18.91	65.26	23.57	67.64
E	18.35	68.83	24.44	60.24
F	18.67	58.99	25.16	53.59

これまでの認識実験の結果を、しきい値パターンＣにおいて図６にまとめてみる。

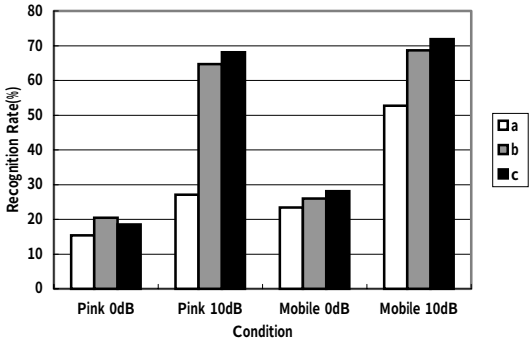


図６　認識率の変化

ａ：変形を加えない場合の認識率

ｂ：谷を付加した場合の認識率（しきい値パターンＣ）

ｃ：平滑化を行った認識率（９フレームで平滑化）

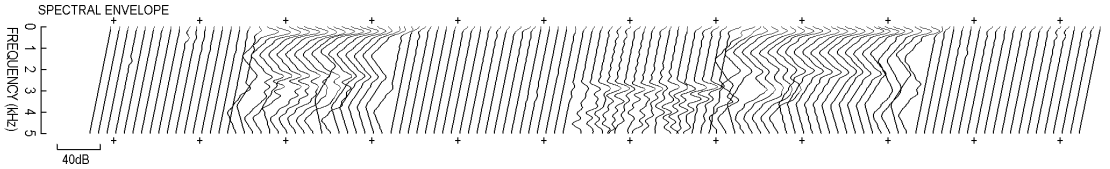
上図のように、谷を付加したものに重み関数を用いた平滑化により、谷の誤付加解消された場合の認識率は全体的に上昇したことがわかる。

さらに、重み関数を用いて平滑化による認識率の上昇を求めるため、平滑化フレーム数Ｌの値を変化させることにより実験を行った。これまでは、平滑化フレーム数は Ｌ＝９によってのみ行っていたが、Ｌの値を５，７，９，１１と変化させ、認識率の変化を調べた。その結果を表４に示す。また、５フレーム、１１フレームで平滑化されたスペクトル包絡を図７に示す。

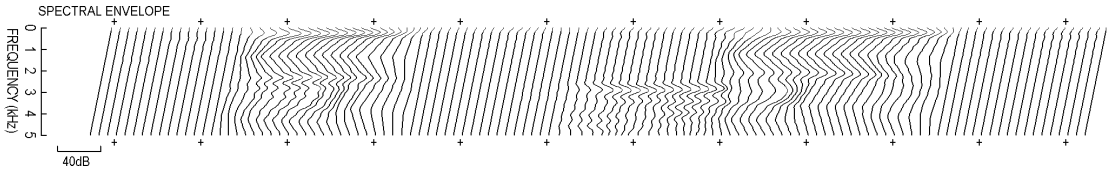
表４　フレーム長Ｌの変化による認識率

（単位は%）

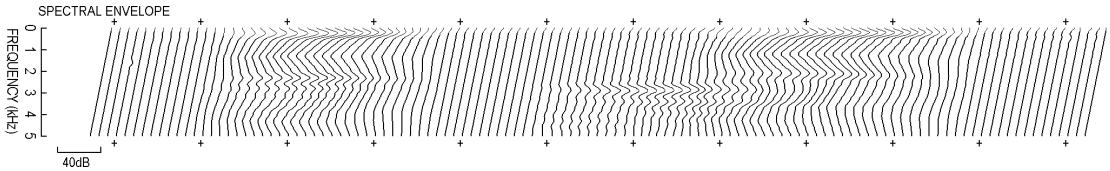
Frame	Pink		Mobile	
	0dB	10dB	0dB	10dB
5	19.4	69.05	27.96	71.59
7	18.85	68.93	28.31	71.85
9	18.57	68.06	28.12	71.63
11	17.34	66.75	27.76	70.44



しきい値パターンＣで谷を付加



平滑化フレーム L=５で平滑化



平滑化フレーム L=１１で平滑化

図７　平滑化フレーム数Ｌの変化によるスペクトル包絡の時間変化

全体的に７フレームがよい結果になったが、９フレーム､11 フレームの場合に注目すると、平滑化をやりすぎることによって認識率がかえって減少してしまうことが分かる。

５．まとめ

本実験では、雑音が重畳されたスペクトル包絡に谷を付加する実験を行い、続いて重み関数の設定によるスペクトル包絡の平滑化によって、谷の誤付加の修正を行って認識率の向上を図った。

その結果、しきい値設定による規則にしたがって谷を回復する場合、ノイズによって適切なしきい値があるということがわかった。また、時間方向の重み関数による平滑化により、谷の誤付加によるスペクトル包絡の乱れを解消し、さらに認識率が上昇させることができるということがいえる。しかし、この平滑化を行い過ぎると、かえって認識率は減少してしまうという結果が得られた。

今後の課題としては、平滑化を行う重み関数を、他の式によって実験を行うこと、また周波数軸の変換や帯域制限等との併用、スペクトル包絡のピーク値の強調によって認識率の向上を図る、などのことが検討される。

参考文献

（１）　秋田、大倉：信学技法　EA95-57(1995)
（２）　坪江、秋田、緑川：秋音講論 1-Q-10(2001)