

雑音環境音声認識用スペクトル規則変形法の改善

奥山雅彦 秋田昌憲 緑川洋一
(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^[1]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^[2]、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^{[3][4]}。これらの中でスペクトルに谷を付加する方法は有効性は大きいですが、誤った谷を付加して認識率に悪影響を及ぼす場合も見られた。ここでは、その規則について検討し、改善した場合の認識実験結果例を示す。

2. 雑音環境スペクトル包絡の規則による谷付加によるスペクトル補正

有声部スペクトル包絡に対し、まず式(1)のように一次関数 $th(k)$ を用いて簡易スペクトルサブストラクションを行う。ここで k は離散的周波数を表し、 N はフーリエ変換の点数を表している。これをスペクトル低レベル制限と呼んでいる。

$$S'(k) = \begin{cases} S(k) & (S(k) > th(k)) \\ th(k) & (S(k) \leq th(k)) \end{cases}$$

$$th(k) = TH1 + 2 \frac{TH2 - TH1}{N} k \quad (1)$$

その後、図1のようにスペクトルと式(1)の一次関数の交差で二つの谷を付けるものとする。ただし谷の深さは図1に示す $TH3, TH4$ による一次関数で定義する。

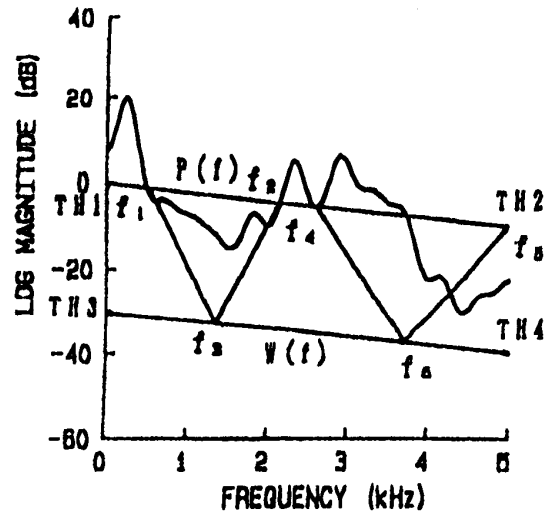


図1 スペクトル包絡への谷の付加例

また、従来法として図1に示した方法と併用して谷を最大で3個付加したり、1つ目の谷の付加に制限を加える^[5]という方法も使用してきた。その付加例を図2に示す。

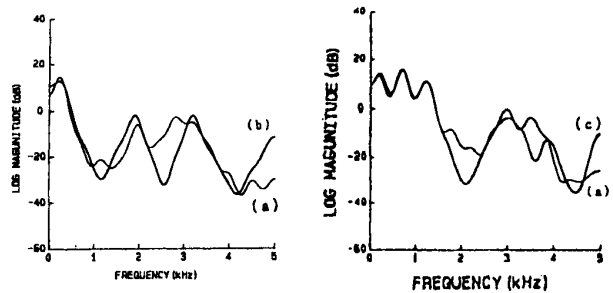


図2 従来法でのスペクトルへの谷の付加例

- (a) 無雑音スペクトルの原包絡
- (b) 谷を最大で3個付加
- (c) 1つ目の谷の付加を制限したもの

しかし、これらの方法では無雑音スペクトルの包絡と谷を付加したスペクトルでの距離の開きがあるので、これを改善する目的で今回は次の3つの実験を行った。

*Improvement of the Spectral Transform Method for Speech Recognition Under the Noise Environment, Masahiko Okuyama and Masanori Akita and Youichi Midorikawa (Oita University)

(方式1)

これまで、下側のしきい値 $TH3, TH4$ の値を一定として実験を行ってきたが、その値も上側のしきい値と同様に变化させて実験を行った。その付加例を図3に示す。

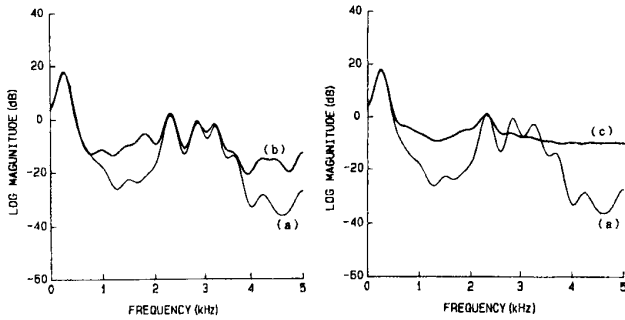


図3 方式1による谷の付加例

- (a) 原包絡
- (b) ピンクノイズ 10dB スペクトル
- (c) 谷を付加したピンクノイズ 10dB スペクトル

(方式2)

先ほどの方式1に加えて今度は認識の際に標準パターンとして用いる無雑音のスペクトル包絡にも変形を加えて実験を行った。その付加例を図4に示す。

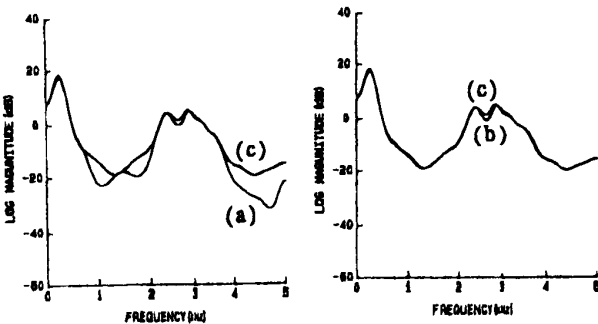


図4 方式2による谷の付加例

- (a) 無雑音スペクトルの原包絡
- (b) 谷を付加した無雑音スペクトル
- (c) 谷を付加したピンクノイズ 10dB スペクトル

図からも分かるように方式1では標準と入力との距離が広がっているのに対し、方式2では標準パターンのスペクトルと入力パターンの

スペクトルとの距離が接近していることが分かる。まずはこの2つの方法を用いて認識実験を行った。

3. 認識実験

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、再帰式方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。標準パターンに無雑音のものを入力パターンに雑音付加したものをを用いている。

まず、方式1と方式2による認識実験を行った。そのしきい値のパターンを表1に、方式1による結果を表2に、方式2による結果を表3に示す。従来法は $TH3 = -30dB, TH4 = -40dB$ の時の最高値である。

表1 しきい値のパターン

パターン	しきい値 (dB)			
	TH1	TH2	TH3	TH4
A	5	0	-10	-10
B	0	-5	-10	-10
C	-5	-10	-10	-10
D	5	-5	-10	-10
E	0	-10	-10	-10

表2 方式1による認識結果

単位 (%)

パターン	自動車ノイズ		ピンクノイズ	
	0dB	10dB	0dB	10dB
A	17.62	18.77	20.83	19.21
B	19.82	24.52	19.52	26.43
C	22.22	32.90	17.68	31.87
D	20.38	23.13	20.26	23.73
E	21.96	30.71	18.43	29.58
従来法	32.32	65.32	32.08	55.02

表3 方式2による認識結果

単位 (%)

パターン	自動車ノイズ		ピンクノイズ	
	0dB	10dB	0dB	10dB
A	49.86	79.66	59.74	80.04
B	31.98	88.71	57.72	87.82
C	22.30	88.27	45.85	92.50
D	48.21	83.04	63.65	81.19
E	32.48	88.91	46.94	90.24
従来法	32.32	65.32	32.08	55.02

この結果から、方式1では標準のスペクトルと入力との距離が開いてしまい、認識率の低下を招いているが、方式2では標準と入力との距離が小さくなったことにより認識率が向上していると考えられる。

(方式3)

これまでの実験から認識率の低下を招いている理由の1つとして高周波数域でのピークが考えられる。そこで選択平滑化法¹⁶⁾を用いた実験を行った。まず、ケプストラムを周波数変換して低周波数域を広げリフタ (lifter) で次数を従来の25次から12次に下げて細かいピークをとり、次に、逆周波数変換をして周波数

軸を元に戻した。これにより、スペクトル包絡の低周波数域のピークはそのままにして、高周波数域の余分なピークを除去することができる。その変形例を図5に示す。

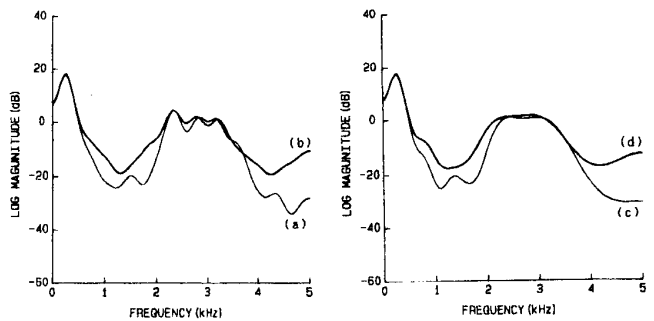


図5 変形後のスペクトル包絡

- (a) 雑音重畳なし
- (b) 谷を付加したピンクノイズ 10dB
- (c) 選択平滑化法による変形を加えた無雑音
- (d) 選択平滑化法による変形を加えた谷を付加したピンクノイズ 10dB

この変形を方式2による変形を施したものに対して行い、認識実験を行った。その結果を表4に示す。

表4 方式3による認識結果

単位 (%)

パターン	自動車ノイズ		ピンクノイズ	
	0dB	10dB	0dB	10dB
A	50.14	80.12	60.99	80.10
B	31.29	88.57	57.92	87.62
C	21.69	88.49	45.85	92.46
D	48.89	83.21	64.56	80.85
E	32.04	88.95	46.51	90.18
従来法	32.32	65.32	32.08	55.02

この結果と表3による結果を比べると、全体のパターンではないが認識率の向上しているものもある。

表5にこれまで行ってきた各々の実験方式による認識率の最高値を示し、これにより比較検討を行う。

表5 各実験における認識率の最高値
単位 (%)

変形法	自動車ノイズ		ピンクノイズ		平均
	0dB	10dB	0dB	10dB	
従来法	32.32	65.32	32.08	55.02	46.19
方式1	22.22	32.90	20.83	31.87	26.96
方式2	49.86	88.91	63.65	92.50	73.73
方式3	50.14	88.95	64.56	92.46	74.03

従来法に対して方式2と方式3では認識率が向上しているのが分かる。逆に方式1での認識率の低下は標準と入力のスpekトル包絡の開きが原因だと思われる。

4. まとめ

今回の実験は、有色雑音環境下での数字音声の認識を行い、雑音により低下した認識率の向上を、その音声のスpekトル包絡に谷を付加する適応変形と選択平滑化法によって達成することを目的として行った。方式1として下側のしきい値を浅くとした場合、認識率は従来法に比べて低下してしまった。これは標準パターンに対する入力パターンのスpekトル包絡の距離が逆に開いてしまった為だと考えられる。そこで方式2として標準パターンのスpekトル包絡にも谷を付加する変形を行ったら従来法に比べ大きく認識率が向上した。これは方式1とは逆に包絡間の距離が近くなった為だと考えられる。方式3の選択平滑化法では著しい認識率の向上は見られなかったが、様々な方式への応用が考えられる。今後の課題としてしきい値の適切な値の検討、各々の条件における最適な変形法についての詳細な検討、谷の付加様式

の自動選択、そして下側のしきい値を従来法に比べて深くとした場合の検討や各々の方式における平滑化などが検討される。

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、図師：“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)
- [5]奥山、秋田：“雑音環境音声認識のためのスペクトル変形法の改善”、電気関係学会九支連大 650(1998)
- [6]秋田：“選択的リフタリングによる雑音環境音声認識の検討” 日本音響学会全国大会講演論文集 1-7-1pp.1-2(1994)