3—P—2　ノイズの変調 Wavelet を用いた音声認識特徴量の選択的利用 *

◎岡田 賢治　荒井 隆行 (上智大・理工)
金寺 登 (石川高専)　百村 裕智　村原 雄二 (上智大・理工)

1. はじめに

音声認識技術において、あらゆる環境において有効な耐雑音技術が必要である。前処理の特徴量抽出の段階で、特徴量の時間軸方向に対して周波数解析処理を施し、特定の周波数成分を抽出することによって認識率が向上することが報告されている[1][2]。Okada et al.[1][2]では、PLP 係数の時間軌跡に対して Wavelet 変換を行ない、特定の変調周波数帯域の成分を抽出することで、認識率が向上することを確認している。この手法を変調 Wavelet 変換と呼んだ。

本研究では変調 Wavelet 変換で得られた音声の平均変調スペクトル成分と、変調 Wavelet 変換で得られたノイズの平均変調スペクトル成分を考慮し、クリーン環境下で得られた HMM の分散値を変化させることによって認識率の向上を試みる。

2. 帯域成分を考慮した単語音声認識実験

変調 Wavelet 変換を用いた従来法と、今回提案した、音声・ノイズ特徴量の平均変調スペクトル成分による影響を考慮した新しい方法を比較するため、単語音声認識実験を行なった。実験環境は表 1の通りである。

表 1. 実験環境

タスク	Bellcore digit (0-9, oh, yes, no の 13 種類 200 人発話の 2600 個)
標本化周波数	8 kHz
フレーム	25 ms
シフト	10 ms
学習	150 人話者 (男性 75 人, 女性 75 人)
評価	50 人話者 (男性 25 人, 女性 25 人)

まず、PLP 係数の時間軌跡に対して Wavelet 変換を施し、音声認識に重要である変調周波数帯 1-10 Hz を分割した。抽出された変調周波数帯域は図 1 の通りである。用いた変調 Wavelet 変換の Mother Wavelet は Meyer 型で、用いたスケールは [32 18 10] である。特徴量は 9 次の PLP 係数を 3 帯域に分けた、計 27 次の特徴量である。

認識・学習には HMM ToolKit (HTK[3]) を利用し、単語毎に状態数6、混合数2の HMM を用いた。雑音は、NOISEX-92 database[4]を利用し、その中の babble, buccaneer1, buccaneer2, destroyerengine, destroyerops, f16, factory1, factory2, hfchannel, leopard, m109, machinegun, pink, volvo, white の雑音を用いた。雑音は、SNR が 10dB になるように混ぜ合わせている。

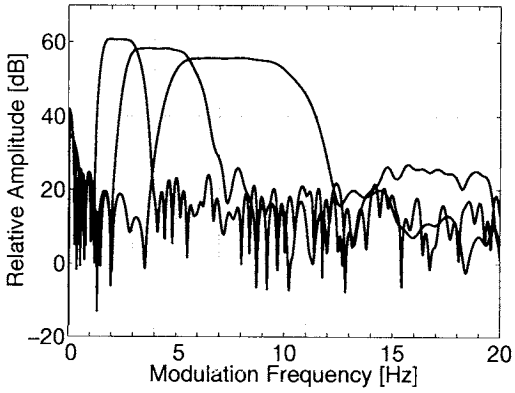

図 1. 変調 Wavelet 変換により抽出される帯域

3. 認識実験とその結果

3.1 音声の平均変調スペクトル成分による音響モデルの補正

音声を変調 Wavelet 変換した特徴量の平均 (RMS) のみを用いて音響モデルの分散値を変化させた場合の効果を確認した。特徴量の RMS が大きいときは、その特徴量はノイズの影響を受けにくいと考えられる。逆に特徴量の RMS が小さいときはノイズの影響を受けやすいと考えられる。よって、RMS が小さい特徴量に対応する分散値を大きくすることによってノイズの影響を軽減できるかどうか調査した。

具体的には、まず各特徴量の RMS の中で最大の RMS 値を求める。次に各特徴量に対応する音響モデルの分散値に（補正係数）×（最大の RMS 値）／（各特徴量の RMS 値）を掛け合わせた。

結果は表 2に示す通りである。補正係数を小さくするに従い認識率は向上したが、従来法を用いた場合の認識率に達することはなかった。

* Selective usage of features for automatic speech recognition using the modulation Wavelet transform of noise

By Kenji Okada, Takayuki Arai (Sophia University), Noboru Kanedera (Ishikawa National College of Technology), Yasunori Momomura, Yuji Murahara (Sophia University)

表 2. 音声の平均変調スペクトル成分を考慮した比較 (単語誤り率 [%])

ノイズ	従来	補正係数			
		0.75	1.0	1.25	1.5
babble	16.9	19.0	18.5	18.8	20.1
buccaneer1	13.8	14.8	16.4	18.9	21.8
buccaneer2	14.9	17.0	17.5	19.5	21.6
destroyerengine	14.5	16.8	17.7	18.9	21.5
destroyerops	13.8	15.1	15.2	16.2	18.4
f16	13.3	15.3	16.2	17.2	19.1
factory1	13.0	14.9	15.4	17.2	19.2
factory2	10.7	12.3	12.5	14.8	17.1
hfchannel	13.7	13.5	15.8	17.4	20.6
leopard	13.3	15.7	14.8	15.5	17.0
m109	19.5	12.6	12.6	13.6	16.0
machinegun	34.5	31.3	31.5	32.6	34.9
pink	11.8	13.5	15.2	17.3	20.2
volvo	6.9	9.0	10.7	13.3	16.1
white	14.8	16.3	19.1	21.7	24.7
平均	15.0	15.8	16.6	18.2	20.6

表 3. ノイズの平均変調スペクトル成分を考慮した比較 (単語誤り率 [%])

ノイズ	従来	補正係数			
		0.75	1.0	1.25	1.5
babble	16.9	16.9	15.4	14.8	16.2
buccaneer1	13.8	13.5	12.9	14.2	15.7
buccaneer2	14.9	15.5	14.0	15.0	16.4
destroyerengine	14.5	14.7	14.6	15.0	16.7
destroyerops	13.8	13.5	12.4	12.6	14.0
f16	13.3	13.7	13.1	13.3	13.8
factory1	13.0	13.6	14.0	17.0	20.4
factory2	10.7	10.8	11.7	13.8	17.1
hfchannel	13.7	14.1	14.9	16.6	19.5
leopard	13.3	11.9	11.1	11.2	13.0
m109	19.5	9.9	9.5	10.2	11.0
machinegun	34.5	31.3	30.1	30.6	33.1
pink	11.8	12.7	11.8	12.0	12.6
volvo	6.9	6.8	7.3	9.2	11.6
white	14.8	15.4	14.7	15.2	16.3
平均	15.0	14.3	13.8	14.7	16.5

3.2 ノイズの平均変調スペクトル成分による音響モデルの補正

ノイズを変調 Wavelet 変換した特徴量の平均 (RMS) を用いて音響モデルの分散値を変化させた場合の効果を確認した。ノイズを変調 Wavelet 変換した特徴量の RMS が大きいときは、その特徴量はノイズの影響を受けやすいと考えられる。よって、ノイズを変調 Wavelet 変換した特徴量の RMS が大きい特徴量に対応する分散値を大きくすることによってノイズの影響を軽減できるかどうか調査した。

具体的には、まずノイズを変調 Wavelet 変換した各特徴量の RMS の中で最小の RMS 値を求める。次に各特徴量に対応する音響モデルの分散値に (補正係数) × (各特徴量の RMS 値) ／ (最小の RMS 値) を掛け合わせた。

結果は表 3に示す通りである。従来法と比較した場合、補正係数が 0.75 と 1.0 の時に改善が見られた。補正係数が 1.0 の時が最高の認識率であり、補正係数を大きくしても小さくしても、1.0 の認識率より悪化した。

4. 考察

まず、音声成分を考慮した場合を見てみると、認識率は従来法より悪化している。これは音声の平均変調スペクトル成分を考慮した方法に対して、補正係数が大きいからと考えられる。

次に、ノイズ成分を考慮した場合を見てみると、1.0 以上 1.25 までは改善が見られたが、1.5 になると悪化した。これは、1.5 になると補正係数が大きくなるからと考えられる。また、ノイズ成分を考慮した手法は、HMM を変化させる比率をノイズ毎に計算するため、改善がみられたものと考えられる。このことは、ノイズの平均変調スペクトルを定期的に取得し音響モデルを随時適応させれば、ノイズ環境の変動にも動的に対応できる可能性を示唆している。

5. まとめ

雑音に対して頑強な特徴量抽出の方法について調べた。従来の変調 Wavelet 変換を用いた方法と、特徴量の音声・ノイズ成分の平均変調スペクトル成分を考慮した方法とを比較した。音声の成分を考慮した方法は改善がみられなかったが、ノイズの成分を考慮した方法は改善が見られた。これから、ノイズの成分を考慮する方法で、更に補正係数の検討や、音声とノイズの両方の成分を検討することで、認識率の向上が期待できる。

参考文献

[1] K. Okada, T. Arai, N. Kanedera, Y. Momomura and Y. Murahara, "Using the modulation wavelet transform for feature extraction in automatic speech recognition," *ICSLP 2000*, Vol. 1, pp. 337 - 340, 2000.

[2] 岡田 賢治, 荒井 隆行, 金寺 登, 百村 裕智, 村原 雄二, "自動音声認識の特徴量抽出への変調 Wavelet 変換の応用," 音講論集, Vol. 1, pp. 53 - 54, 2000.9.

[3] S. Young, D. Kershaw, J. Odell, D. Ollason, V. Valtchev, P. Woodland, "The HTK Book," Ver. 2.2, Entropic, 1999.

[4] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, Vol. 12, No. 3, pp. 247 - 251, 1993.