1—P—13　変調スペクトルを用いた環境音の分析

◎石田妙子，荒井隆行，村原雄二（上智大・理工）

1. はじめに

我々の周りを取り巻く環境音は、生活に障害を与えるような騒音の他、さまざまな機器やシステムの性能劣化させる雑音として捉えることもできる[1]。例えば、携帯電話が持つ音声認識の機能を野外で使用する場合を考える。その際、認識の対象となる音声信号は周囲の環境音によって変形され、認識率が低下すると同時にシステムの性能が劣化する[1]。このような場合の対処法として、雑音に対してロバストな認識アルゴリズムの研究[3]などが多数なされている。

そのようなロバストな音声認識システムの構築を目的に開発された特徴量抽出法の一つとして、変調スペクトル (modulation spectrum) によるものがあるが、そこでは環境音の影響を軽減し認識性能を改善できることが報告されている[2]。この変調スペクトルは、対象とする音響信号に対し複数の帯域ごとにエネルギーの時間変化を求め、その時間変化をさらに周波数領域で表現するというものであり、音声情報[4]や話者情報[5]を担っていることが知られている。特に、音声認識に際しては、特定の変調周波数成分を特徴量として用いることによって、認識率の改善を実現している[2]。

そこで本研究では、この変調スペクトルを環境音についても適用し、異なった数種類の環境音が変調スペクトル上でどのような違いを生み出すかについて比較、検討する。

2. 原理

変調スペクトルは、対象とする音響信号に対しそれぞれの帯域ごとにエネルギーの時間変化（時間包絡）を求め、その時間変化をさらに周波数領域で表現することによって得られる。分析対象とする環境音に対し、図1のブロックダイアグラムに沿って各帯域における変調スペクトルを求めた。

まず、8 kHz でサンプリングされた入力信号に対しHamming窓によってフレーム分け（$N = 256$）を施し、フレームは75%オーバーラップするようにシフトする。そしてそれぞれのフレームにおいて、256点高速フーリエ変換 (FFT) を行う。次に、その結果として得られるスペクトルを大きく4つの帯域（第1帯域: 0～500 Hz, 第2帯域: 500～1000 Hz, 第3帯域: 1000～2000 Hz, 第4帯域: 2000～4000 Hz）に分割し、各帯域内でエネルギーの総和を求めた。これにより、各帯域内におけるエネルギーの時間軌跡が得られる。これを各帯域ごとに時間方向にFFTを施しdB値に変換することによって変調スペクトルを求めた。なお、入力信号は実効値で正規化した。

3. 結果

分析対象とする環境音データには、NTTアドバンストテクノロジーの “Ambient Noise Database for Telephonometry 1996” の中から鉄道ガード下道路、高速道路、商店街の環境騒音を用いた。Fig.2にこれら3種類の環境音に対して各帯域ごと求めた変調スペクトルを示す。

4. 考察

3種類のそれぞれの環境音について、その主な音源を以下に示す：

- 鉄道ガード下道路…電車の走行音、振動音、車の走行音
- 高速道路…車の走行音
- 商店街…通行人のざわめき声、足音

それぞれの環境音に関して得られた変調スペクトルを比較したところ、似たような傾向を示したものの、詳細では次のような違いが確認された。それぞれの環境音について各帯域ごとにそのエネルギーの

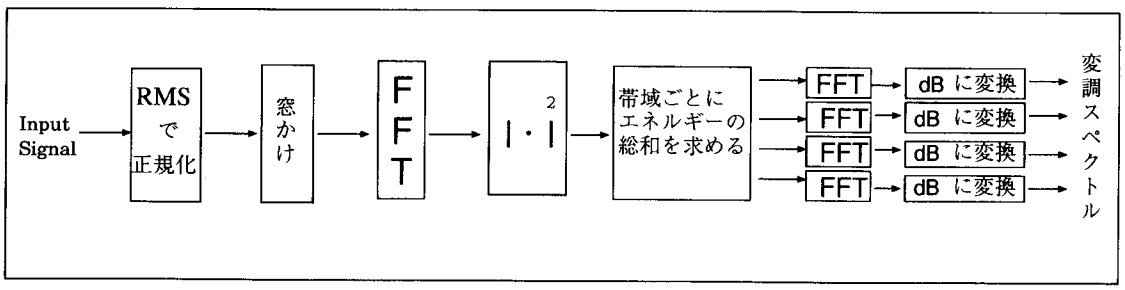

Figure 1: Block diagram.

---

* Analysis of Environmental Noises using the Modulation Spectrum

By Taeko Ishida, Takayuki Arai and Yuji Murahara (Sophia Univ.)

総和を求め、周波数の低い方から50%に達する周波数を求めたところ、平均で約15 Hzとなった。そこで、0～15 Hzの変調周波数域における変調スペクトルの傾きを比較した結果、高速道路ではどのバンドにおいてその傾斜が最も緩やかであった。また、商店街ではその傾斜が最も急であったことから、ゆっくりとした時間変化の成分が多いことが分かった。ガード下道路では変調スペクトル上の帯域幅も広く、エネルギーの時間変化の速い成分も多く含まれることが分かった。

5. おわりに

3種類の環境音に対しその変調スペクトル表現を求め、それらを比較、検討した。その結果、変調スペクトル上の形状は似たような傾向が確認されたものの、その帯域幅や傾きにそれぞれの特徴を捉えることができた。

参考文献

[1] 飛田瑞広, 管村昇, "音声認識における周囲環境の影響," 日本音響学会誌, 51巻4号, pp.331-335, 1995.

[2] N. Kanedera, T. arai, H. Hermansky, and M. Pavel, "On the relative importance of various components of the modulation spectrum for automatic sppech recognition," Speech Communication, vol. 28, pp.43-55, 1999.

[3] H. Hermansky, N. Morgan, "RASTA Processing of speech," IEEE Trans. on Speech and Audio Processing, Vol.2, No.4, pp.578-589, 1994.

[4] T. Arai, M.Pavel, H. Hermansky, and C. Avendano, "Syllable intelligibility for temporally filtered LPC cepstral trajectories," The Journal of the Acoustical Society of America, vol.105, No.5, pp.2783-2791, 1999.

[5] T. Arai, M. Takahashi, and N. Kanedera, "On the important modulation-frequency bands of speech for human speaker recognition", ICSLP, vol.3, pp.774-777, 2000.

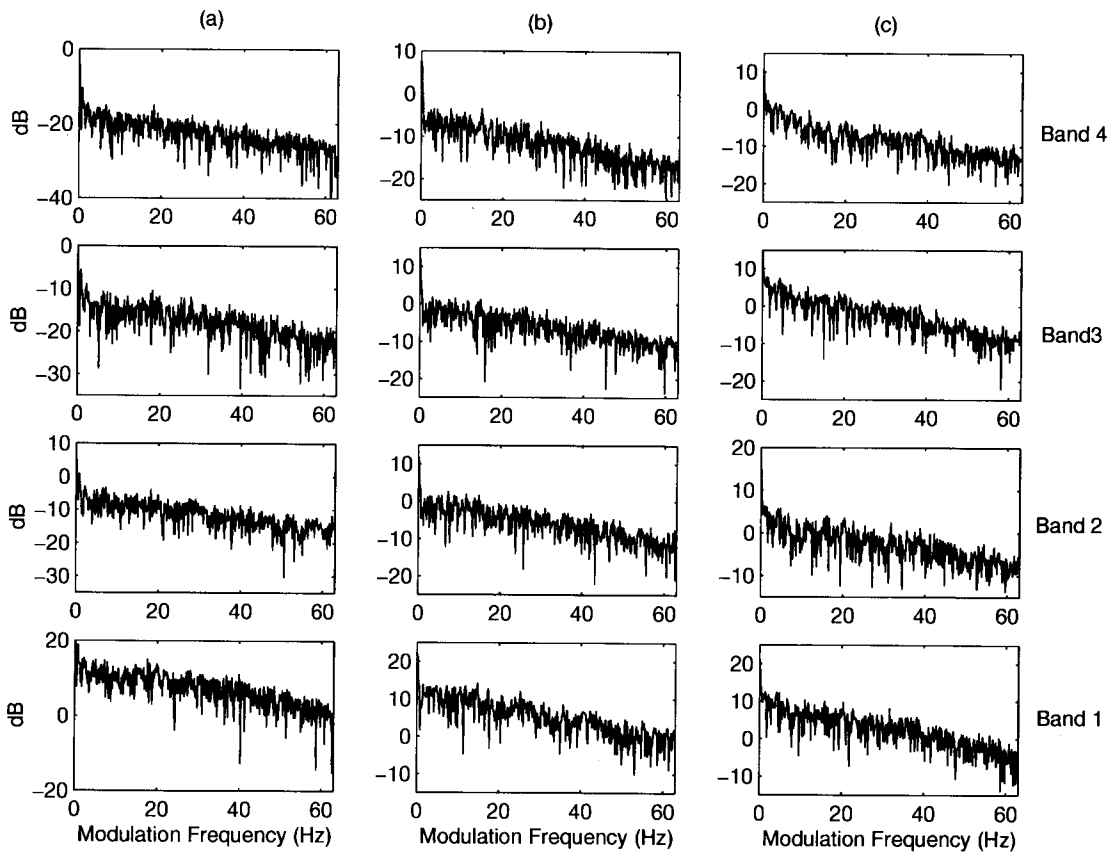

Figure 2: 変調スペクトル (a):鉄道ガード下道路、 (b):高速道路、 (c):商店街