

最大エントロピー法による音声スペクトルの推定

石原 学・白滝 順・家入 勝吾

Estimation of Speech spectrum by Maximum Entropy Method

Manabu ISHIHARA, Jun SHIRATAKI and Shogo IEIRI

Abstract

In this paper we describe a computer simulation study of Maximum entropy method (MEM). When rearranging the computer simulation results,

- (1) We can estimate speech spectrum from MEM.
- (2) Apparently FPE (Final Prediction Error) low coefficient is also true of the case of Power Spectrum.

1. はじめに

音声のホルマント周波数の推定法には、分析合成によって求める A-b-S 法や、ピーク値検出と、その 2 次式の内挿法により、近似的にホルマント周波数と、その帯域幅を求めるピークピッキング法など多数が提案されている^{1,2)}。しかし、A-b-S 法は、ホルマント周波数の正確な推定ができる反面、計算時間が長いという欠点がある。また、ピークピッキング法は、スペクトルに基づく推定法であり、推定過程が比較的簡単で高速推定が可能であるために、現在もよく用いられている³⁾。しかし、ピークピッキング手法は、ホルマント周波数が接近した場合には、推定不可能となることが指摘され、その解決方法の一つが発表⁴⁾されている。

本研究では、最大エントロピー法によって音声スペクトル周波数の推定を行った。最大エントロピー法は、情報エントロピーを最大にするようなスペクトルが最も確からしいスペクトルと考えたものである。この最大エントロピー法を利用してパーソナルコンピュータにより音声ホルマント周波数の推定を行った。その結果、音声ホルマント周波数がある程度表現できると考えられる。このとき、サンプリング数を 500, 700, 800, 1000, 1200 と変化させたが、音声ホルマント周波数は、それぞれについて推定を行うことができる。しかし、FPE 次数の設定によってスペクトルの振幅が不安定

であり、この点が今後の課題である。

2. MEM スペクトル推定法⁷⁾

Burg は、Wiener-Khintchin の関係を制約条件とした上で情報エントロピーを最大とするようなスペクトルが最も確からしいスペクトルであるとして、スペクトル推定法を提案した⁵⁾。これが、最大エントロピー法 (MEM) である。また、赤池が Burg より 2 年あとに、自己回帰モデルに基づくスペクトル推定法を提案した⁶⁾。これらの 2 つの方法では、理論的な導出の仕方は異なるが、結果的には、ほぼ同じアルゴリズムによる推定法となっている。本論文では、赤池法による MEM 理論で分析を行っている。

赤池法の理論において重要な点は、観測データに対し“自己回帰モデル”を仮定することにある。自己回帰モデルは、

$$x_k = -\sum_{i=1}^m a_{mi} x_{k-i} + n_k \quad (1)$$

ただし、 x_k は観測された時系列データ

n_k は $x_l (l < k)$ と独立な定常白色雑音

m は自己回帰モデル次数

a_{mi} は次数 m における自己回帰係数

ここで、時系列データ x_k の自己相関関数を

$$R_i = R(i\Delta t) \equiv E\{x_k x_{k-i}\} \quad (2)$$

ただし、 $E\{\}$ は期待値

と表し、式(1)の両辺にそれぞれ x_k を掛けて期待値をとると

$$\begin{aligned}
 R_0 &= E\{x_k^2\} \\
 &= -\sum_{i=1}^m a_{mi} \cdot E\{x_k x_{k-i}\} + E\{x_k n_k\} \\
 &= -\sum_{i=1}^m a_{mi} R_i + E\{n_k^2\}
 \end{aligned} \tag{3}$$

が得られる。

n_k は $x_i (i < k)$ とは独立であるから、式(1)の両辺にそれぞれ $x_{k-1}, x_{k-2}, \dots, x_{k-m}$ を掛けて期待値をとると、

$$\begin{bmatrix} R_0 & R_1 & \dots & R_m \\ R_1 & R_0 & & R_{m-1} \\ \vdots & & \ddots & \vdots \\ R_m & \dots & \dots & R_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_{m1} \\ \vdots \\ a_{mm} \end{bmatrix} = \begin{bmatrix} P_m \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{4}$$

ただし、 P_m は定常白色雑音の分散なる行列方程式が得られる。式(4)に Wiener-Khinchine の公式を適用すると

$$S(\omega) = \frac{P_m \Delta t}{|1 + \sum_{i=1}^m a_{mi} e^{-j\omega i \Delta t}|^2} \tag{5}$$

が得られる。式(5)より、自己回帰モデル $\{a_k\}$ とパワー・スペクトル $S(\omega)$ の関係が求められる。

また、最大エントロピー法では推定の次数をいくつにとるかが問題となる。通常用いられる規準は赤池の FPE (Final Prediction Error) である。ここでは FPE を次式で定義する。

$$Q = \left(1 + \frac{m+1}{N}\right) \left(1 - \frac{m+1}{N}\right)^{-1} E_m^2$$

ただし、

$$E_m^2 = \sum_{k=m+1}^N \left(x_k - \sum_{i=1}^m \hat{a}_{mi} x_{k-i}\right)^2$$

FPE が最小値を取るような m で漸化式を打ち切る手法が取られるが、 x_k の値によっては明瞭な極小を取らない場合がある。しかし極端な高次の推定を行った場合、まぎらわしいにせのスペクトルピークを作ることがあり、極小を示さない場合でもある程度以上推定次数を上げない方が有益である。この目安は実数データに対するスペクトル推定の場合、データ点数の半分程度とされている⁷⁾。本論文では、データ点数を変化させた場合も、実験を行った。

3. 実験および検討

本シミュレーション実験のフローチャートを Fig. 1 に示す。

また、音声データとしては、著者らの開発した音声分析システム⁸⁾によって得られた単音節データを利用した。この音声データは、12 bit の A/D 変換器を利用し、サンプリング周波数 10 kHz で入力されたデータである。

本論文では分析の一例として単音節データの単母音 /a/ について述べる。

サンプリング間隔は、0.1 ms とした。また音声データのサンプリング数は、希望サンプリング数を入力することにより、パーソナルコンピュータが自動的に

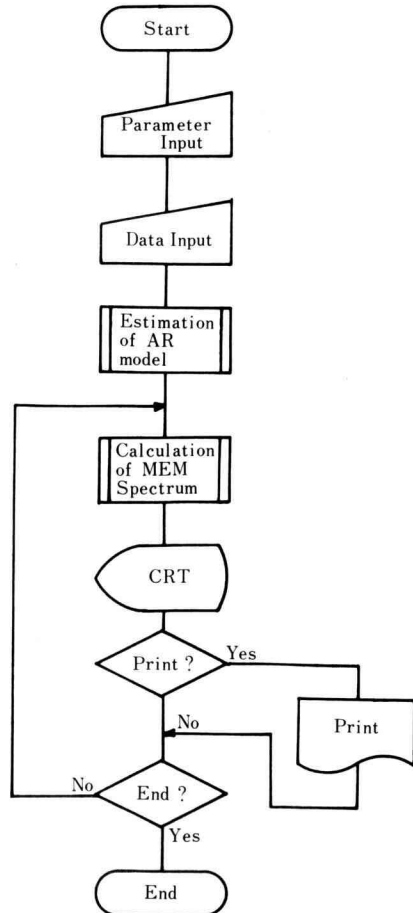


Fig. 1. Flow chart of program for MEM spectrum.

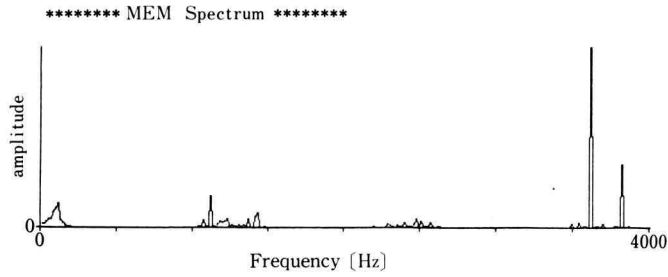


Fig. 2. MEM SPECTRUM of vowel /a/. (Sample points 500 and FPE 6).

データからカウントする。このとき、音声データのサンプリング数をすべて使用すると演算時間が長くなるおそれがある。そこで、このサンプリング数をある程度、間引きをして演算を行った。例えば、この /a/ のサンプリング数を実際に全データとして利用すると約 3000 程度であり、このサンプリング数で演算を実行すると、PC-9801/VX 21, クロック周波数 10 MHz, ノンウェートで動作させても、FPE 次数の設定によっては数時間の演算時間を必要とする。

ここでは、サンプリング数を 500, 700, 800, 1000, 1200 の 5 種類について FPE 次数を変えて実験を行っ

た。

3.1 サンプリング数“500”の場合

サンプリング数を 500 点と設定し、FPE 次数を変化させて単音節 /a/ のスペクトルを推定した。そのときの周波数は、0~4 kHz の範囲である。

FPE 次数を 6 および 8 としたときのスペクトル推定を、それぞれ Fig. 2 および 3 に示す。これらの図に見られるように、音声ホルマント周波数がよく表現されている。1 kHz 付近, 2.5 kHz 付近, 3.7 kHz 付近にホルマント周波数があることがわかる。

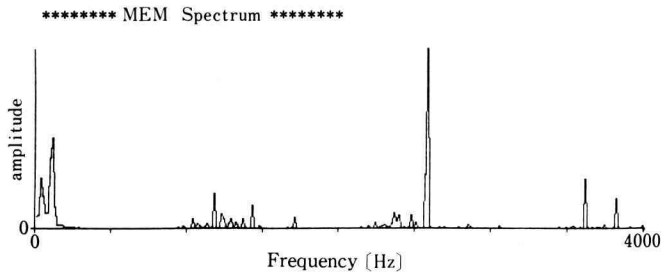


Fig. 3. MEM SPECTRUM of vowel /a/. (Sample points 500 and FPE 8).

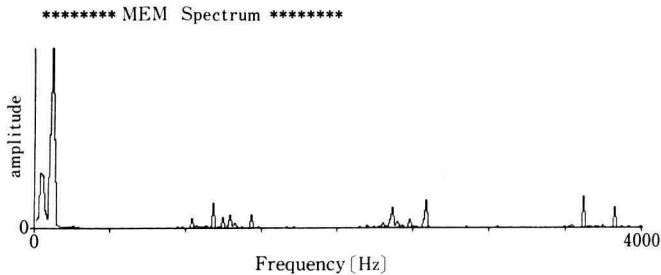


Fig. 4. MEM SPECTRUM of vowel /a/. (Sample points 800 and FPE 10).

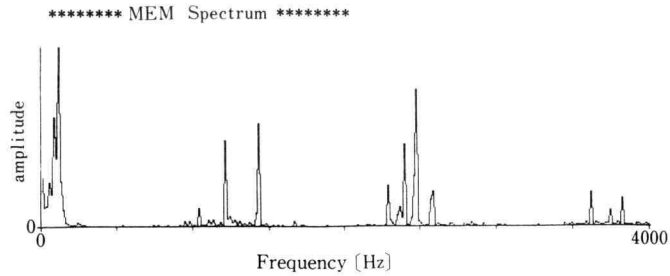


Fig. 5. MEM SPECTRUM of vowel /a/. (Sample points 800 and FPE 12).

3.2 サンプリング数“800”の場合

サンプリング数を 800 点と設定し、FPE 次数を変化させて単音節 /a/ のスペクトルを推定した。そのときの周波数は、0~4 kHz までの範囲である。

FPE 次数を 10 および 12 としたときのスペクトル推定を、それぞれ Fig. 4 および 5 に示す。

これらの図に見られるように、音声ホルマント周波数が、1 kHz 付近、2.5 kHz 付近、3.7 kHz 付近にあることがわかる。特に、FPE 次数が 12 の場合には、ホルマント周波数がよく表現されている。

3.3 サンプリング数“1000”の場合

サンプリング数を 1000 点と設定し、FPE 次数を変化させて単音節 /a/ のスペクトルを推定した。そのときの周波数は、0~4 kHz までの範囲である。

FPE 次数を 6 および 8 としたときのスペクトル推定を、それぞれ Fig. 6 および 7 に示す。Fig. 6 および 7 に見られるように、音声ホルマント周波数が、1.2 kHz 付近、2.5 kHz 付近、3.7 kHz 付近にあることがわかる。

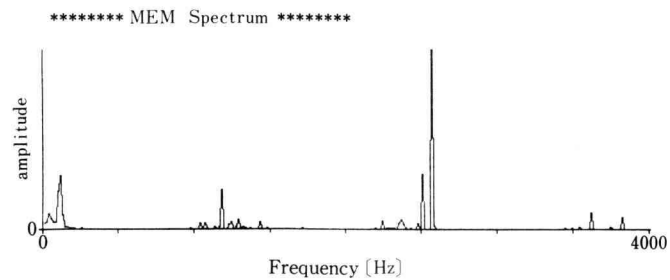


Fig. 6. MEM SPECTRUM of vowel /a/. (Sample points 1000 and FPE 6).

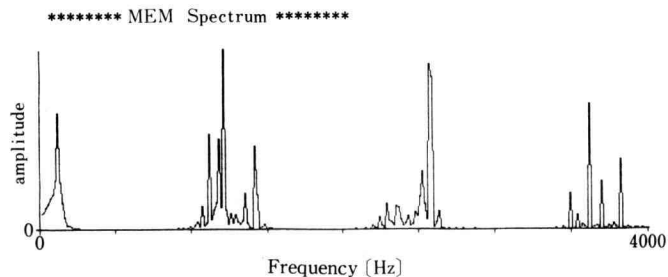


Fig. 7. MEM SPECTRUM of vowel /a/. (Sample points 1000 and FPE 8).

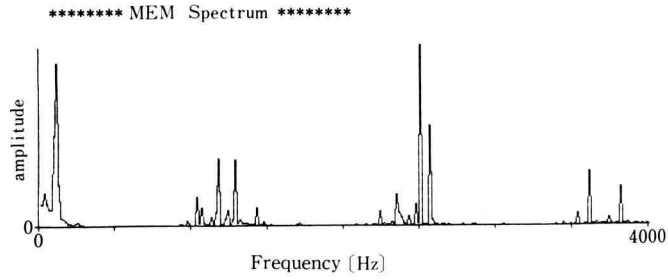


Fig. 8. MEM SPECTRUM of vowel /a/. (Sample points 1200 and FPE 8).

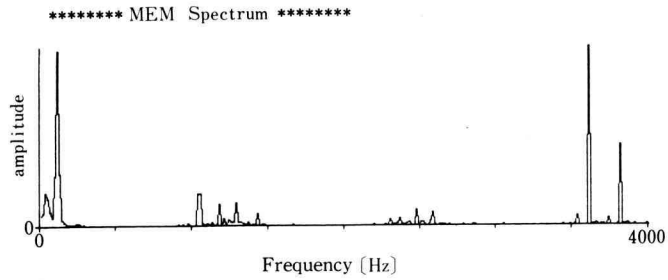


Fig. 9. MEM SPECTRUM of vowel /a/. (Sample points 1200 and FPE 10).

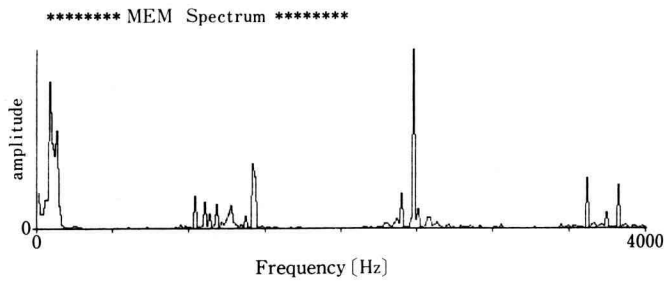


Fig. 10. MEM SPECTRUM of vowel /a/. (Sample points 1200 and FPE 12).

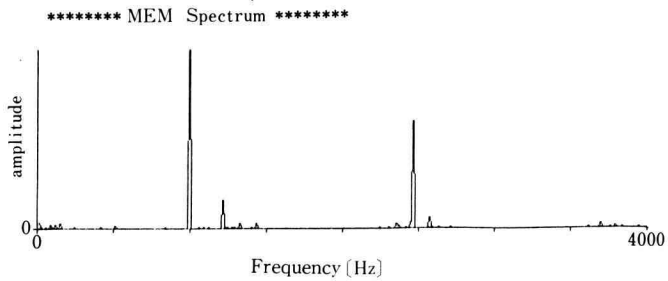


Fig. 11. MEM SPECTRUM of vowel /a/. (Sample points of 1000 and FPE 500).

3.4 サンプリング数“1200”の場合

サンプリング数を1200点と設定し、FPE次数を変化させて単音節 /a/ のスペクトルを推定した。そのときの周波数は、0~4 kHz までの範囲である。

FPE 次数を8, 10 および12としたときのスペクトル推定を、それぞれ Fig. 8, 9 および10 に示す。

このサンプリング数においても、音声ホルマント周波数が、1.2 kHz 付近, 2.5 kHz 付近, 3.7 kHz 付近にあることがわかる。

3.5 FPE 次数がサンプリング数の半分程度の場合

サンプリング数が1000, FPE 次数を500と設定したときの最大エントロピー法によるスペクトル推定を行った。そのときのスペクトルを Fig. 11 に示す。ここで示されるように、1 kHz 付近でスペクトル表示があり、これがホルマントと考えられる。また、1.2 kHz 付近にやはりスペクトル表示が見うけられる。これもホルマントと考えられ、周波数についてよく分離できている。

4. あとがき

最大エントロピー法を用いて、音声信号のパワースペクトルの推定を試みた。その結果、音声ホルマント周波数の推定が可能であることを示した。また、これは線形予測分析に用いられているシステムモデルの自己回帰過程と等価であるから、全極モデルの音声スペクトル推定には有効であると考えられる。また、線形予測法で音声スペクトルを推定する場合には、線形予測係数を10程度として経験的に求めているが、本方法では最大モデル次数を決定することにより自動的に自己回帰モデル次数を求めることができる。

また、最大エントロピー法は、スペクトル推定が主であって、振幅については保証されていないために、大

きさの関係が不明瞭である。

今後、最大推定次数の決定数を最適にすること、および演算時間と推定精度の関係を明らかにすることが課題である。

最後に、本研究の一部に本学昭和62年度学術研究費を充当使用したのでここに記して謝意を表す。

参 考 文 献

- 1) 例えば、三浦種敏監修：“新版 聴覚と音声”，pp. 323-329, 電子通信学会（昭和55年）。
- 2) Markel, Gray：“音声の線形予測，鈴木久喜訳”，第1~5章，コロナ社（昭和55年）。
- 3) 安居院，中嶋：“コンピュータ音声処理”，第2~3章，産報出版（昭和55年）。
- 4) 鎌田，石田，小川：“指定した偏角付近にある極のみ推定する方法によるホルマントに対応した極の高速推定について”，信学論（D），J70-D, No. 2, pp. 376-386（昭和62年2月）。
- 5) Burg, J.P.：“Maximum Entropy Spectral Analysis”，presented at the 37th Annu. meeting Soc. of Exploration Geophysicists, Oklahoma City, OK.
- 6) Akaike, H.：“Fitting Autoregressive Models for Prediction”，Annals of Institute Statistical Mathematics, vol. 21, pp. 243-247（1969）。
“Power Spectrum Estimation Through Autoregressive Model Fitting”，Annals of Institute Statistical Mathematics, vol. 21, pp. 407-419, 1969.
- 7) 南 茂夫編著：“波形データ処理”，第9章，pp. 166-180, CQ 出版社（昭和61年）。
- 8) 白滝，石原：“パーソナルコンピュータを利用した音声分析システム”，幾徳工業大学研究報告B，第11号，pp. 127-131（昭和62年）。