

# 最大エントロピー法による音声スペクトルの推定 (第2報)

石原 学・白滝 順・家入 勝吾・志方 泰\*

Estimation of Speech Spectrum by Maximum Entropy Method (2nd Report)

Manabu ISHIHARA, Jun SHIRATAKI, Shogo IEIRI and Yutaka SHIKATA\*

## Abstract

Using the computer simulation, we have studied the relation between the numbers of terms and those of data points of the prediction error filter, those are parameters in the Maximum Entropy Method (MEM).

We have obtained the experimental results that (1) the computer simulation gives the satisfactory agreement with the speech spectrum and (2) MEM is more available than FFT for the spectrum estimation when data points are small in numbers.

## 1. はじめに

われわれは、最大エントロピー法を用いた音声スペクトルの推定についてすでに報告している<sup>1)</sup>。今回、音声スペクトルを求めるときにサンプル数と推定次数を変えて計算機シミュレーションを行って音声スペクトルを求めた。最大エントロピー法のスペクトル推定次数を変えて求めた音声スペクトルとFFTで求めた音声スペクトルとを比較検討したので報告する。

## 2. MEM スペクトル推定法<sup>2)</sup>

Burg は、Wiener-Khintchine 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。これが、最大エントロピー法 (MEM) である。また、赤池は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。これらの2つの方法は、理論的な導出の方法に

ついては異なるが、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。赤池の理論において重要な点は、観測データに対して“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、次式のように表されている。

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

ただし、

$x_k$  = 観測された時系列データ  
 $n_k = x_l (l < k)$  と独立な定常白色雑音  
 $m$  = 自己回帰モデル次数  
 $a_{mi}$  = 次数  $m$  における自己回帰係数

ここで、時系列データ  $x_k$  の自己相関関数は式 (2) のように表される。

$$R_i = R_{(idt)} = E \{ x_k x_{k-i} \} \quad (2)$$

ただし、 $E \{ \dots \}$  = 期待値、

式 (1) の両辺にそれぞれ  $x_k$  を掛けて期待値をとると、

平成元年 10 月 2 日受理

\* Department of Electrical Engineering, Faculty of Science and Engineering, Meisei University.

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

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  $x_{k-1}, x_{k-2}, \dots, x_{k-m}, \dots$  を掛けて期待値をとると、

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

ただし、 $R_m =$  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、

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

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。通常用いられる基準は赤池の最終予測誤差 (FPE: Final Prediction Error) である。FPE は次式で定義されている。

$$Q = \left(1 + \frac{m+1}{N}\right) \left(1 - \frac{m+1}{N}\right)^{-1} E_m^2 \tag{6}$$

ただし、

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

$\hat{a}_{mi}$  は時系列  $x_k$  から次数  $m$  で推定した自己回帰係数 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、まぎらわしいにせよ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、ある程度以上に推定次数を上げない方が有利である。この目安は実数データに対するスペクトルの場合、データ点数の半分程度とされている。われわれは、データ点数を変化させて音声スペクトルの推定を行った。以下に、計算機シミュレーションによって得られた結果について報告する。

### 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、さらに、詳細なフローチャートを付録を示す。また、プログラムは C 言語 (MS-C) で作成されている。

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<sup>3)</sup>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 の変換器を利用し、サンプリング周波数 10 kHz で入力されたデータである。

分析の一例として単音節データの単母音/a/を用いた。

サンプリング間隔は、0.1 ms とした。また音声データの分析データ数は、希望のデータ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計数を行う。このとき、音声データ数と推定次数を変化させて求めた音声スペクトルと FFT で求めたスペクトルとを比較検討した。

以下に、スペクトル推定に関する実験結果について述べる。

#### 3.1 予測誤差フィルタの次数とデータ数の関係

—データ数が 128 の場合—

ここでは、予測誤差フィルタの次数を 5, 10, 15 と変

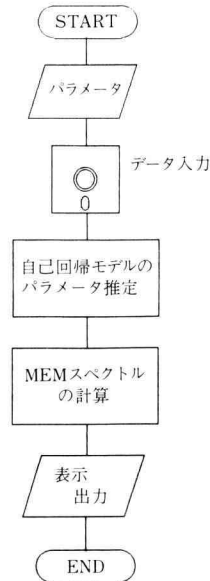


Fig. 1 Flowchart of Simulation

化させた場合の結果とFFTとの関係について検討した。Fig. 2(1)は、予測誤差フィルタの次数が5, 10, 15のときのスペクトルを示している。Fig. 2(2)は、MEMとFFTから得られたスペクトルを同時に示している。予測誤差フィルタの次数が5, 10のときには、1 kHz付近にスペクトルのピークが現れているが、予測誤差フィルタの次数が15のときには、2つのスペクトルのピークが得られている。

#### —データ数が256の場合—

予測誤差フィルタの次数を5, 10, 15と変化させた場合の結果とFFTとの関係について検討した。Fig. 3(1)は、予測誤差フィルタの次数が5, 10, 15のときのスペクトルを示している。Fig. 3(2)は、MEMとFFTから得られたスペクトルを示している。データ数128の場合に得られたスペクトルより、スペクトルのピーク部分がよりはっきり現れている。

#### —データ数が512, 1024, 2048の場合—

予測誤差フィルタの次数が5, 10, 15と変化させた場合の結果とFFTとの関係について検討した。Fig. 4(1)は、予測誤差フィルタの次数が5, 10, 15のときの

スペクトルを示している。Fig. 4(2)は、MEMとFFTから得られたスペクトルを同時に表している。データ数256のときに得られたスペクトルより、ピーク部分がはっきりと現われている。さらにデータ数を増加させると、スペクトルの位置がはっきり現われる。しかし、データ数を増加させると演算時間が長くなる。データ数が512程度でもスペクトルが良く現われると考えられる。データ数が1024の場合のスペクトルをFig. 5(1)および(2)に、また、データ数が2048の場合のスペクトルをFig. 6(1)および(2)にそれぞれ示す。

予測誤差フィルタの次数は15程度あればFFTよりも良くスペクトルが表現できると考えられる。データ数を変化させた場合のMEMとFFTとのスペクトルを比較するために、両スペクトルをFig. 7(1)~(5)を示す。

#### 3.2 予測誤差フィルタの次数を大きくした場合

予測誤差フィルタの次数をそれぞれ、20, 40, 60と大きくしてシミュレーションを行うと、予測誤差フィルタの次数が大きくなれば、情報のエントロピーの推

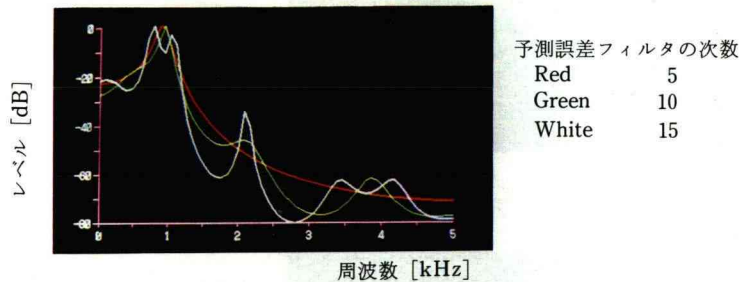


Fig. 2(1) Speech Spectrum estimated by MEM (DATA 128)

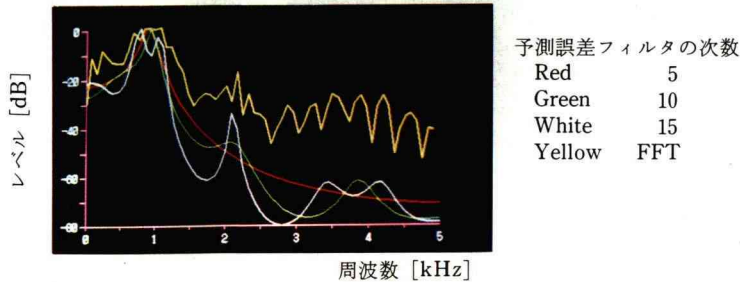


Fig. 2(2) Speech Spectrum estimated by MEM and FFT (DATA 128)

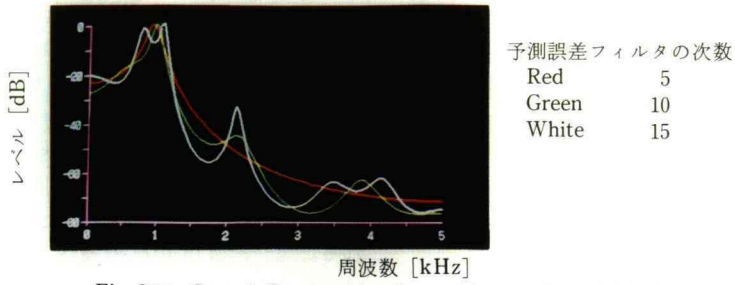


Fig. 3(1) Speech Spectrum estimated by MEM (DATA 256)

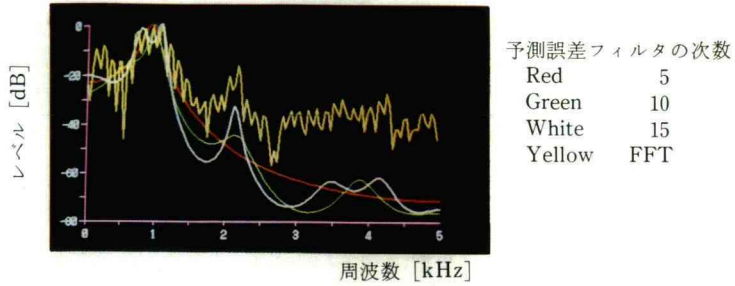


Fig. 3(2) Speech Spectrum estimated by MEM and FFT (DATA 256)

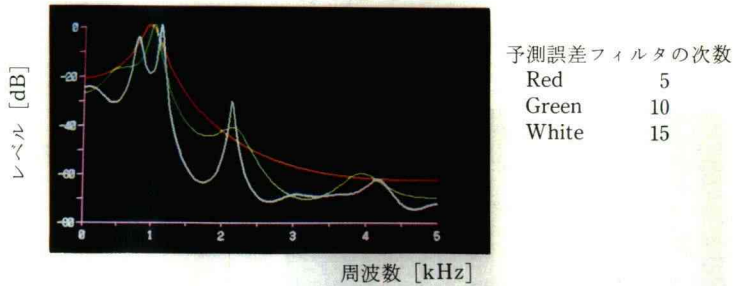


Fig. 4(1) Speech Spectrum estimated by MEM and FFT (DATA 512)

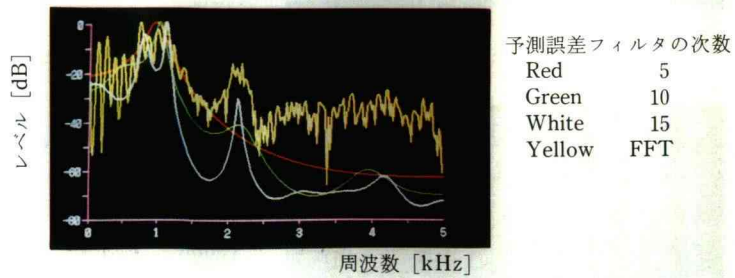


Fig. 4(2) Speech Spectrum estimated by MEM and FFT (DATA 512)

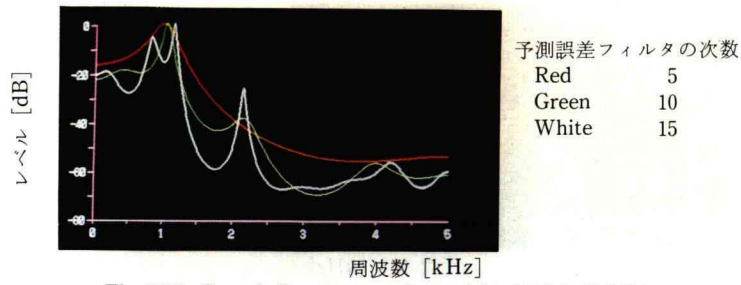


Fig. 5(1) Speech Spectrum estimated by MEM (DATA 1024)

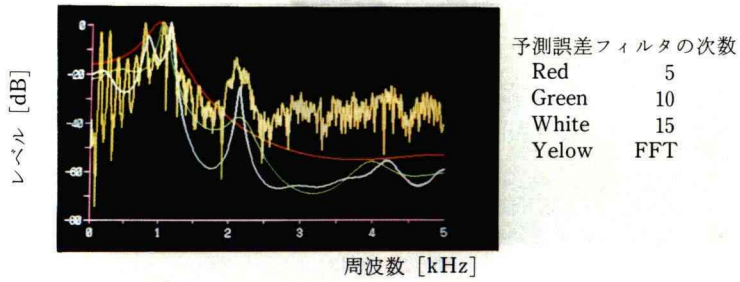


Fig. 5(2) Speech Spectrum estimated by MEM and FFT (DATA 1024)

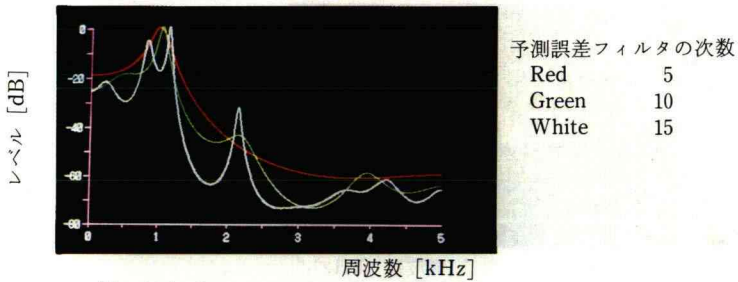


Fig. 6(1) Speech Spectrum estimated by MEM (DATA 2048)

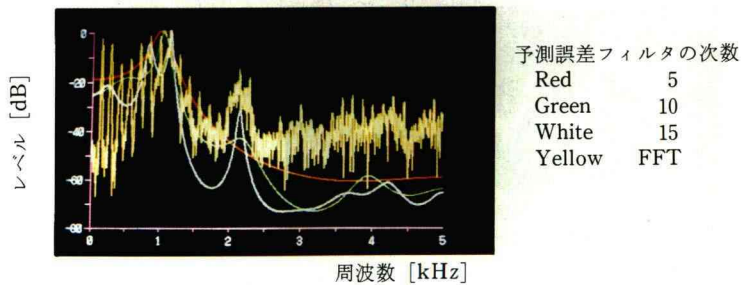


Fig. 6(2) Speech Spectrum estimated by MEM and FFT (DATA 2048)

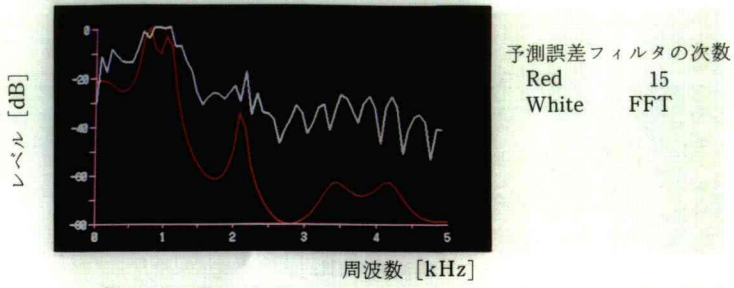


Fig. 7(1) Speech Spectrum estimated by MEM and FFT (DATA 128)

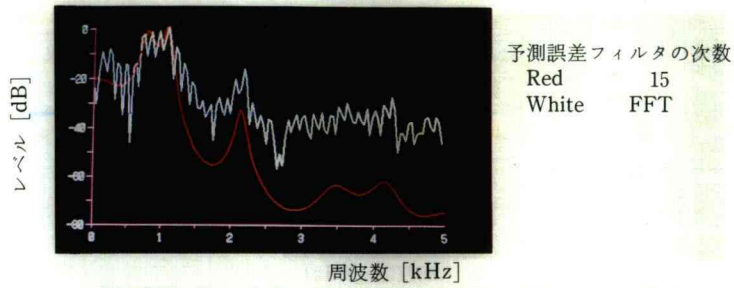


Fig. 7(2) Speech Spectrum estimated by MEM and FFT (DATA 256)

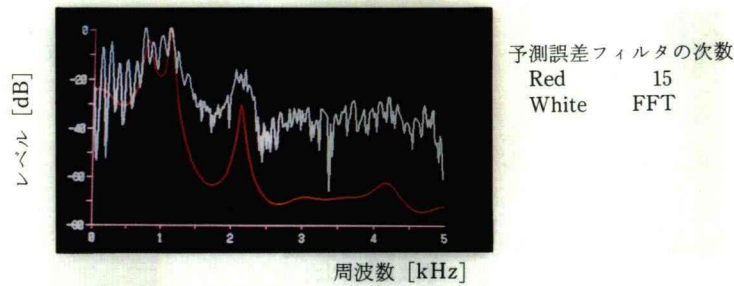


Fig. 7(3) Speech Spectrum estimated by MEM and FFT (DATA 512)

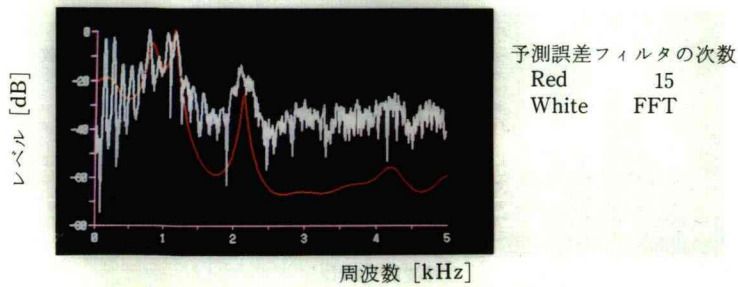


Fig. 7(4) Speech Spectrum estimated by MEM and FFT (DATA 1024)

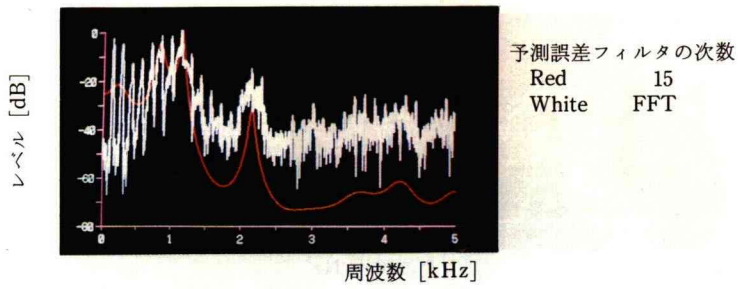


Fig. 7(5) Speech Spectrum estimated by MEM and FFT (DATA 2048)

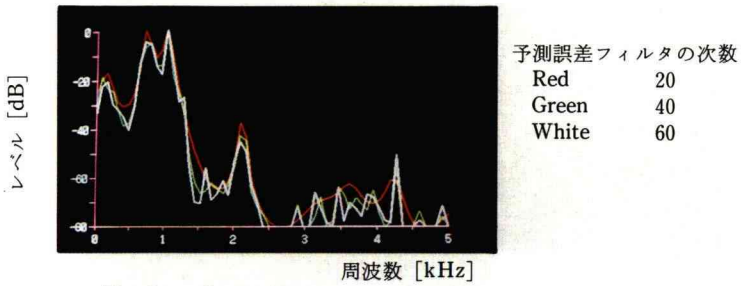


Fig. 8(1) Speech Spectrum estimated by MEM (DATA 128)

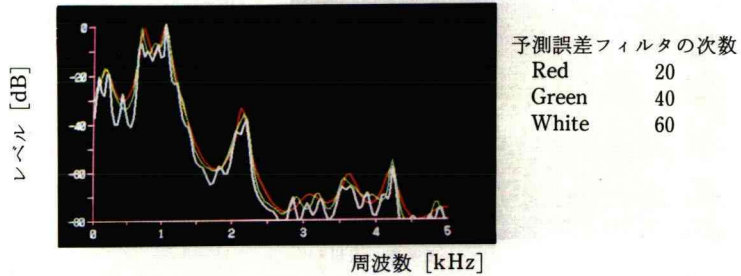


Fig. 8(2) Speech Spectrum estimated by MEM (DATA 256)

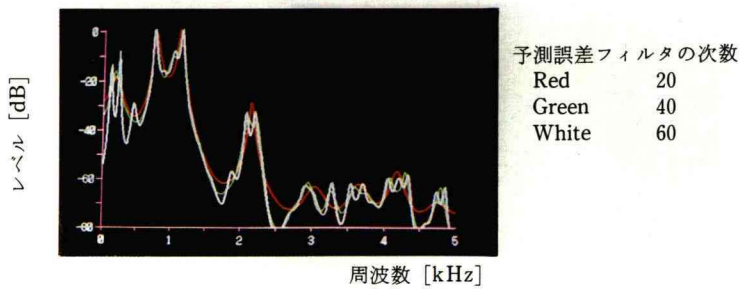


Fig. 8(3) Speech Spectrum estimated by MEM (DATA 512)

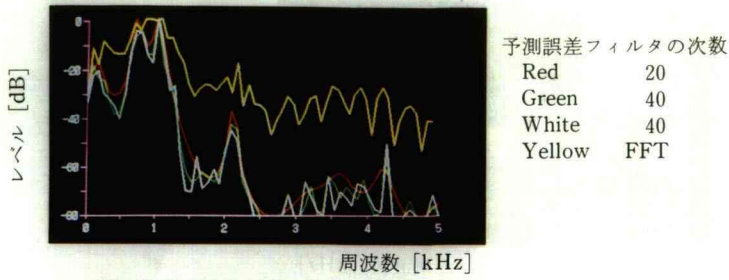


Fig. 8(4) Speech Spectrum estimated by MEM and FFT (DATA 128)

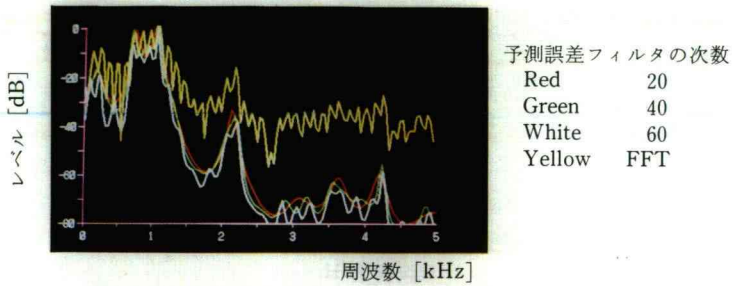


Fig. 8(5) Speech Spectrum estimated by MEM and FFT (DATA 256)

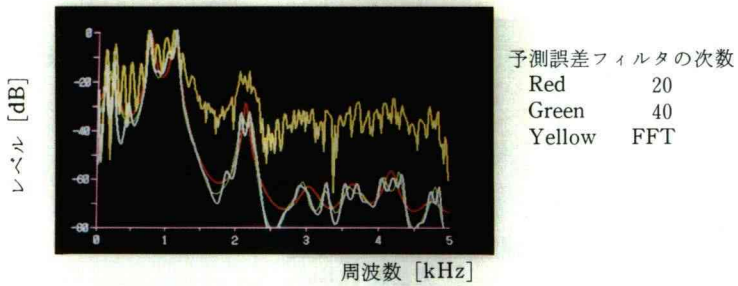


Fig. 8(6) Speech Spectrum estimated by MEM and FFT (DATA 512)

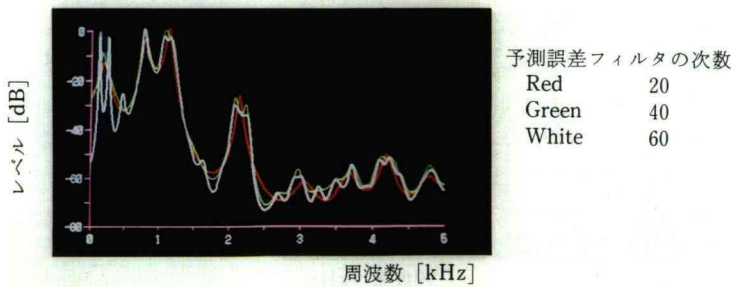


Fig. 9(1) Speech Spectrum estimated by MEM (DATA 1024)



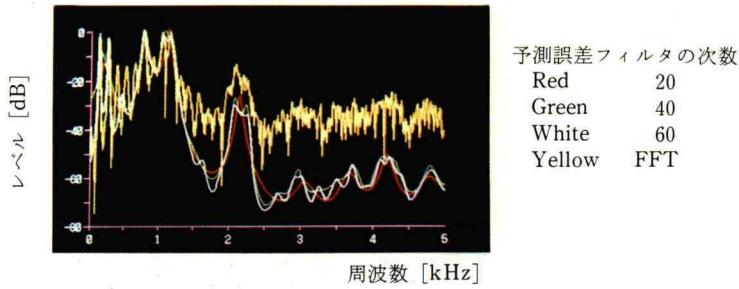


Fig. 9(2) Speech Spectrum estimated by MEM and FFT (DATA 1024)

定が正確になる。しかし、音声のスペクトルを表現する場合には、情報の位置がはっきりしすぎ不向きと考えられる。データ数が128, 256および512で、かつ予測誤差フィルタの次数が20, 40および60の場合について、MEMによって推定された音声スペクトルをそれぞれFig. 8(1)~(3)に、またFFT法によって求められたスペクトル、MEMによるスペクトルとともにFig. 8(4)~(6)に示している。さらに、データ数が512以上になると、3 kHz以上の周波数で細かいスペクトルが現れてくる。Fig. 9(1)は、データ数が1024のときで、予測誤差フィルタの次数を20, 40, 60としたときを示す。Fig. 9(2)は、MEMとFFTによって推定された音声スペクトルを示す。

#### 4. おわりに

われわれは、予測誤差フィルタの次数とデータ数の関係について計算機シミュレーションを行った。その結果、データ数が少ない場合には、FFTよりもMEM

がスペクトルの推定には有効であろうと考えられる。また、予測誤差フィルタの次数が大きすぎると、音声スペクトルの特徴は必ずしも有効に表現されないようである。また、音声スペクトルを有効に表現するには、予測誤差フィルタの次数が、15程度あれば、十分であると考えられる。

#### 参考文献

- 1) 石原 学, 白滝 順, 家入勝吾: “最大エントロピー法による音声スペクトルの推定”, 幾徳工業大学研究報告B理工学編, 第12号, pp. 135-140, (昭和63年).
- 2) 例えば, 日野幹雄著: “スペクトル解析”, pp. 210-225, 朝倉書店 (昭和52年)
- 3) 白滝 順, 石原 学: “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理工学編, 第11号, pp. 127-131 (昭和62年).

付 録 Flowchart of Simulation

