準同形フィルタリングによる音声の極—零点推定

横　山　正　人*

Estimation of Poles-Zeros of Speech by Homomorphic Filtering

Masato YOKOYAMA

ABSTRACT

Homomorphic prediction is the speech analysis method which considers influences of the spectral fine structure. However, when the spectral fine structure fluctuates radically, this method cannot eliminate the influences sufficiently.

This paper proposes two new techniques for a improvement of accuracy of poles-zeros estimation by Homomorphic prediction; first for window processing, and second for homomorphic filtering. As a result, the influences of the spectral fine structure are effectively eliminated, and polez-zeros can be estimated with the high accuracy.

英文要旨

ホモモルフィック予測はスペクトル微細構造の影響を考慮した音声分析手法である。しかしながら，スペクトル微細構造が急激に変動する場合，その影響を十分に除去することは不可能である。

本論文ではホモモルフィック予測による極—零点推定に対する精度改善のための2つの新しい手法を提案している。第1の改善は窓処理に対してなされている。第2はホモモルフィックフィルタリングに対してなされている。その結果，スペクトル微細構造の影響を効果的に除去し，高精度で極—零点を推定することが可能となる。

1　まえがき

線形予測（LPC）に基づく音声分析処理技術の発展と共に，音響的特徴としてのパラメータの選択，更には，その精度が問題となってきた。LPCは音声の生成系を全極形（自己回帰〔AR〕）モデルで近似した分析手法であり，明らかに零点が存在する鼻子音等の子音分析では，その音韻の有する特徴を十分に記述することは不可能である。又，声道特性と音源特性の分離が不十分なため，特徴パラメータの抽出精度に大きく影響を及ぼす場合が生じる。

このようなLPCにおける問題点を改善すべく，種々の改良手法が提案されている中で，Oppenheimらは音声生成系を極—零形（ARMA）モデルに拡張し，更に，音源の周期構造成分の除去を考慮したホモモルフィック予測法を提案している[1),2)]。本手法の特徴は，ケプストラムとLPCの2つの概念を組み合せた点にあり，LPCの欠点を改善する有効な手法の1つと考えられる。しかしながら，実際問題として生じるスペクトル微細構造の低レベル部分の急激なレベル変化には，比較的弱い欠点を有する。すなわち，音源の周期構造成分を考慮した手法にもかかわらず，声道インパルス応答を推定する際に影響を受け，極—零点の推定精度を低下させる場合が生じる。

この改善策として，2つのアプローチを想定することができる。第1はホモモルフィックフィルタの前処

*電気工学教室
1985年6月15日受付

理の段階で，事前にスペクトル微細構造の影響を軽減しておく方法である。第2はスペクトル微細構造の急激なレベル変化に追従しないよう，ホモモルフィックフィルタ自体を改良することである。

そこで本論文では，より効果的なホモモルフィック予測法の適用を図るため，上記改善策の具体的方法として2つの改良手法を提案する。まず第1の改良手法として，事前にスペクトル微細構造の影響を軽減するため，分析区間窓に着目した手法を提案する[3]。本手法では分析始点をピッチに同期させ，半 Hamming 窓を音声信号にかけることによって，スペクトルの急激なレベル変化を軽減させている。第2の手法としては，スペクトル微細構造の急激なレベル変化に対する追従を軽減するため，声道インパルス応答を推定する際，平滑化スペクトルが対数スペクトルの頂点近似によって平滑化を行うよう，ホモモルフィックフィルタに修正を施す手法を提案する[4]。又，極－零点を推定するための方法としては，逆フィルタ法による展開を試み精度向上を図る[5),6)]。なお，本論文では合成音，自然音声を用いて，上記改良手法の有効性を明らかにすると共に，極—零分析次数の推定精度への影響度についても検討を行っている。

2　準同形フィルタリング

ホモモルフィック予測は音声信号から直接極—零点の推定を行っているのではなく，準同形処理（ホモモルフィックフィルタ）によって，音源周期構造成分と声道成分との分離 (Deconvolution) を図り，推定された声道インパルス応答（声道成分）を用いて極—零点の推定を行っている手法である[1)]。以下にその処理過程を示す。

いま，標本化周期Tで標本化された音声を$S(n)$とすると，有声音の場合，駆動音源を周期 T_p のパルス列で近似するならば，$S(n)$は式(1)のように表わされる。

$$S(n)=\left[\sum_{m=-\infty}^{\infty}\delta(n-mT_p)\right]*V(n) \qquad (1)$$

$$(n=0,1,2,\cdots\cdots,N-1)$$

但し　$V(n)$　：声道インパルス応答

$\sum_{m=-\infty}^{\infty}\delta(n-mT_p)$　：周期 T_p のパルス列

$*$　：畳込み積分

ここで音声波形$S(n)$に窓関数$W(n)$をかけたものを$X(n)$とすると，窓の長さNを比較的長く，$V(n)$に比べて$W(n)$がゆるやかに変化するものと仮定するならば，式(2)のように近似することができる。

$$X(n)=P_w(n)*V(n) \qquad (2)$$

但し，　$P_w(n)=\left[\sum_{m=-\infty}^{\infty}\delta(n-mT_p)\right]*W(n)$

窓処理された信号$X(n)$を，離散的フーリェ変換したものを$X_f(k)$とすると，式(3)のようになる。

$$X_f(k)=\sum_{n=0}^{N-1}X(n)\exp\left(-j\frac{2\pi kn}{N}\right) \qquad (3)$$

$$(k=0,1,2,\cdots\cdots,N-1)$$

ここで音声生成系を最小位相系と仮定するならば，$X_f(k)$を式(4)のように対数化し，式(5)に示す離散的逆フーリェ変換を施すことによって，ケプストラム$C(n)$を求めることができる。

$$Y(k)=\log|X_f(k)| \qquad (4)$$

$$C(n)=\frac{1}{N}\sum_{k=0}^{N-1}Y(k)\exp\left(j\frac{2\pi kn}{N}\right) \qquad (5)$$

複素ケプストラム$C_p(n)$は最小位相系の仮定によって，式(6)のようになる。

$$C_p(n)=\begin{cases}2C(n) & 1\leqq n<N/2\\ C(n) & n=0,N/2\\ 0 & N/2<n\leqq N-1\end{cases} \qquad (6)$$

複素ケプストラムにおける$P_w(n)$と$V(n)$の寄与は重なることなく，$P_w(n)$成分は周期T_pのパルス列として，又，$V(n)$成分は一般的に早く減衰し，$n=0$付近に存在する。したがって，この低域成分のみを抽出することによって，$P_w(n)$成分を除外した，声道特性のみの抽出が可能となる。ここでは，式(7)のリフタによって声道成分の抽出を行う。

$$V_c(n)=\begin{cases}2C(n) & 0<n\leqq M-1\\ C(n) & n=0\\ 0 & M\leqq n\leqq N-1\end{cases} \qquad (7)$$

なお，Mは低域成分を抽出するためのカットオフ点を表わし，音声のピッチ周期T_pよりも短くなければならない。上記手順に従い，$P_w(n)$と$V(n)$の分離がなされると，以下の処理を行うことにより，声道インパルス応答を推定することができる。まず，式(8)に示すように，$V_c(n)$に離散的フーリェ変換を施す。なお，平滑化スペクトルは，この$Y_f(k)$によって求められる。

$$Y_f(k)=\sum_{n=0}^{N-1}V_c(n)\exp\left(-j\frac{2\pi kn}{N}\right) \qquad (8)$$

更に，式(4)，(5)に対応するよう，$Y_f(k)$を指数化し，離散的逆フーリェ変換することによって，声道インパルス応答数列$V(n)$は推定される。

$$V_f(k)=\exp[Y_f(k)] \tag{9}$$

$$V(n)=\frac{1}{N}\sum_{k=0}^{N-1}V_f(k)\exp\left(j\frac{2\pi kn}{N}\right) \tag{10}$$

$V(n)$が推定されると，2段目の処理である極—零解析を行うことになる。Fig. 1に一連の手順の構成を示す。図中，$D*[\ ]$と$D*^{-1}[\ ]$は逆対の関係にあり，準同形フィルタリングにおける特性システムと逆システムの関係を意味する[1]。

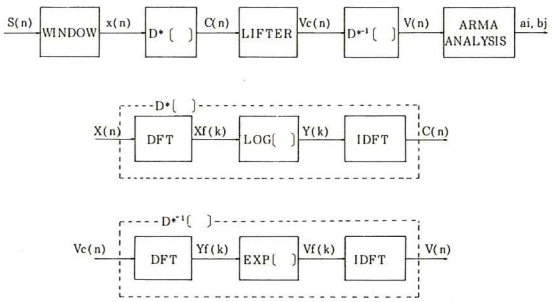


Fig. 1　Block diagram for homomorphic filtering.

3　改善手法 I

3．1　分析アルゴリズム

通常音声波形を分析処理する前に，式(2)に示すように，Hamming 窓等の窓関数による窓処理を行う。すなわち，ピッチ周期と分析区間との相対的位置関係の変動による分析出力の変動を抑制している。この$P_w(n)$と$V(n)$はスペクトル領域においては，式(11)の形で表わされる。

$$\log[X(w)]=\log[P_w(w)]+\log[V(w)] \tag{11}$$

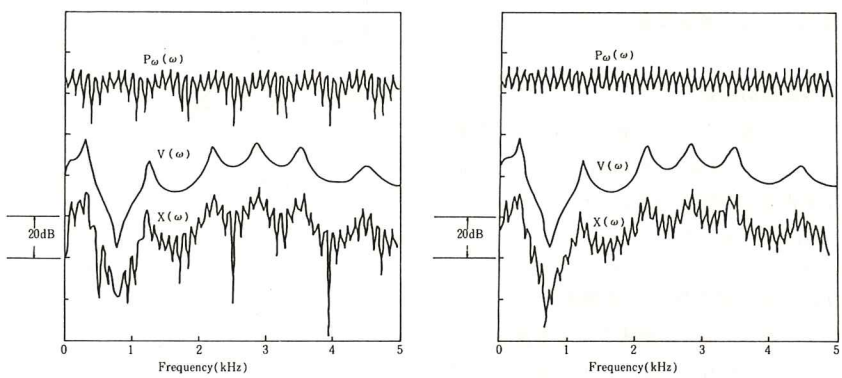
Fig. 2 (a)に Hamming 窓を掛けた合成音信号によるスペクトルの例を示す。スペクトル微細構造成分を表わす$P_w(w)$には，ピッチ周期に対応したパルス的レベル変化と共に，比較的周期性を持った急激なレベル変化が低レベル部分に生じている。この原因は音源のパルス列と区間長との位置関係，すなわち，ピッチ周期と分析区間長，およびその位置関係による変動を窓処理によって十分抑制していないために生じると考えられる。したがって，合成音信号の対数スペクトル$X(w)$にも$P_w(w)$と同様，$P_w(w)$の変化に対応して低レベル部分において急激なレベル変化が生じている。特にスペクトルの谷部と対応している所では，その変化が強調されている。ホモモルフィック予測では，声道インパルス応答を推定する際，ケプストラム処理を行っているため，式(8)の$Y_f(k)$を求める際に影響を受けてしまう。

そこで，上述の点を考慮して新たな窓関数を定義することによって，急激なスペクトルレベル変化を軽減させる。

まず，分析始点を決定するため音声波形$S(i)$よりL個のサンプルを抽出する。次に$S(i)$より，式(12)なる最大ピーク点$H\max, i$を検出する。

$$H\max, i=\max\{S(i)\} \quad [i=0, 1, 2, \cdots\cdots, L-1) \tag{12}$$

すなわち，分析始点として，i番目の信号を用いることによってピッチと同期させる。なお，ここでは最大



(a)　By Hamming window　　(b)　By half Hamming window

Fig. 2　Spectrums of synthetic signal.
($T=0.1$msec, $T_p=9.0$msec, $N=25.6$msec)

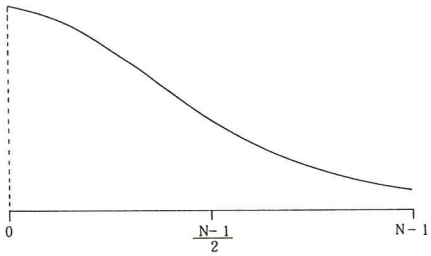


Fig. 3　Window function.
(half Hamming window)

ピッチ周期 T_p, max を 12msec と仮定して，$L=120$（サンプリング周期 $T=0.1$msec）とする。次に分析始点より分析区間として，音声波形 $S(n)$, $(n=0, 1, 2, \cdots\cdots, N-1)$ を抽出する。なお，$N=256$ (25.6msec) である。

更に，窓関数として Fig. 3 に示す半 Hamming 窓 $W(n)$ を提案し，$S(n)$ に乗じる。式(13)に窓関数を定義する。

$$X(n)=S(n)\left\{0.54+0.46\cos\frac{k\pi}{N-1}\right\} \qquad (13)$$

ここで，上記窓処理によるスペクトルを考察する。Fig. 2 (b)は(a)で示したスペクトルを有する合成音信号に提案した窓処理を行った場合のスペクトルである。この場合，$P_w(w)$ におけるレベル変化は，Fig. 2 (a)と比べて比較的一定しており，急激なレベル変化は生じていない。したがって，対数スペクトル $X(w)$ にも急激なレベル変化は現れず，声道特性を表わすスペクトル $V(w)$ の良い近似を与えている。このように従来の窓処理に比べて，スペクトル微細構造の影響を事前に軽減することができ，更に，窓処理された信号を一連の準同形処理に通すことによって近似の良い声道インパルス応答数列 $V(n)$ が推定可能となる。

3．2　極—零点の推定

上記のようにして推定された声道インパルス応答数列 $V(n)$ より，音声生成系を ARMA モデルで近似し，式(14)に示す声道伝達関数の推定を行う。

$$V(z)=\frac{1+b_1z^{-1}+b_2z^{-2}+\cdots\cdots b_qz^{-q}}{1+a_1z^{-1}+a_2z^{-2}+\cdots\cdots a_mz^{-m}}=\frac{1}{G(z)} \qquad (14)$$

但し，$V(z)$ は声道伝達関数の z 変換表示を示し，$G(z)$ は $V(z)$ の逆フィルタを意味する。本論文では，式(14)の推定に，$G(z)$ を用いた逆フィルタ法による展開を試みる。いま，単一インパルスを $U(n)$ とすると，その z 変換表示 $U(z)$ を用いて式(15)が成り立つ。

又，その時系列変換は式(16)のようになる。

$$V(z)=\frac{1+b_1z^{-1}+b_2z^{-2}+\cdots\cdots b_qz^{-q}}{1+a_1z^{-1}+a_2z^{-2}+\cdots\cdots a_mz^{-m}}U(z) \qquad (15)$$

$$V(n)=-\sum_{i=1}^{m}a_iV(n-i)+\sum_{j=0}^{q}b_jU(n-j) \qquad (16)$$

ここで，$G(z)$ のインパルス応答数列を $g(i)$, $(i=0, 1, \cdots\cdots, p)$ とすると，式(14)，(15)の関係から，式(17)が成り立つ。

$$U(n)=V(n)+g(1)V(n-1)+g(2)V(n-2)+\cdots\cdots+g(p)V(n-p)+E(n) \qquad (17)$$

$E(n)$ は残差を意味する。まず，式(17)より逆インパルス応答 $g(i)$ の推定を行う。式(17)を $n=1, 2, \cdots\cdots, N$ のサンプル値において行列式で表わすと式(18)のようになる。

$$\boldsymbol{U}=\boldsymbol{V}_0+\boldsymbol{V}\boldsymbol{G}+\boldsymbol{E} \qquad (18)$$

但し，$\boldsymbol{V}_0=(V(1), V(2), \cdots\cdots V(N))^T$

$\boldsymbol{U}=(U(1), U(2), \cdots\cdots U(N))^T$

$\boldsymbol{G}=(g(1), g(2), \cdots\cdots g(N))^T$

$\boldsymbol{E}=(E(1), E(2), \cdots\cdots E(N))^T$

$$\boldsymbol{V}=\begin{bmatrix} V(0) & V(-1) & \cdots\cdots V(-p+1) \\ V(1) & V(0) & \cdots\cdots V(-p+2) \\ \vdots & \vdots & \\ V(N-1) & V(N-2) & \cdots\cdots V(N-p) \end{bmatrix}$$

更に，残差 $E(n)$ を最小にするよう，最小二乗法によって，式(18)を展開することにより，式(19)，(20)を導くことができる。

$$\boldsymbol{V}^T(\boldsymbol{U}-\boldsymbol{V}_0)=\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{G} \qquad (19)$$

$$-\boldsymbol{V}^T\boldsymbol{V}_0=\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{G} \qquad (20)$$

したがって，式(20)を式(21)のように整理することによって，$g(i)$ は推定される。

$$\begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(p) \end{bmatrix}=\begin{bmatrix} R(0) & R(1) & \cdots\cdots R(p-1) \\ R(1) & R(0) & \cdots\cdots R(p-2) \\ \vdots & \vdots & \\ R(p-1) & R(p-2) & \cdots\cdots R(0) \end{bmatrix}\begin{bmatrix} g(1) \\ g(2) \\ \vdots \\ g(p) \end{bmatrix} \qquad (21)$$

但し，$R(i)=\sum_{n=1}^{N-1}V(n)V(n-i)$

$(i=0, 1, 2, \cdots\cdots, p)$

また，ai, bj, $g(i)$ との間には次式が成り立つ。

$$\sum_{j=0}^{i}bjg(i-j)=\begin{cases} ai & (0\leqq i\leqq m) \\ 0 & (i>m) \end{cases} \qquad (22)$$

但し，$a_0=b_0=g(0)=1$

$bj=0 \quad (j<q)$

この式(22)から，ai, bj を推定できるわけだが，ここでは，下記の手順に従い，2回の反復推定によって求め

ている。まず，ai の初期推定を系を全極モデルと仮定して $V(n)$ より線形予測を用いて推定する。また，式(22)より式(23)が与えられる。

$$\boldsymbol{H}=\boldsymbol{DB}+\boldsymbol{E} \tag{23}$$

但し，$\boldsymbol{B}=(b_1, b_2, \cdots\cdots, b_p)^T$

$\boldsymbol{H}=(a_1-g(1), a_2-g(2), \cdots\cdots, a_m-g(m), \cdots\cdots, g(p))^T$

$\boldsymbol{E}=(e_1, e_2, \cdots\cdots, p)^T$

$$\boldsymbol{D}=\begin{bmatrix} 1 & 0 & \cdots\cdots & 0 \\ g(1) & 1 & \cdots\cdots & 0 \\ \vdots & \vdots & & \\ g(p-1) & g(p-2) & \cdots\cdots & g(p-n) \end{bmatrix}$$

ここで，式(23)より bj の最小二乗推定は式(24)で与えられる。

$$\boldsymbol{D}^T\boldsymbol{DB}=\boldsymbol{D}^T\boldsymbol{H} \tag{24}$$

次に，ai, bj の2次推定を行うため，式(25)に示される新たな信号 $V'(n)$ を考える。

$$V'(n)=V(n)-\sum_{j=1}^{q} bjV'(n-j) \tag{25}$$

$$V'(1)=V(1) \qquad (n=1, 2, \cdots\cdots, N)$$

式(25)は $V(n)$ より bj の影響を意味し，この式より $V'(n)$ は全極形の系による応答数列とみなすことができる。したがって，$V'(n)$ より，1次推定と同様に，まず ai の2次推定を行い，更に，bj の2次推定を行う。このようにして，ai, bj が推定されると，式(14)の分子，分母を各々零とおき，高次方程式の根を求めることによって，極—零点を推定することができる。共振，反共振の周波数，及び帯域幅は方程式の根を zi とすることによって，式(26)で与えられる。

$$\begin{cases} Fi=arg(zi)/2\pi\Delta T \\ BWi=\log|zi|/2\pi\Delta T \end{cases} \tag{26}$$

3．3　合成音を用いた分析精度の評価

ここで，本改善手法の有効性を検討するため，合成音によるシミュレーション実験を行う。合成音信号は極点6個，零点1個を有するモデルを想定して，実際には，音源として周期 T_p のインパルス列を用い，$m=14$, $q=3$ を有する ai, bj を与えることによって作成している。なお，標本化周期 $T=0.1$msec，分析区間長 $N=25.6$msec (256サンプル) とする。また，逆インパルス応答数列 $g(n)$, $(n=1, 2, \cdots\cdots, p)$ の長さは，計算量及び精度を考慮して，$p=80$ とする。更に，式(7)に示すリフタの長さ M は，ピッチ周期の長さを考慮して，式(27)のように決定する。

$$\begin{cases} M=50 & T_p \geqq 5.0(\text{msec}) \\ M=10\,T_p-5 & T_p<5.0(\text{msec}) \end{cases} \tag{27}$$

① ピッチ周期による影響

まず，ピッチ周期によって分析精度への影響度について検討する。Fig. 4 (a), (b)にピッチ周期を順次変化させた場合のフォルマント，アンチフォルマント，及びその帯域幅の推定誤差評価を示す。評価式は式(28)を用いる。ここで，$F(i)$ は真のフォルマント，及びアンチフォルマント周波数を，$Fe(i)$ は各々の推定値を表わす。帯域幅についても同様である。

$$\varepsilon=\sqrt{\frac{1}{7}\sum_{i=1}^{7}(Fe(i)-F(i))^2} \tag{28}$$

Fig. 4 において比較すると，フォルマント，及び帯域幅共に，従来の手法ではピッチ周期によって誤差変動の著しいことに気づく。この原因はピッチ周期によってスペクトル微細構造のレベル変化が異なり，特に，急激なレベル変化が生じる位置，度合いによって，スペ

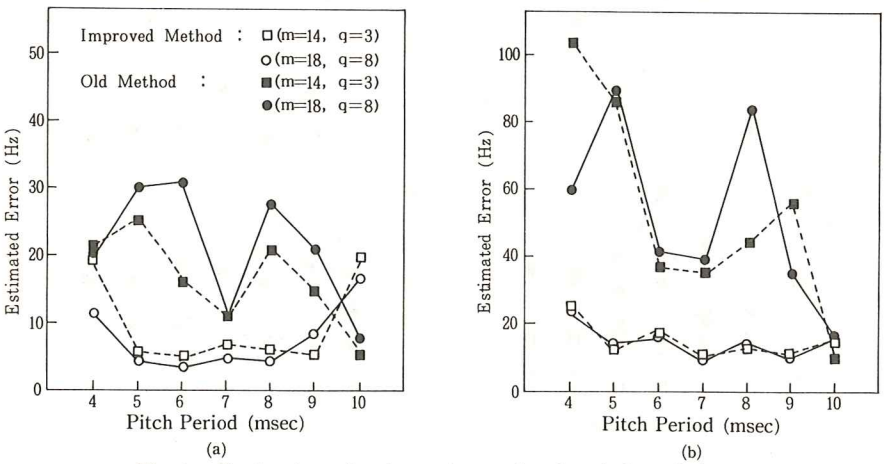


Fig. 4　Evaluation of estimated error by the pitch periods.
(a) Formants and antiformants frequency
(b) Band width

クトル包絡の追従のしかたが異なってくることに起因するものと考えられる。一方，本手法ではスペクトル微細構造の急激なレベル変化を比較的抑制しているため，その追従の度合いを無視することができ，従来の手法に比べてかなり誤差が少なくなっており，又誤差変動も減少している。特に Fig. 4 (b)に見るように，帯域幅に関しては安定した精度を保っている。フォルマントの誤差評価（Fig. 4 (a)）において，ピッチ周期が4,10msec の場合，誤差が他に比べて大きくなっている。この原因は，零点の位置とスペクトル微細構造の谷部のレベル変化が一致したため，零点の推定誤差が劣化したため生じたもので，フォルマント周波数のみの推定精度は他のピッチ周期の場合と同様に改善されている。このように，本改善手法による分析では，ピッチ周期に比較的影響されない高精度の分析が期待できる。

② 極—零次数による影響

極—零次数による推定精度への影響度について検討を行う。Fig. 5 (a), (b)に Fig. 4 と同様にフォルマント，及びアンチフォルマントの推定誤差評価を示す。(a)は極次数を変化させた場合の評価である。従来の手法では零次数が真値 $q = 3$ のとき，極次数 m が増加するにつれて推定誤差が大きくなっている。又，過大に仮定した $q = 8$ のときでは逆に減少している。この原因は，$q = 3$ のときでは余分な極でスペクトルの谷部を近似しようとするため，推定値に影響を与え誤差が大きくなり，逆に $q = 8$ の場合では余分な極零を互いに打消しあって誤差が減少するものと考えられる。これに対して，本手法では比較的次数による影響が少なく，安定した精度を保っている。Fig. 5 (b)は零次数を変化させたときの誤差評価である。本図においても，従来の手法に比べて，精度が改善されていることがわかる。特に，従来の手法では，$m = 14$のとき真値 $q = 3$ 以外の場合，余分な零による誤差の増大が著しい。又，$m = 18$のときは，零次数を過大にしても余分な極零を互いに打ち消しあって，$m = 14$のときに比べて誤差が減少している。しかしながら，本手法に比べると誤差は大きい。この原因は Fig. 6 に示すスペクトル包絡によく特徴が現れている。Fig. 6 (b)は $m = 18$, $q = 8$ のときの従来の手法によるスペクトル包絡の推定値である。この場合，対数スペクトルには，2.4kHz, 3.9kHz 付近に急激な微細構造のレベル変化が現れており，スペクトル包絡はこのレベル変化に追従して，偽零点を形成している。すなわち，Fig. 6 (a)に示す真のスペクトル包絡に比べてスペクトルが変形している。このため，偽零点の両側のフォルマントは大きく影響を受けることになり，誤差が増大するものと考えられる。又，この偽零点は，零点を推定する際，真の零点との判別を困難にする恐れがある。この合成音（ピッチ周期9msec）においても，真の零点の帯域幅が123Hz と推定しているのに対して，偽零点の推定値は251Hz, 159Hz と比較的狭い帯域幅となっている。

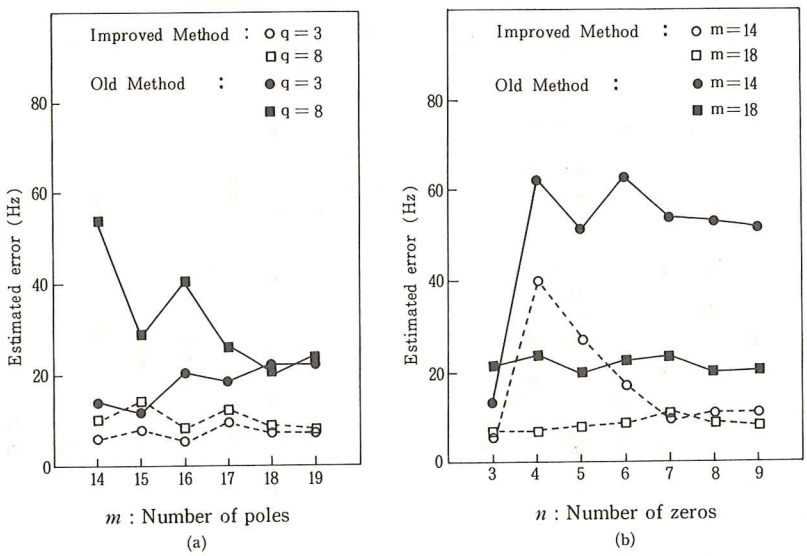


Fig. 5　Evaluation of estimated error by the number of poles-zeros.

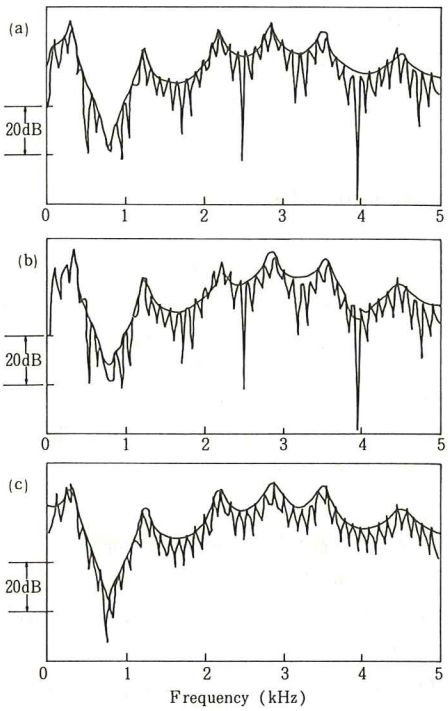


Fig. 6 Log spectrums and spectral envelopes.
(a) Real spectral envelope (14 poles/3 zeros)
(b) Old method ($m=18$, $q=8$)
(c) Proposed method ($m=18$, $q=8$)

以上のように，従来の手法では余分な極—零によって誤差が比較的大きくなり，かつ偽零点を抽出する恐れがあるといった問題が生じるのに対して，本手法ではこれらの問題を解決することができ，高精度の分析が可能と考えられる。更には，極・零次数が未知である音声の場合，比較的次数を大きく仮定しておけば，品質の良い分析が期待できる。

3．4 自然音声への適用

前節までに述べた分析アルゴリズム，及び合成音による有効性を確認するため，自然音声（鼻子音）分析への適用を試みた。分析用音声は，5kHz の低域フィルタに通した後，10kHz サンプリング，12bit 量子化を行っている。更に，低域の零点と高域の極を強調するため，プリエンファシス処理を施している。

Fig. 7 に鼻子音を含む/VCV/型音節の単語/ima/，/ana/の/m/，/n/についての分析結果を示す。なお，極—零次数は比較的過大に仮定し，$m=16$，$q=12$とした。他の条件は合成音の場合と同様である。Fig. 7 より，ピッチ周期に影響されることなく，鼻子音の特徴である/m/における1kHz，3kHz 付近の零点，及び，/n/における2kHz付近の零点を明確に推定していることがわかる。又，過大な次数の仮定にもかかわらず，

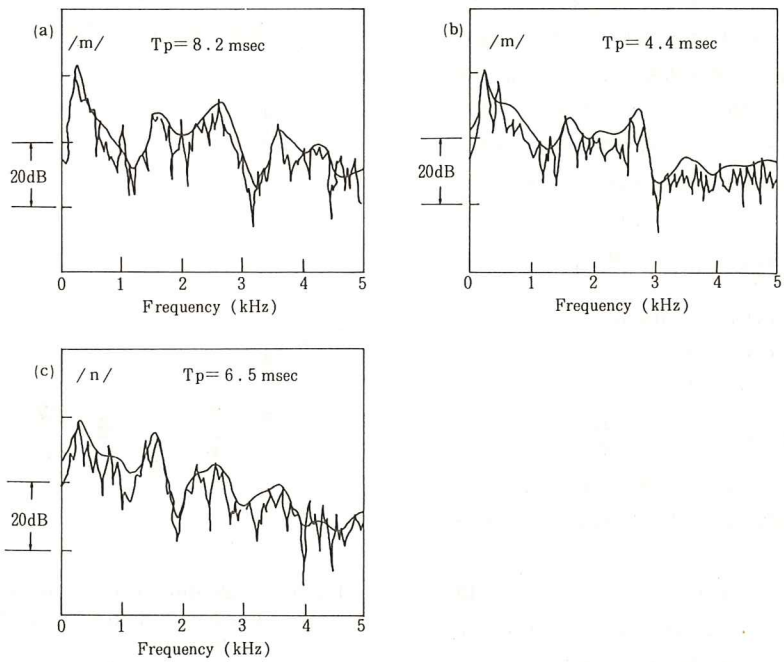


Fig. 7 Log spectrums and spectral envelopes of natural speech. (/m/, /n/)

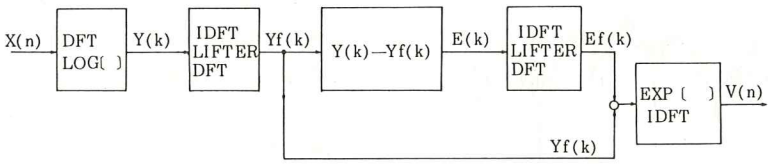


Fig. 8　Block diagram of improved method II.

スペクトル微細構造による深い谷部の形成もなく，安定した極—零点の抽出がなされている。しかしながら，環境雑音によるスペクトルの乱れが対数スペクトルに現れており，より高精度な推定を期待するならば，この雑音情報の適切な除去が必要であろう。

4　改善手法II

4．1　分析アルゴリズム

ここでは声道インパルス応答を推定する際，微細構造のレベル変化に平滑化スペクトルが追従しないよう，以下の手順に従って，ホモモルフィックフィルタを修正する。従来の手法では，式(3)〜(8)の手順に従って $Y_f(k)$ を推定している。しかしながら，本手法では，式(4)に示す $Y(k)$ のピーク包絡値に $Y_f(k)$ が出来るだけ近似するよう $Y_f(k)$ を修正している。すなわち，$Y_f(k)$ は $Y(k)$ に比べて比較的レベルが低いと仮定して，まず式(29)のように $E(k)$ を設定する。次に $Y(k)$ から $Y_f(k)$ から $Y_f(k)$ を推定したときと同様に，$E(k)$ の平滑化 $E_f(k)$ を求める。スペクトル的に言うならば，$E_f(k)$ は対数スペクトルと平滑化スペクトルの誤差の平滑化スペクトルと考えることができる。

$$E(k)=\begin{cases} Y(k)-Y_f(k) & Y(k)\geqq Y_f(k) \\ 0 & Y(k)<Y_f(k) \end{cases} \tag{29}$$

$$C_e(n)=\frac{1}{N}\sum_{k=0}^{N-1}E(k)\exp\left(j\frac{2\pi kn}{N}\right) \tag{30}$$

$$E_c(n)=\begin{cases} 2C_e(n) & 0<n\leqq M-1 \\ C_e(n) & n=0 \\ 0 & M\leqq n\leqq N-1 \end{cases} \tag{31}$$

$$E_f(k)=\sum_{n=0}^{N-1}E_c(n)\exp\left(-j\frac{2\pi kn}{N}\right) \tag{32}$$

上記のようにして $E_f(k)$ が求められると，$Y_f(k)$ は $Y(k)$ のピーク値に近づくように式(33)のように修正することができる。

$$Y_e(k)=Y_f(k)+E_f(k) \tag{33}$$

したがって，声道インパルス応答 $V(n)$ は，式(34)，(35)に示すように，指数変換，続いて離散的逆フーリェ変換を行うことによって推定することができる。

$$V_f(k)=\exp[Y_e(k)] \tag{34}$$

$$V(n)=\frac{1}{N}\sum_{k=0}^{N-1}V_f(k)\exp\left(j\frac{2\pi kn}{N}\right) \tag{35}$$

声道インパルス応答が推定されると，3章で述べた手順に従い，極—零点の推定を行う。また，Fig. 8 に一連の手順を示す。

4．2　合成音を用いた分析精度の評価

①　修正反復回数

本手法では，式(29)〜(33)の手順に従い，$Y_f(k)$ が出来るだけ $Y(k)$ のピーク値に近似するよう修正を施し，$Y_e(k)$ を設定した。しかしながら，実際には $E(k)$ の平均値で修正を行っているため，1回の修正では充分な修正近似がなされているとは言難い。そこで修正を数回反復することで，より頂点への近似度を深めることを試みる。最適反復回数の決定を行うため，ピッチ周期 $T_p=9.0$msec の合成音を用い，反復回数による推定誤差への影響度を調べる。評価量は，式(28)にしたがう。

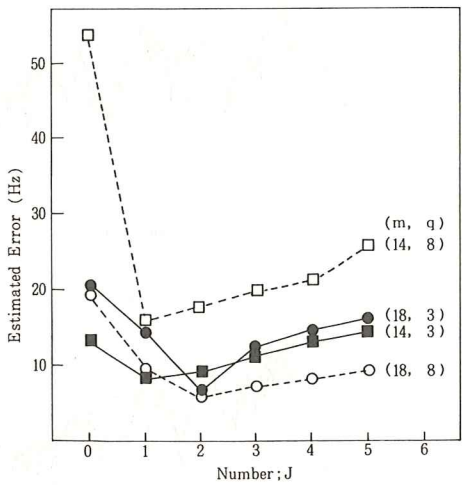


Fig. 9　Evaluation of estimated error by the iterative number of times for modification.

Fig. 9 に評価結果を示す。極－零次数を$m=14$, 18, $q=3$, 8と仮定して，その組合せによる評価を行ったが，全体的に$J=1$又は2の時に誤差が最低となっている。逆に$J \geqq 3$では誤差が増大している。この原因は反復回数を増すことで過度の修正を行うことになり，スペクトルが飽和して誤差が拡大するものと考えられる。この現象は他の極—零次数においても同様な傾向を示しており，以上の結果から反復回数は$J=2$が最適と考えられる。

② 合成音による分析例

本改差手法においても，極—零次数を過大にしたとき，分析精度が十分高精度に役まれていることが重要な問題となる。そこで，Fig. 5 と同様に極—零次数による推定精度への影響度について検討を行う。Fig. 10 に誤差評価を示す。Fig. 10 (a)は極次数を変化させた場合の評価である。零次数は$q=3$に固定している。予想通り，従来の手法では，余分な極によって誤差が増加しているのに対して，本手法では次数による影響が小

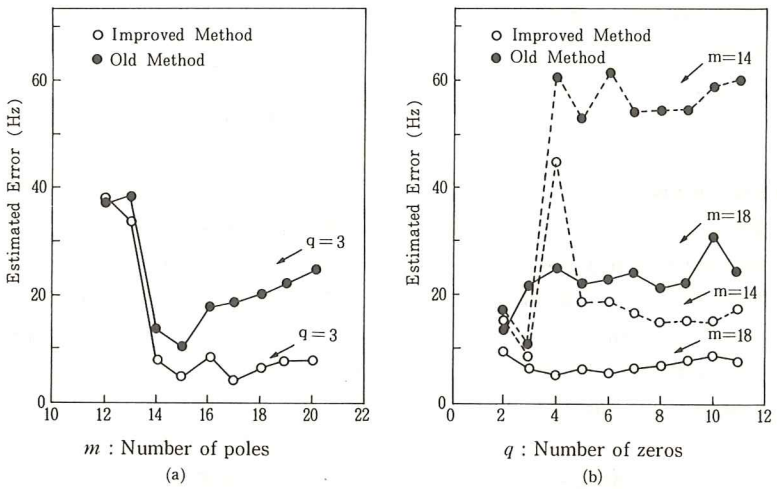


Fig. 10 Evaluation of estimated error by the number of poles-zeros.

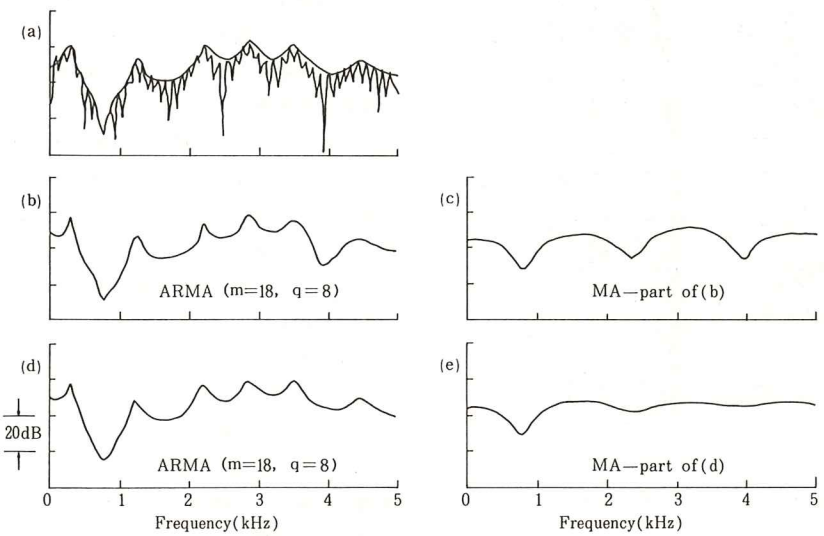


Fig. 11 Estimated spectral envelopes.
(a) Real log spectrum and spectral envelope
(b), (c) Spectral envelope by old method
(d), (e) Spectral envelope by improved method

さく，高精度に保たれていることがわかる。又，Fig. 10 (b)は零次数を変化させた場合の評価であるが，同様な傾向にあり，本手法では安定した精度を保っており，分析精度が改善されていることがわかる。以上の結果は，Fig. 11 に示すスペクトルの例にも良く現れており，従来の手法では余分な極・零によって，スペクトル微細構造のレベル変化に追従した深い谷部の形成が見られるのに対して，本手法では近似良いスペクトル包絡の指定を行っている。

上記のように本手法では，改善手法Ⅰと同様に極—零次数による影響を比較的受けず推定することが可能となり，極—零次数が未知な場合においても，次数を比較的大きく設定しておくことで品質の良い音声分析が期待できる。

4．3　自然音声への適用

本節では自然音声への有効性を確認するため，鼻子音を含む/VCV/型音節の単語/ima/を用いて連続分析への適用を試みる。分析フレーム周期は6.4msec，極—零次数は過大に仮定して$m=16$，$q=12$とする。他の分析条件は，3章と同様である。Fig. 12 に連続分析によるスペクトル包絡値を示す。Fig. 12 (b)は従来の手法による結果であるが，フレーム軸にそったスペクトルの乱れを生じており，スペクトル微細構造の影響を受けていると思われるフレームにおいては，不規則に深い谷部を生じている。これに対して，本手法では比較的安定した連続スペクトル値を示しており，鼻子音部，母音部の特徴を明確に表わしている。又，Fig. 12 (a)における母音部においても比較的深い谷部を生じている部分がある。これは鼻音化，更には声帯音源による零点と考えられる。

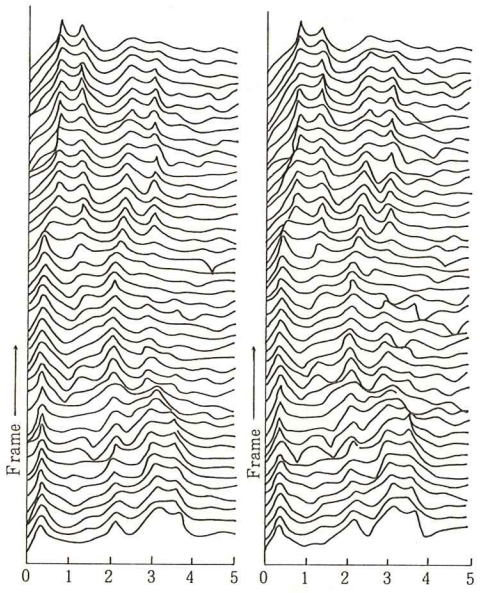


Fig. 12　Spectral envelope series for real speech signals /ima/.
(a) Improved method
(b) Old method

以上の結果，本手法は自然音声においてもその効果を十分発揮することができ，連続分析にも有効なことが明らかとなった。

5　むすび

本論文ではホモモルフィック予測による高精度極—零形音声分析を目的として，2つの改善手法を提案しその有効性について検討を行った。すなわち，従来の手法の欠点である，音源周期構造成分の影響による極—零点の精度劣化を防ぐため，(Ⅰ)窓処理に関する改良，並びに(Ⅱ)ホモモルフィックフィルタの改良についての改善手法を提案した。その結果，従来の手法に比べて両者共に分析精度の向上を図ることができ，スペクトル微細構造のレベル変化による影響を軽減した極—零点の推定が可能となった。下記に，合成音及び自然音声を用いて考察した，各改善手法の特徴を示すことによって両者の比較を行う。

(1)　改善手法Ⅰはスペクトル微細構造のレベル変化を事前に窓処理によって軽減しているため，従来の手法に比べてかなり分析精度を改善することができる。特に，ピッチ周期，及び極—零次数にかかわらず，比較的安定した精度良い分析が可能であり，計算量も従来の手法と同量であることから有効な分析手法と考えられる。しかしながら，連続分析を行う場合，本手法は分析始点をピッチに同期させる必要があり，分析始点の決定方法に多少の不便さを残している。

(2)　改差手法Ⅱはホモモルフィックフィルタを修正することで分析精度の向上を図った手法であるが，ピッチ周波数が極めて高くない場合は，極—零次数の影響を軽減した高精度分析が期待できる。更には，スペクトル微細構造のレベル変化に対する影響度が少ないため，連続分析にもその効果を十分発揮することができる。しかしながら，ピッチ周波数が極めて高い場合，帯域幅の推定を過大評価することがあり，多少の

不安定さを残している。

上記のように，両者の改善手法は高精度分析に十分有効な手法と考えられる。しかしながら，より高精度な音声分析を期待するならば，目的とする分析処理によって適宜選択していくことが必要であろう。

参考文献

1) Oppenheim, A. V., Kopec, G. E. and Tribolet, J. M.: "Signal analysis by homomorphic Prediction", IEEEE Trans. Acoust., Speech & Signal Process., ASSP-24, pp. 327－332 (1976)

2) Oppenheim, A. V., Kopec, G. E. and Tribolet, J. M.: "Speech analysis by homomorphic prediction", IEEE Trans. Acoust., Speech & Signal Process., ASSP-25, pp 40－49 (1977)

3) 横山，井上；"ホモモルフィック予測による音声分析の精度改善"，信学論(A)，J 67－A，№9，pp. 928－929 (1984)

4) 横山，井上；"改良ホモモルフィック予測法による音声の極・零点推定"，信学論(A)，J 63－A，№5，pp. 454－461 (1982)

5) 深林，鈴木："極・零形線形モデルによる音声分析"，信学論(A)，J 58－A，№5，pp. 270－277(1975)

6) J. D. Markel; "Application of a Digital Inverse Filter for Automatic Formant and Fo Analysis", IEEE Trans. on Audio and Electroacoustics, Vol. AU-21, No. 3, pp. 149－153 (1973)