

An Effect of the Restriction of Predictive Coefficients on Linear Prediction Method

メタデータ	言語: jpn 出版者: 公開日: 2017-10-03 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	http://hdl.handle.net/2297/9616

線形予測法における予測係数拘束の効果

三好義昭

An Effect of the Restriction of Predictive Coefficients on Linear Prediction Method

Yoshiaki MIYOSHI

1. まえがき

情報社会の到来といわれて久しいが、今日の高度情報化社会の基盤をなしているのはデジタル信号処理技術のハード・ソフト両面での飛躍的な進歩・普及と言える。デジタル信号処理は、周知のように高精度処理が可能、品質の劣化がなく特性が均一かつ安定、時分割処理が可能などの特徴を有することから、われわれの身近で主要な情報源である音や映像といった本来はアナログ(連続量)の情報もデジタル(離散量)に変換して処理する時代となり、現代はまさにデジタル信号処理の時代と言える。このデジタル信号処理の中でも、観測信号の周波数スペクトルに関する情報が10個程度の係数に集約される線形予測法[1]が特に音声信号処理に広く活用されている[2][3]。

本論文では、通常の線形予測法における予測係数とこの予測係数を基に得られる観測系の推定極の周波数及び帯域幅との関係の考察結果に基づき、予測係数の自由度を拘束した拘束線形予測法を提案し、本方法による観測信号の極周波数推定の有効性を合成音ならびに自然有声破裂音の特徴抽出に適用して実験的に示す。

以下、2.において、線形予測法の予測係数と推定極の関係を明らかにし、3.において、予測係数拘束の効果を示す。そして、4.において、合成音のシミュレーションにより音声の極周波数であるホルマント周波数の推定精度を通常の線形予測法と比較して示し、5.では、実際に自然有声破裂音のホルマント周波数推定

に適用して、本方法の有効性を示す。

2. 線形予測法によるデジタル信号処理

線形予測法とは、任意の時点の観測値をそれ以前の観測データの線形一次式で予測する手法である。すなわち、観測信号値列 $(y_1, y_2, \dots, y_n, \dots)$ の第 n 番目の観測値 y_n (第 n 標本値)の予測値 \hat{y}_n をその時点から過去 p 個の観測値 $(y_{n-1}, y_{n-2}, \dots, y_{n-p})$ の線形一次式、

$$\hat{y}_n = -(\alpha_1 y_{n-1} + \alpha_2 y_{n-2} + \dots + \alpha_p y_{n-p}) \quad (1)$$

で予測できるものとし(ここで負の符号を付けるのは後の式を簡潔にするためである)、観測値 y_n とその予測値 \hat{y}_n の誤差 ε_n ,

$$\begin{aligned} \varepsilon_n &= y_n - \hat{y}_n = y_n - (-\sum_{k=1}^p \alpha_k y_{n-k}) \\ &= y_n + \sum_{k=1}^p \alpha_k y_{n-k} \end{aligned} \quad (2)$$

の自乗平均 $\bar{\varepsilon}_n^2$,

$$\bar{\varepsilon}_n^2 = \frac{1}{N} \sum_{n=1}^N (y_n + \sum_{k=1}^p \alpha_k y_{n-k})^2 \quad (3)$$

最小基準により得られる、

$$\sum_{k=1}^p \phi_{ik} \alpha_k = -\phi_{0i} \quad (4)$$

但し、 $\phi_{ik} = E\{y_{n-i} y_{n-k}\}$, $i=1, 2, \dots, p$

なる p 元連立一次方程式（この式を正規方程式と称する）の解として係数 $\{\alpha_i\} \ i=1,2,\dots,p$ を算出する手法である。ここで、係数 $\{\alpha_i\} \ i=1,2,\dots,p$ を予測係数、 p を予測次数と称する。ところで、式(2)より、

$$y_n = -\sum_{k=1}^p \alpha_k y_{n-k} + \varepsilon_n \quad (5)$$

となる。すなわち、線形予測法は観測系を予測誤差 ε_n を入力とする全極型モデルで記述したのと等価であり、この観測系の伝達関数 $H(z)$ は、式(5)の両辺の z 変換より、

$$H(z) = \frac{Y(z)}{\varepsilon(z)} = \frac{1}{1 + \sum_{k=1}^p \alpha_k z^{-k}} \quad (6)$$

となる。

線形予測法における予測次数 p は一般には 10 前後が用いられており、比較的少数の予測係数で、観測信号の特性を精度良く表すことができ、かつ必要に応じて、この予測係数を係数とする p 次方程式(式(6)の分母=0)の根より、観測系の極情報も推定できることから、今日、線形予測法がデジタル信号処理に広く活用されている。

ここで、予測次数 p が偶数の場合、式(6)は、

$$H(z) = \frac{1}{\prod_{i=1}^{p/2} (1 + a_i z^{-1} + b_i z^{-2})} \quad (7)$$

$$\text{但し、} \ a_i = -2e^{-\pi B_i T} \cos(2\pi F_i T)$$

$$b_i = e^{-2\pi B_i T}$$

なる共役複素極の積の形で記述できる。ただし、 F_i : 第 i 極の周波数、 B_i : 第 i 極の帯域幅、 T : 標本化周期である。したがって、式(6)と式(7)の分母が恒等的に等しいことより、

$$\alpha_1 = \sum_{i=1}^{p/2} a_i = -2 \sum_{i=1}^{p/2} e^{-\pi B_i T} \cos(2\pi F_i T) \quad (8)$$

$$\alpha_p = \prod_{i=1}^{p/2} b_i = \prod_{i=1}^{p/2} e^{-2\pi B_i T} \quad (9)$$

となる。すなわち、通常の線形予測法により得られる第 1 予測係数 α_1 は推定極の周波数と帯域幅の比較的単純な関数となっており、特に第 p 予測係数 α_p は推定極の帯域幅のみの関数となる。

3. 予測係数拘束の効果

今、式(7)の分母=0 とおいた第 i 項 $1 + a_i z^{-1} + b_i z^{-2} = 0$ の根を z_i とすると $|z_i| = \sqrt{b_i}$ より、

$$b_i = 1 \quad (10)$$

但し、 $i=1,2,3,\dots,p/2$

すなわち、式(9)より $\alpha_p = 1$ とすれば、 $H(z)$ の極はすべて z 平面の単位円上に存在することになる。この時、式(7)の分母は、

$$\prod_{i=1}^{p/2} (1 + a_i z^{-1} + z^{-2}) = 1 + c_1 z^{-1} + c_2 z^{-2} + \dots + c_2 z^{-p+2} + c_1 z^{-p+1} + z^{-p} \quad (11)$$

$$\text{但し、} \ c_i = f(a_j), \ i, j=1,2,3,\dots,p/2$$

と、 z_i の係数 c_i と z^{-p+i} の係数 c_{p-i} が等しくなる。したがって、式(11)より、式(6)において、

$$\alpha_i = \alpha_{p-i} \quad (12)$$

$$\text{但し、} \ \alpha_0 = \alpha_p = 1, \ i=0,1,2,\dots,p/2$$

すなわち、 $\{\alpha_i\}, i=0,1,2,\dots,p$ を $p/2$ 番目の予測係数 $\alpha_{p/2}$ を中心とした対称形に拘束することにより、 $H(z)$ の極を全て z 平面の単位円上に拘束したことになる。このとき、

$$1 + \sum_{k=1}^p \alpha_k z^{-k} = 0 \quad (13)$$

は式(12)より相反方程式となり、式(13)を解くことは $x = z + z^{-1}$ に関する $p/2$ 次方程式を解くことと、 z に関する 2 次方程式 $z^2 - xz + 1 = 0$ を解くことに帰着する[4]。そして、予測係数 α_i

も式(12)のもとで、予測誤差の自乗平均最小の条件より、 $p/2$ 個の予測係数 $\{\alpha_1, \alpha_2, \dots, \alpha_{p/2}\}$ のみを求めればよいことになる。

以上のように、予測係数 α_i を式(12)に拘束することにより、推定極の位置が z 平面の単位円上に拘束されるため、推定極の帯域幅の情報は得られないが、正規方程式の次元ならびに高次方程式の次数をそれぞれ半減することができる。この数値演算上のメリットもさることながら、本方法では全ての極の位置が z 平面の単位円上に拘束されていることから、定性的には推定極の位置が予測誤差に大きく影響するため、予測誤差の自乗平均最小の条件より、観測信号の極情報を担った安定な極推定が可能と言える。このことを解析的に明らかにするのは困難であるため、以下、線形予測法が広く活用されている分野の一つである音声信号処理に適用し、その有効性を検証する。

4. 合成音によるシミュレーション結果

音声の重要な特徴パラメータであるホルマント周波数(声道伝達関数の極周波数)推定における予測係数拘束の効果を、声道(声帯から唇までの音響的空間)の断面積がほぼ一定で、ホルマントに極端な偏りのない母音/e/により検証する。合成条件は、標準化周波数 10kHz、励振源:ピッチ周期 8ms の Rosenberg 波[5]、ホルマント周波数: $F_1 = 437.5\text{Hz}$, $F_2 = 1812.5\text{Hz}$, $F_3 = 2687.5\text{Hz}$, $F_4 = 3437.5\text{Hz}$, $F_5 = 4437.5\text{Hz}$, 放射特性: 6dB/oct である。

4. 1 ホルマント周波数推定精度

本手法により得られる周波数スペクトルを通常の線形予測法と対比して図1に示す。ただし、前処理として一階差分後、分析窓長 $T_u = 25.6\text{ms}$ 、図1上段は分析次数 $p=10$ 、下段は $p=12$ とした場合の結果で、図中の破線は合成音の声道特性である。なお、本方法による推定極は分析次数の如何に拘わらず全て z 平面の単位円上に

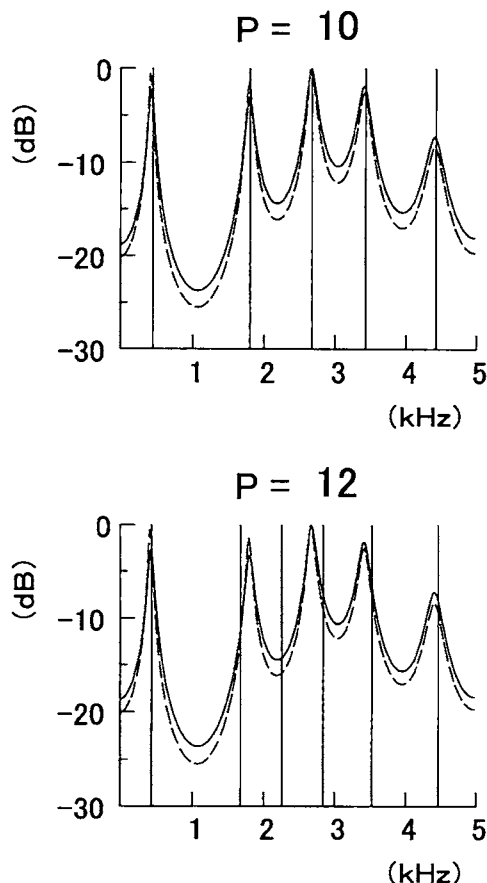


図1 周波数スペクトルの比較(合成母音/e/)

存在するため推定極の帯域幅は零、すなわち線スペクトルとなる[6][7]。

今の場合、前述の5個のホルマントを用いて声道特性を設定しているので、声道伝達特性の次数 $p_0=10$ となる。したがって、分析次数 $p=10(=p_0)$ の場合(図1上段)、本手法ならびに通常の線形予測法ともホルマント周波数を精度良く推定できており、特に通常の線形予測法は声道特性もほぼ正確に推定できていると言える。

一方、 $p=12(>p_0)$ とした場合、本手法は z 平面の単位円上に存在する6個($=p/2$)の共役複素極として推定するため、図1下段に示されているように、2.2kHz付近にホルマントに対応しない極が推定され、この極の影響で特に第2ホ

ホルマントに対応する線スペクトルが低域に、第 3 及び第 4 ホルマントに対応する線スペクトルが高域にそれぞれシフトしていると言える。これに対して、通常の線形予測法では $p=12(>p_0)$ とした場合でも、 $p=10(=p_0)$ とほぼ同等の周波数スペクトルが推定されており、本手法のような推定極のシフトは起こらず各ホルマント周波数を精度良く推定できていると言える。ただし、通常の線形予測法においても $p=12(>p_0)$ とした場合、一般には 6 個($=p/2$)の共役複素極が推定されており、本合成音のホルマントは 5 個であることから、ホルマントに対応しない極(擬似ホルマントと称する)が 1 個存在するが、この擬似ホルマントの帯域幅が大きくなることにより周波数スペクトルへの影響が軽減されているからである。したがって、 $p>p_0$ とした場合、通常の線形予測法では、ホルマントに対応する極の選定問題が生じる。この問題に関しては後述する。

図 1 より、本手法のホルマント周波数推定精度は分析次数 p に大きく依存すると言えるので、ホルマントの中でも音声認識等において特に重要となる第 1～第 3 ホルマント周波数推定誤差 E の分析次数 p 依存性を図 2 に示す。ただし、前処理等の分析条件は図 1 の場合と同じで、フレームシフト間隔 0.2ms で 1 周期に渡って分析した計 40 フレームの平均値で、○印: 本手法、△印: 通常の線形予測法の結果である。

図 2 より、分析次数 $p=8(<p_0)$ では、ホルマント周波数推定誤差は両方法とも大きくなるが、 $p=10(=p_0)$ の場合、通常の線形予測法による誤差は 0.9% であるのに対して、本手法による誤差は 1.0% と若干劣るが、本手法でも通常の線形予測法とほぼ同等の精度でホルマント周波数が精度良く推定可能であると言える。しかしながら、通常の線形予測法は $p \geq 14$ においてホルマント周波数推定誤差が若干増大してはいるが、 $p \geq p_0$ であれば、ホルマント周波数推定誤差に及ぼす分析次数依存性はほとんどないのに対して、本手法は $p > 10$ において、ホルマント周

波数推定誤差が分析次数に依存して変動し、通常の線形予測法とは大きく異なった特性となる。すなわち、 $p > p_0$ の場合には、両方法ともホルマントに対応しない擬似ホルマントが生じるが、通常の線形予測法では擬似ホルマントの帯域幅が一般に大きな値として推定されるため、ホルマントに対応する極への影響が小さく、分析次数を $p > p_0$ としても、ホルマント周波数推定精度が急激に悪くなることはないのに対し、本方法では全ての極の位置が z 平面の単位円上に拘束されているため、擬似ホルマントの存在が大きく影響し、 $p > p_0$ となるとホルマント周波数推定精度が悪くなると言える。

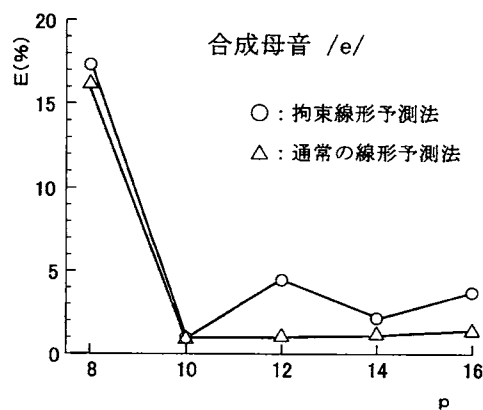


図 2 ホルマント周波数推定誤差 E の分析次数 p 依存性

4. 2 極周波数推定値の頑健性

図 2 から明らかなように、分析次数 p を $p < p_0$ となるとホルマント周波数推定誤差が極端に悪化する。原理的には $p = p_0$ に設定するのが適切であるが、実音声の正確な p_0 は未知であることから、一般には少し大きめに設定し $p > p_0$ となるようにしている(標本化周波数を 10kHz すなわち周波数帯域を 5kHz までに限定した場合、成人の発声による実音声にはこの帯域内に 5 個前後のホルマントが存在すると推測できるので、一般に $p=12$ が使用されている)。しかしながら、通常の線形予測法において分析次

数 p を $p > p_0$ とした場合、母音定常部においても分析位置により、特に擬似ホルマントの位置が大きく変動するため、得られる極の全てを音声の特徴パラメータとして利用することができず、ホルマントに対応する極の選定が必要となる[8]。一方、本方法は3. で述べたように全ての極の位置が z 平面の単位円上に拘束されているため、定性的には推定極の位置が予測誤差に大きく影響し、予測誤差の自乗平均最小の条件より、母音定常部のようなホルマント周波数がほぼ一定とみなせる音声区間においては、 $p > p_0$ の場合でもホルマントに対応しない極を含めて極周波数推定値が分析位置等の影響で大きく変動することはないと言える。

一例として、前述の合成母音/e/ における推定極の分析位置依存性を図3に示す。ただし、分析次数 $p=12 (> p_0)$ 、フレームシフト間隔 0.3ms とし、その他の分析条件ならびに図中の記号等は図2と同じである。なお、図3の上段には本合成音/e/の波形を示す。

図3より、通常の線形予測法による推定極(Δ 印)の内、ホルマントに対応する極は分析位置に拘わらずほぼ一定である

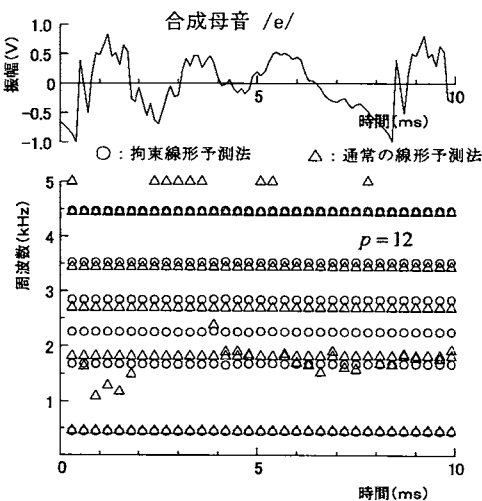


図3 推定極周波数の分析位置依存性 (合成母音 /e/)

が、擬似ホルマントが 1.2kHz 付近(分析位置：1ms~2ms の区間)、あるいは第2ホルマント前後(分析位置：5.5ms 以降)、さらにはときおり 5kHz と非常に不安定に生じているのに対して、本手法による推定極(\circ 印)はホルマントに対応する極であるか否かに拘わらず分析位置依存性はほとんどないと言える。このことを定量的に評価するため、本合成母音/e/ において、分析次数 $p=12$ として得られる6個($= p/2$)の極周波数各々の標準偏差を平均した式(14)で定義する極周波数推定値の平均標準偏差を算出した結果、従来の線形予測法では 50.4Hz であったのに対して、本手法は 3.6Hz と一桁以上改善していることが明らかとなった。ただし、前処理として一階差分後、分析窓長 $T_a = 25.6ms$ 、フレームシフト間隔 0.2ms で1周期に渡って分析した計 40 フレームの内、通常の線形予測法において極周波数推定値が 0Hz または 5kHz となった8フレームは除外したので、通常の線形予測法では式(14)の $M=32$ であるが、本手法では、そのような分析フレームは生じなかったので $M=40$ である。

$$\bar{\sigma} = \frac{2}{p} \sum_{i=1}^{p/2} \sqrt{\frac{1}{M} \sum_{j=1}^M (F_{ij} - \bar{F}_i)^2} \quad (14)$$

$$\text{但し、} \bar{F}_i = \frac{1}{M} \sum_{j=1}^M F_{ij}$$

F_{ij} : 第 j 分析フレームの第 i 極周波数

M : 分析フレーム数

p : 分析次数

5. 自然音声への適用例

前節の結果は、母音定常部のようなホルマント周波数が時間的にほぼ一定とみなせる音声区間においての結果であり、実音声ではホルマント周波数は時々刻々変化しており、音韻によっては急激に変化している。このような実音声においても同様のことが言えるかどうかをホルマント周波数が時間的に急変している代表的音声

である自然有声破裂音の極周波数推定に適用し、本手法の有効性を検証する。

成人男性が発声した有声破裂音/de/の極周波数推定例を図4に示す。ただし、前処理として一階差分を行い、分析次数 $p=12$ 、分析窓長 $T_u=25.6\text{ms}$ 、フレーム間隔 2ms で分析した結果であり、図中の○印および△印の意味は図3と同じである。

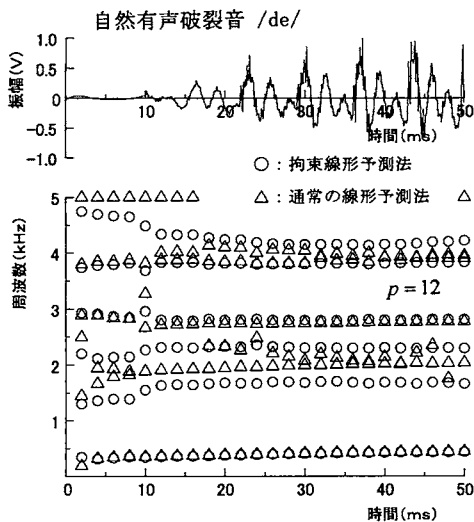


図4 推定極周波数の分析位置依存性
(自然有声破裂音 /de/)

図4より、ホルマント周波数が急激に変化する破裂時点(今の場合、10msの時点)前後での推定極の時間的変化が通常の線形予測法では不明確(特に、1.8kHz付近の推定極の時間的変化)であるのに対し、本方法では何れの推定極もフレーム間の連続性を保持した極が推定されており、ホルマント周波数の時間的変化をより正確に追尾していると言える。そして、音波形的にはまだ過渡区間ではあるが、ホルマント自体は時間的にそれほど変化していないと思われる25ms以降において、通常の線形予測法では1.5kHz以上の推定極のフレーム間安定性に難点があるのに対し、本方法ではフレーム間の連続性を保持した安定な極が推定されており、本

手法は実音声の極推定にも有効であると言える。

しかしながら、自然音声では真のホルマント周波数が未知であるので、その誤差を定量的に評価できない。したがって以下、破裂時点での第2、第3ホルマント周波数を特徴パラメータとしてホルマント空間での自然有声破裂音識別を行い、本手法の有効性を検討する。具体的には、第2-第3ホルマント周波数空間での/b/、/d/、/g/各クラスの重心からの距離による識別率ならびに類間分散と類内分散の比である分散比の良さで評価した。

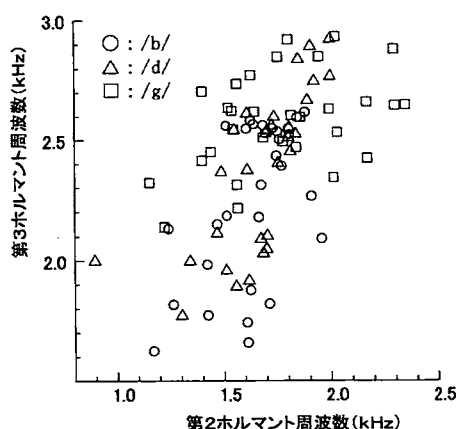
図5に第2-第3ホルマント周波数空間での/b/、/d/、/g/の分布図を示す。ただし、前処理として一階差分を行い、分析次数 $p=12$ 、分析窓長 $T_u=25.6\text{ms}$ で破裂時点を分析して得られる極周波数のうち単純に小さい順に第2番目及び第3番目の極をそれぞれ第2ホルマント周波数、第3ホルマント周波数と選定した場合の分布図で、それぞれ図5(a)が通常の線形予測法、同図(b)が本方法による分布図である。そして、図中の○、△及び□印はそれぞれ/b/、/d/及び/g/の位置を示す。なお、音声資料は、電子協日本語共通音声データベース中の20代及び30代の男性30人の単音節/be/、/de/、/ge/(ただし、2回目の発声)計90個である。

図5より、通常の線形予測法では、第2ならびに第3ホルマント周波数のバラツキが大きいのにに対して、本方法により、それらが大幅に改善しており、とりわけ、第2ホルマント周波数のバラツキが大きく改善していると言える(縦軸のスケールは(a)(b)両図とも同じであるが、横軸のスケールが異なることに留意)。

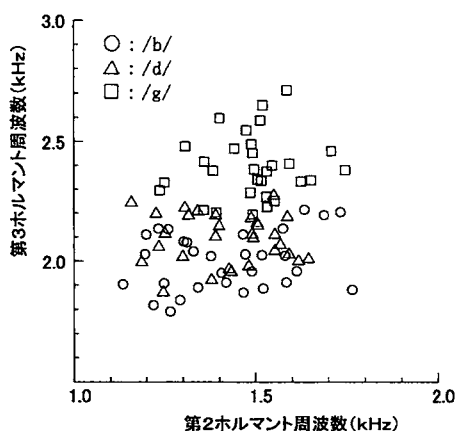
また図(a)より、通常の線形予測法では/g/(□印)の分布にバラツキはあるものの/g/の第3ホルマント周波数が相対的に高く推定され比較的まとまった分布となっ

ているのに対して、/b/(○印)、/d/(△印)の分布が相互に重複し、明確なクラスターを形成していないと言える(特に、/d/が/g/の領域にも広がっていると言える)。これに対して、本方法では、/g/の分布のまとまりが、より明確になると共に/b/、/g/それぞれの分布の様子が大幅に改善されていると言える。

この分布の違いを定量的に評価するため、第2-第3ホルマント空間において、



(a) 通常の線形予測法



(b) 拘束線形予測法

図5 ホルマント空間における有声破裂音(後続母音/e/)の分布

/b/, /d/, /g/相互を各重心からのユークリッド距離により識別した結果を表1に示す。表1(a)より、通常の線形予測法では、/g/において音声資料30個中4個が/b/、別の4個が/d/にそれぞれ誤識別されているが、残り22個は/g/と識別され識別率が73.3%と比較的正確に識別されている。しかしながら、上述したように/d/の分布の広がりが大きい為、正しく/d/と判定されるのは音声資料30個中4個しかなく、識別率は13.3%と極端に悪化している。

表1 有声破裂音の識別結果
(後続母音 /e/)

(a) 通常の線形予測法

	/b/	/d/	/g/	識別率(%)
/b/	13	5	12	43.3
/d/	11	4	15	13.3
/g/	4	4	22	73.3
分散比	0.15			平均 43.3

(b) 拘束線形予測法

	/b/	/d/	/g/	識別率(%)
/b/	20	7	3	66.7
/d/	11	18	1	60.0
/g/		3	27	90.0
分散比	0.85			平均 72.2

そして、/b/に関しては/d/ほどではないが、識別率は43.3%にしかならず、/b//d//g/の平均識別率は43.3%に留まっている。これに対して、本方法により、/g/が/b/に誤識別されることはなくなり、30個中3個のみ/d/と誤識別されるだけとなり、/g/の識別率が73.3%から90.0%へと改善すると共に、/b/、/d/それぞれの識別率も改善し、特に/d/の識別率が13.3%から60.0%に大幅に改善している。その結果、/b//d//g/の平均識別率が43.3%から

72.2%に大きく改善し、かつ分散比(= 類間分散/類内分散)も 0.15 から 0.85 に改善していることから、本方法の有効性が示されていると言える。

6. むすび

ディジタル信号処理手法として広く活用されている線形予測法により得られる予測係数と観測系の極の関係を考察することにより、予測係数間に簡単な関係を付与すれば、推定極の位置を z 平面の単位円上に拘束することができ、正規方程式の次元ならびに高次方程式の次数を通常の線形予測法の 1/2 に半減できると共に観測信号の極情報を担った安定な極周波数が得られることを示した。

本方法の有効性を合成音ならびに自然有声破裂音のホルマント周波数推定に適用して検討した結果、分析次数を適切に設定する必要があるが、通常の線形予測法とほぼ同等の精度でホルマント周波数推定が可能であることが合成音のシミュレーションにより明らかとなり、また合成音はもとより自然音声においても極の位置の拘束が極周波数推定値の頑健性にも大きく寄与することが実験的に示された。そして、自然有声破裂音(成人男性 30 名の単音節/be//de//ge/計 90 個)の破裂時点での第 2, 第 3 ホルマント周波数による有声破裂音識別に適用した結果、通常の線形予測法による識別率が 43.3%であった

のが、本方法により、72.2%に向上した。この大幅な改善は本方法により、音声のホルマント構造を担った安定した極周波数が得られることを意味し、本方法により得られる極周波数は音声認識における有効な特徴パラメータとなり得ることが明らかとなった。

文 献

- [1]J. Makhol : "Linear Prediction : A Tutorial Review", IEEE Proc., 63-4, pp. 561-580(1975).
- [2]板倉, 斎藤 : "統計的手法による音声スペクトル密度とホルマント周波数の推定", 信学論(A), 53-A, 1, pp. 35-42(1970).
- [3]B. S. Atal and S. L. Hanauer : "Speech analysis and synthesis by linear prediction of the speech wave", J. Acoust. Soc. Amer., 50, pp. 637-655(1971).
- [4]日本数学会編 : 岩波数学辞典(第 3 版), p. 705, 岩波書店(1990).
- [5]A. E. Rosenberg : "Effect of glottal pulse shape on the quality of natural vowels", J. Acoust. Soc. Am., 49, pp. 583-590(1971).
- [6]嵯峨山, 板倉 : "複合正弦波モデルによる音声スペクトルの分析", 信学論(A), J64-A, 2, pp. 105-112(1981).
- [7]菅村, 板倉 : "線形予測係数の線スペクト表現とその統計的性質", 信学論(A), J64-A, 4, pp. 323-330(1981).
- [8]粕谷, 和田, 岡田 : "線形予測分析法で得られる極周波数からのホルマント周波数選択アルゴリズム", 信学論(A), J66-A, 11, pp. 1144-1145(1983).