音声のホルマント周波数推定の高精度化に関する研究

昭和63年9月

三　好　義　昭

内容梗概

本論文は，筆者が大阪大学産業科学研究所 角所 収教授の指導のもとに行った研究の内，音声のホルマント周波数推定の高精度化に関する研究の成果をまとめたものであり，次の6章をもって構成されている．

第1章は序論で，音声研究の歴史的経緯と社会的背景を概説し，本研究の目的ならびにその音声工学上の意義および位置付けについて述べている．

第2章では，人間の音声生成過程とそのモデル化について概説し，生理学的には喉から唇または鼻孔までの空間—これを声道と称する—が音声生成に重要な役割を担っており，声道伝達特性，特にその極周波数であるホルマント周波数が音声の伝送・認識において重要であることが示されている．そして，このホルマント周波数推定手法として，今日広く活用されている線形予測分析について概説し，通常の線形予測分析では，(1) 鼻子音のように声道伝達特性に零点のある音声，(2) ピッチ周期の短いいわゆる高ピッチ音声，(3) 急激な声道形状の変化を伴う音声の過渡部，などではホルマント周波数を正確に推定することが困難であることを示し，後の章への準備としている．

第3章では，零点のある音声のホルマント周波数を高精度で推定する手法について述べている．零点の存在する音声のホルマント周波数を正確に推定するためには，基本的には零点をも考慮した極零型モデルに基づく分析，すなわち自己回帰移動平均（ＡＲＭＡ）モデルに基づく分析手法を採用するのが妥当であると言えるが，このＡＲＭＡモデル分析を行う場合，音声生成系のＡＲ部ならびにＭＡ部の正確な次数，ならびに系への入力信号に関する情報が必要となる．本章では，先ず音声波の変形共分散行列の固有値に基づくＡＲ部の次数の推定法を提案している．次に，音声では系への入力である励振波形が観測できないことを考慮し，音声生成系が本質的に極零型であると考えられる音声においても励振源の情報を必要としない全極型モデルに基づく分析により，ホルマント周波数を高精度で推定する手法について述べている．

第4章では，高ピッチ音声のホルマント周波数を高精度で推定する手法について述べている．通常の線形予測分析ではホルマント周波数推定に励振源の影響が現われる

ので，特にピッチ周期の短いいわゆる高ピッチ音声のホルマント周波数を正確に推定することが困難であることを解析的に示し，このような音声のホルマント周波数を高精度で推定する一手法として，線形予測分析における残差情報の大局的な特徴を考慮して線形予測モデルに適合する音声標本の選択を行い，かつこの処理を2段階行う2段標本選択線形予測分析について述べている．

第5章では，音声の過渡部のホルマント周波数を高精度で推定する二つの手法について述べている．一つは，短時間周波数スペクトルに及ぼす分析窓の位置および窓長の影響の詳細な検討結果に基づいて，分析窓長を有声音の1ピッチ周期未満に短縮した1ピッチ周期内周波数分析であり，他の一つは，通常の線形予測分析によるホルマント周波数推定値の分析窓長依存性の過渡モデル音による解析結果に基づいて，分析窓の任意の点を固定して窓長を漸減させた一連の分析の結果から，窓長が零になる場合の値を外挿する窓長漸減型線形予測分析である．

第6章は結論で，本研究で得られた結果をまとめ，音声の高精度分析の分野に関する今後の展望と残された課題について述べている．

関連発表論文

A．学会誌掲載論文

1 三好，大和，角所："有声音の1ピッチ周期内周波数分析によるホルマント周波数抽出"，電子通信学会論文誌，**J61-A**, 7, pp.633-640(1978).

2 Y. Miyoshi, K. Yamato and O. Kakusho : "Order estimation of speech production model based on the eigenvalue ratios of quasi-covariance matrix", J. Acoust. Soc. Jpn.(E), **4**, 1, pp.45-47(1983).

3 三好，大和，柳田，角所："自己相関行列の近似再構成による極周波数の精密推定"，電子通信学会論文誌，**J68-A**, 12, pp.1389-1397(1985).

4 三好，大和，柳田，角所："2段標本選択線形予測法による高ピッチ音声の分析"，電子情報通信学会論文誌，**J70-A**, 8, pp.1146-1156(1987).

5 Y. Miyoshi, K. Yamato, R. Mizoguchi, M. Yanagida and O. Kakusho : "Analysis of speech signals of short pitch period by a sample-selective linear prediction", IEEE Trans. Acoust., Speech & Signal Processing, **ASSP-35**, 9, pp.1233-1240(1987).

6 三好，大和，柳田，角所："窓長漸減型線形予測分析による過渡的音声のホルマント周波数抽出"，電子情報通信学会論文誌，A分冊（１９８８年１０月掲載予定）

B．国際会議発表論文

1 Y. Miyoshi, K. Yamato and O. Kakusho: "Order setimation of AR model based on eigenvalues of covariance matrix of speech", 10th International Congress on Acoustics, A1-10.5(1980).

2 Y. Miyoshi, K. Yamato, M. Yanagida and O. Kakusho: "Analysis of speech signals of short pitch period by the sample-selective linear prediction", Proceeding of International Conference on Acoustics,Speech, and Signal Processing, pp.1245-1248(1986).

C．研究会発表論文

1 三好，大和，角所："音声の短時間スペクトル分析に関する考察"，電子通信学会技術研究報告，**EA73**-48, pp.45-52(1974).

2 三好，大和，角所："ＦＲＡＰＳ法によるフォルマント周波数の抽出"，電子通信学会技術研究報告，**EA74**-47, pp.17-24(1975).

3 三好，大和，角所："音声の線形予測分析に関する考察"，電子通信学会技術研究報告，**EA75**-56, pp.23-28(1976).

4 三好，大和，角所："線形予測法による有声音の１ピッチ周期内分析"，電子通信学会技術研究報告，**EA76**-53, pp.9-16(1977).

5 三好，大和，角所："音声波の共分散行列の固有値に基づく生成モデルの次数推定"，電子通信学会技術研究報告，**EA77**-53, pp.53-58(1978).

6 三好，大和，角所："音声波の固有値解析による生成モデルの次数推定"，電子通信学会技術研究報告，**EA78**-74, pp.25-30(1979).

7 三好，大和，角所："音声波の固有値解析による生成モデルの次数推定－標本化周波数の検討－"，電子通信学会技術研究報告，**EA79**-86, pp.31-36(1980).

8 三好，大和，柳田，角所："自己相関行列のスペクトル分解を用いた線形予測分析"，日本音響学会音声研究会資料，**S84**-7, pp.49-54(1984).

9 三好，大和，柳田，角所："標本選択線形予測法による高ピッチ音声の分析"，日本音響学会音声研究会資料，**S85**-22, pp.161-166(1985).

10 三好，大和，柳田，角所:"分析窓長漸減型線形予測分析による過渡部のホルマント周波数抽出"，電子情報通信学会技術研究報告，**SP87**-13, pp.41-48(1987).

目　次

第1章　序　　論

近年のディジタル信号処理技術の飛躍的な進歩による高度情報化社会への移行に伴い，計算機を主体としたいわゆる”知的情報システム”と人間とのコミュニケーション手段の高度化にはますます社会的なニーズが高まってきている．この人間と知的情報システムとの対話媒体として，現在では主に文字や図形が広く用いられているが，もう一つの重要な対話媒体と言える「音声」は人間相互の間の情報伝達の手段として，文化の発展過程で発達してきたものであり，人間にとって最も根源的な情報伝達手段であることから，音声を用いたヒューマンインターフェースの高度化は，人間と知的情報システムとの対話を最も自然な形で実現するものとして期待を集めている．この音声を用いたヒューマンインターフェースの高度化のためには，知的情報システムに具備すべき音声認識ならびに音声合成の機能の高度化が必要不可欠であるが，本研究はこの音声認識機能の高度化の基礎となる音声の精密分析に関する研究を行ったものである．

音声の工学的な研究は，1779年に，C.G.Kratzenstein が人間の声道（声帯から唇または鼻孔までの口の中の空間）を模擬した音響管を用いて音声の合成の実験を行ったとの記録があるように[1]，その歴史はかなり古くまでさかのぼることができるが，今日の音声研究の基礎は，1940年前後の H.Dudley による音声の帯域圧縮を目的としたチャンネル型ボコーダの研究[2]，R.K.Potter らによる音声の周波数スペクトルの時間的変化を2次元平面上に表現できる ”Visible Speech” の研究[3]にその端を発すると言える．すなわち，これらの研究により，音声波の中で言語的情報を担っているのは音声波の周波数スペクトル中のエネルギーが集中している部分，すなわち，声道伝達特性の極であるホルマントであることが実験的にも明らかとなり，またそれを視覚的に観測できるようになった意義は大きいと言える．そして，1955年の K.N.Stevens らによる調音レベルでのパラメータ表示に関する研究[4]，1960年の G.Fant による音声スペクトル生成に関する近似理論[5]により音声生成の理論的基礎が確立した．

そして，この時期に相前後して，音声の重要な特徴パラメータであることが明らかとなったホルマント周波数の抽出方法に関する研究[6]~[10]，ならびに種々の音声認識

に関する研究[11]~[19]が精力的に行なわれたが，信号処理技術の未発達による音声の特徴パラメータの自動抽出の困難さ等により，実用に耐える音声認識装置は遂に実現されず，音声認識の研究は一時停滞した．しかし，1965年の Cooley らによる高速フーリエ変換アルゴリズムの開発[20]に見られるように，1960年代後半から1970年代にかけてのディジタル信号処理技術のハード及びソフト両面での飛躍的な発展，そして板倉らによる音声の最尤スペクトル分析[21]，B.S.Atal らによる音声の線形予測法[22]の提案により，音声の特徴を少数のパラメータで記述でき，しかも比較的簡単な計算でホルマント周波数を求めることができる分析手法の確立，ならびに音声の時間軸を非線形に伸縮する時間正規化に動的計画法（ＤＰ：Dynamic Programing）の導入[23]等により，語彙を限定した単語音声認識に関しては，現在，一部実用化されるまでになり[24]，音声認識の研究は新たな段階へ進展した．

音声認識は認識対象別に分類すると，単語音声認識と連続音声認識に大きく分かれ，またそれぞれは話者を限定した特定話者型と誰の音声でも認識できる不特定話者型に区分される．もちろん，音声認識における究極の目標は不特定話者による連続音声認識であるが，現在，実用化されている単語音声認識の認識手法は各単語ごとに標準パターンをあらかじめ用意しておき，入力単語と標準パターンとのマッチング度合に基づき認識するいわゆる単語単位でのパターンマッチング手法をその基本としている．この単語単位でのパターンマッチング手法を基本とした連続音声中の単語音声認識も検討されてはいるが[25]~[27]，認識単語数の飛躍的な増大あるいは連続音声認識を実用化レベルで実現するには，記憶しておくべき標準パターン数ならびにパターンマッチング処理の計算量が膨大となるため，単語単位での認識手法では限界があり，音素単位での認識に基づく手法[28]~[30]を確立することが必要であると言える．

ところで，音声は，人間の発声器官によって生成される，言語として意味のある音響現象であるため，音声の音響的特質は発話者により千差万別ではあるが，言語の表現のための符号系であることから，同一言語体系を持つ人々に共通の性質を持っている．この言語としての共通の性質を音響現象に反映させる上で，生理学的には喉から唇または鼻孔までの空間が重要な役割を担っている．すなわち，人間は声道の共鳴作用を最大限に活用して音声を生成していると言える．したがって，音素単位での音声

認識を行なうためには，声道伝達特性，特にその極周波数であるホルマント周波数のより精密な推定が必要となる．このホルマント周波数推定手法として，近年，線形予測分析が広く用いられているが，そこでは次の３つの条件，

（１）音声生成系が全極型である．

（２）励振源が白色ガウス過程である．

（３）分析区間内において音声生成系が定常である．

が満足されている必要がある．しかし，音声現象では一般にこれらの条件は近似的にしか満たされていないのでホルマント周波数を精度よく推定できないという状況がしばしば生じる．特に，

（１）零点のある音声

（２）ピッチ周期の短いいわゆる高ピッチ音声

（３）音声の過渡部

ではこの状況に陥り易い．本研究はこれらの音声あるいは音声区間における有効な分析手法について検討したものである．以下，

第２章で，音声の線形予測分析の基本となっている音声生成過程とそのモデル化について概説し，線形予測分析によるホルマント周波数推定ならびにその問題点について述べ，後の章への準備とする．

第３章で，零点のある音声のホルマント周波数を高精度で推定する手法として，一般には零点をも考慮した極零型モデルに基づく分析，すなわち自己回帰移動平均（ＡＲＭＡ）モデルに基づく分析が用いられるが，このＡＲＭＡモデル分析において重要となるＡＲ部の次数を音声波の変形共分散行列の固有値に基づいて推定する手法，ならびに自己相関行列の主要な固有値を用いてこれを近似再構成して得られる極情報を担った自己相関係数を用いることにより，零点のある音声に対しても極零型モデルに基づく分析を行なうことなく全極型モデルに基づく分析により，ホルマント周波数を高精度で推定する手法について述べる．

第４章で，高ピッチ音声のホルマント周波数を高精度で推定する手法として，線形予測分析における残差情報の大局的な特徴を考慮して線形予測モデルに適合する音声

標本の選択を行い，かつこの処理を２段階行う２段標本選択線形予測分析について述べる．

第５章で，音声の過渡部のホルマント周波数を高精度で推定する手法として，二つの手法について述べる．一つは，短時間周波数スペクトルに及ぼす分析窓の位置および窓長の影響の詳細な検討結果に基づいて，分析窓長を有声音の１ピッチ周期未満に短縮した１ピッチ周期内周波数分析であり，他の一つは，通常の線形予測分析によるホルマント周波数推定値の分析窓長依存性の過渡モデル音による解析結果に基づいて，分析窓の任意の点を固定して窓長を漸減させた一連の分析の結果から，窓長が零になる場合の値を外挿する窓長漸減型線形予測分析である．

そして，最後に第６章で，本研究で得られた主な成果をまとめ，今後に残された課題を述べる．

第2章　音声生成過程と線形予測分析[31)~(34)]

2.1　緒　　言

人間の発話過程は，相手に伝えたい意味内容の言語的形式への変換，これに基づく発声器官への運動神経指令の発生，発声器官の種々の動きによる言語としての情報を担った音波すなわち音声波の生成に分けられる．すなわち大脳における言語的処理，小脳における運動神経指令，発声器官における音波生成の3段階からなると言えるが，ここでは発声器官における音波生成機構について概説し，音声の合成及び認識等の基礎となる音声生成モデルについて述べる．そして，この音声生成モデルに基づいた音声の線形予測分析について概説し，線形予測分析による声道伝達特性の極周波数であるホルマント周波数推定の問題点について述べ，後の章への準備とする．

2.2　音声生成過程

図2．1に音声生成に直接的に関与する生理器官を示す．これらの器官が発声器官と総称される．これらの器官は図2．1から明らかのように，呼吸あるいは食物の摂取を本来の目的とするが，これらの器官の協調により呼気エネルギーの一部が音響エネルギーに変換され音声波が生成される．発声器官における発声の機構は音源の生成，調音及び放射の3要素から成る．

2.2.1　音源の生成

音源には大きく分けて，声帯振動による呼気の断続をエネルギー源とする声帯音源と，声道（声帯から唇ないしは鼻孔にいたる音響的空間）中に形作られた狭めにおける呼気の乱流をエネルギー源とする乱流音源の2種類があり，前者は有声音の生成のためのもので有声音源，後者は無声音の生成のためのもので無声音源と呼ばれる．

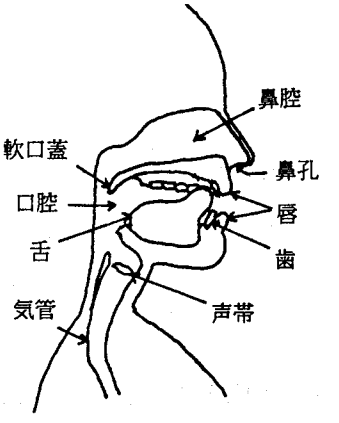


図2．1　発声器官

2.2.2 調　音

音源の生成に基づいて発生された音波に声道の共鳴作用によって周波数的にエネルギーの強弱を付与し，言語情報を担った音波に変換することを調音と言う．この調音に主として寄与する発声器官（舌，唇及び顎）を特に調音器官と呼ばれる．

2.2.3 放　射

調音により生成された言語としての情報を担った音波すなわち音声波は，唇または鼻孔から空間に放射される．この放射特性は無限大バッフルを持った唇の開口面積と等価なピストンからの放射と考えれば単調な約6dB/octの高域強調で近似できる．

以上の発声器官における発声の機構より，音声の言語としての情報を音響現象に反映させる上で，調音が重要な役割を担っていると言える．すなわち，人間は声道の共鳴作用を最大限に活用して音声を生成しており，声道伝達特性，特にその極周波数であるホルマント周波数が音声の伝送・認識において重要な特徴パラメータとなる．

2.3 音声生成モデル

発声器官における発声の基本要素である音源の生成，調音及び放射の特性が互いにそれぞれ独立であると仮定すると，音声波の周波数スペクトル $S(\omega)$ は

$$S(\omega) = U(\omega)H(\omega)R(\omega) \tag{2.1}$$

但し，　$U(\omega)$ ： 音源の周波数スペクトル
　　$H(\omega)$ ： 声道伝達特性
　　$R(\omega)$ ： 放射特性

と記述できる．ところで，放射特性 $R(\omega)$ は単調な約6dB/octの高域強調で近似し得るので，これを声道伝達特性に含めて，

$$H_e(\omega) = H(\omega)R(\omega) \tag{2.2}$$

なる等価声道伝達特性 $H_e(\omega)$ を考えれば，

$$S(\omega) = U(\omega)H_e(\omega) \tag{2.3}$$

となる．また，２．２．１で述べたように音源には大きく分けて２種類ありそれぞれの生成機構を考慮すれば，音源の周波数のスペクトル$U(\omega)$は

$$U(\omega) = U_0(\omega)U_e(\omega) \tag{2.4}$$

但し，　$U_0(\omega)$：パルス列（または白色雑音）のスペクトル
　　$U_e(\omega)$：音源のスペクトル包絡

と表現できる．したがって，

$$S(\omega) = U_0(\omega)U_e(\omega)H_e(\omega) \tag{2.5}$$

となる．以上のことより音声生成系は図２．２のようにモデル化することができる．

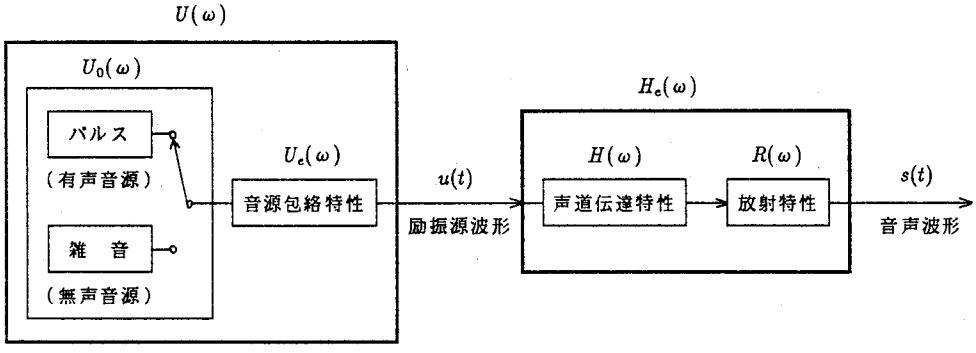


図２．２　音声生成モデル

2.4 音声の線形予測分析

音声の伝送・認識において重要となるホルマント周波数推定手法として，今日，線形予測分析[21],[22]が広く活用されているが，本節ではこの線形予測分析について概説し，通常の線形予測分析によるホルマント周波数推定の問題点について述べる．

2.4.1 線形予測分析によるホルマント周波数推定

今，等価声道伝達特性 $H_e(\omega)$ が全極型で記述できる場合には，音声波の第 n 標本値 s_n は図2．2より，

$$s_n = \sum_{k=1}^{p} \alpha_k s_{n-k} + u_n \tag{2.6}$$

となる．但し，u_n は励振源の第 n 標本値であり，$\{\alpha_k, k = 1,2,3,\cdots,p\}$ は予測係数と呼ばれる．この予測係数の推定値 $\{\hat{\alpha}_k\}$ は予測誤差 ε_n

$$\varepsilon_n = s_n - \sum_{k=1}^{p} \hat{\alpha}_k s_{n-k} \tag{2.7}$$

の2乗平均最小の条件より，

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} & \cdots & c_{1p} \\ c_{21} & c_{22} & c_{23} & \cdots & c_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & c_{p3} & \cdots & c_{pp} \end{bmatrix} \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_p \end{bmatrix} = \begin{bmatrix} c_{01} \\ c_{02} \\ \vdots \\ c_{0p} \end{bmatrix} \tag{2.8}$$

但し，　$c_{ij} = E\{s_{n-i}s_{n-j}\}$

なる正規方程式の解として求められる．ここで，$\{s_1, s_2, s_3, \cdots, s_N\}$ なる N 個の音声標本値から c_{ij} を推定する手法に従って，音声の線形予測分析には次の2種類の方式がある．

（1） 自己相関法

$$c_{ij} = r_{|i-j|} = \sum_{n=1}^{N-|i-j|} s_n s_{n+|i-j|} \tag{2.9}$$

（2） 共分散法

$$c_{ij} = \phi_{ij} = \sum_{n=p+1}^{N} s_{n-i}s_{n-j} \tag{2.10}$$

ところで，等価声道伝達特性の z変換 $H_e(z)$ は式（2．6）より，

$$H_e(z) = \frac{1}{1 - \alpha_1 z^{-1} - \alpha_2 z^{-2} - \alpha_3 z^{-3} - \cdots - \alpha_p z^{-p}} \tag{2.11}$$

となる．したがって，

$$1 - \alpha_1 z^{-1} - \alpha_2 z^{-2} - \alpha_3 z^{-3} - \cdots - \alpha_p z^{-p} = 0 \tag{2.12}$$

の根，$\{z_i, i = 1, 2, \cdots, p\}$ が等価声道伝達特性の極，すなわち音声のホルマントに対応するので，ホルマントの周波数 f_i 及び帯域幅 b_i はそれぞれ，

$$z = e^{sT} = e^{(\sigma + j\omega T)} \tag{2.13}$$

但し，　T ：標本化周期

より，

$$f_i = \frac{\omega_i}{2\pi} = \frac{1}{2\pi T} \arg(z_i) \tag{2.14}$$

$$b_i = \frac{-\sigma_i}{\pi} = \frac{-1}{\pi T} \log |z_i| \tag{2.15}$$

として求められる．なお，式（2．12）は実係数の p 次方程式であるので，その根は一般に[$p/2$]（但し，[　]：ガウス記号）個の共役複素根となる．したがって，式（2．14）により求められる独立なホルマント周波数は一般に[$p/2$]個となる．

2.4.2　ホルマント周波数推定の問題点

線形予測分析によるホルマント周波数の推定値は観測した N 個の音声標本値 $\{s_i, i = 1, 2, 3, \cdots, N\}$ の自己相関係数に基づく正規方程式の解として得られる p 個の予測係数の推定値 $\{\hat{\alpha}_k, k = 1, 2, 3, \cdots, p\}$ を係数とする p 次方程式の根より求められる．このように線形予測分析によれば，音声の伝送ならびに認識において特に重要となるホルマント周波数が明解でかつ比較的簡単な処理によって推定できることより，線形予測分析は今日の音声分析の根幹をなしていると言える．しかしながら，この線形予測分析により，ホルマント周波数が精度よく推定できるためには，以下の３つの仮定が満足される必要がある．

(1)音声生成系が全極型である.

線形予測分析における等価声道伝達特性 $H_e(z)$ が式(2.11)となることから明らかなように,通常の線形予測分析は音声生成系を全極型であると仮定している.

母音のように声道に分岐がなく,かつ励振源が声道端である声門にある場合には,この仮定を十分満足するが,摩擦音のように励振源が声道中にある場合,あるいは励振源が声道端である声門にあっても,鼻子音のように声道に分岐がある場合には,声道伝達特性に零点が存在するため,音声生成系は極零型となり,これらの音声はこの仮定を満足しないと言える.

(2)励振源が白色ガウス過程である.

通常の線形予測分析は予測係数の推定値 $\{\hat{\alpha}_k\}$ を式(2.7)で定義する予測誤差 ε_n の2乗平均最小の条件より求めているが,式(2.7)から明らかなように,予測誤差 ε_n の評価に励振源である u_n の項が含まれていない.したがって,通常の線形予測分析には励振源が白色ガウス過程であるとの仮定が存在する(第4章で詳述).

2.2.1で述べたように励振源には,典型的なものとして声帯振動による有声音源と,声道中の狭めにおる乱流による無声音源があるが,これらの音源包絡特性を等価声道伝達特性に含めて考えれば,励振源はそれぞれパルス列ならびに白色雑音となる(図2.2参照).したがって,無声音源ならびにピッチ周期の比較的長い有声音源の場合には,励振源はほぼ白色ガウス過程と近似できるが,ピッチ周期の短いいわゆる高ピッチ音声の場合には励振源を白色ガウス過程とみるには無理があると言える.

(3)分析区間内において音声生成系が定常である.

通常の線形予測分析は式(2.6)において予測係数 $\{\alpha_k\}$ が n すなわち時間の関数ではないことから明らかなように,分析区間内において音声生成系が定常であるとの仮定がある.

例えば母音定常部のように通常の分析区間(一般に 20 ～ 30ms)内において声道形状がほぼ一定と見なせる場合には,この仮定を十分満足するが,子音から母音への遷移部のように声道形状が急激に変化する音声の過渡部ではこの仮定を満足しないと言える.

以上のように，通常の線形予測分析には，これらの仮定が満足されない音声あるいは音声区間では正確なホルマント周波数推定ができないことがしばしば生じると言った問題がある．

2.5 結　　言

発声器官における音声生成機構に基づく音声生成モデルについて概説した．この音声生成モデルから明らかなように，人間は声道の共鳴作用を最大限に活用して言語情報を担った音波すなわち音声波を生成していると言え，音声認識を行なう上で声道伝達特性のより精密な推定が重要となる．この声道伝達特性の極周波数であるホルマント周波数の推定手法として，今日，広く活用されている線形予測分析について概説し，通常の線形予測分析により，ホルマント周波数を正確に推定できるためには，

(1)音声生成系が全極型である．

(2)励振源が白色ガウス過程である．

(3)分析区間内において音声生成系が定常である．

の3つの仮定を満足する必要があることを述べた．

通常の線形予測分析では，これらの仮定が満足されない音声あるいは音声区間（例えば，極零型と考えるべき鼻子音，励振源を白色ガウス過程と近似するには無理なピッチ周期の短いいわゆる高ピッチ音声，破裂音のように声道形状が急激に変化する音声区間）では正確なホルマント周波数推定ができないことがしばしば生じる．次章以下では，このような音声あるいは音声区間においても正確なホルマント周波数推定を可能とする分析手法について述べ，その有効性を示す．

第3章　零点のある音声の分析

3.1　緒　　言

本章では，鼻子音あるいは鼻音化音のように音声生成系が本質的に極零型であると考えられる場合，すなわち音声生成系をＡＲＭＡ過程で記述するのが適当である場合のＡＲ部の次数の一推定法，ならびに零点のある音声においても極零型モデルに基づく分析を行なうことなく全極型モデルに基づく分析により，ホルマント周波数を正確に推定する分析手法を示す．

3.2　変形共分散行列の固有値に基づくＡＲ部の次数推定[35],[36]

鼻子音あるいは鼻音化音のように零点の存在する音声のホルマント周波数を正確に推定するためには，基本的には零点も考慮した極零型モデルに基づく分析，すなわち自己回帰移動平均（ＡＲＭＡ）モデルに基づく分析を行なえばよいと言える．このＡＲＭＡモデル分析を行なう場合，音声生成系のＡＲ部ならびにＭＡ部の正確な次数を推定する必要がある．現在，種々の次数推定方法が検討されているが[37]~[39]，いずれも残差に対する評価基準最小の条件より，次数推定を行なっている．したがって，予測されるＡＲ部ならびにＭＡ部の全組み合わせに対して残差を評価する必要がある．本節では，有声音の声門閉止区間における音声波の変形共分散行列の固有値を利用すれば，音声生成系のＡＲ部の次数が残差等を求めることなく，かつＭＡ部とは独立に推定できることを示す．以下，３．２．１において，音声波の変形共分散行列の固有値の特徴を示し，３．２．２で，合成音のシミュレーションによりその検証を行なう．そして，３．３．３では，実際に鼻子音のＡＲ部の次数推定に適用し，本方法の有効性を示す．

3.2.1　変形共分散行列の固有値

音声生成過程を式(３．１)に示すＡＲ部の次数が p_0，ＭＡ部の次数が q_0 なるＡＲＭＡ過程と仮定する．

$$s_n = \sum_{k=1}^{p_0} \alpha_k s_{n-k} + \sum_{m=0}^{q_0} \beta_m u_{n-m} \tag{3.1}$$

但し，s_n及びu_nはそれぞれ音声波ならびに励振源の第n標本値，α_kは第kＡＲパラメータ，β_mは第mＭＡパラメータである．今，この音声生成過程から得られるN個の音声標本値$\{s_0, s_1, s_2, \cdots, s_{N-1}\}$を用いて，変形共分散行列$\Phi^{(i_0)}$を式(３．２)のように定義する．

$$\boldsymbol{\Phi}^{(i_0)} = \left(\phi_{ij}^{(i_0)}\right) \tag{3.2}$$

$$\phi_{ij}^{(i_0)} = \sum_{n=p+i_0}^{N-1} s_{n-i_0-i} s_{n-j}, \quad i, j = 1, 2, \cdots, p$$

今，行列$\Phi^{(i_0)}$の次数pがp_0より大きければ，式(３．１)，(３．２)より，

$$\phi_{ij}^{(i_0)} = \sum_{k=1}^{p_0} \alpha_k \phi_{i,j+k}^{(i_0)} + \sum_{m=0}^{q_0} \beta_m \sum_{n=p+i_0}^{N-1} s_{n-i_0-i} u_{n-j-m} \tag{3.3}$$

但し，　$i = 1, 2, \cdots, p$，　$j = 1, 2, \cdots, p - p_0$

となる．ところで，分析区間を$u_n = 0$となる区間に取れば，式(３．３)より行列$\Phi^{(i_0)}$の階数は音声生成系のＡＲ部の次数p_0に等しくなる．有声音の声門閉止区間が$u_n = 0$となる区間の候補となりうる．しかしながら，実際の音声では，声門閉止区間においてu_nが完全に零となっている保証はないが，少なくともこの区間ではu_nは振幅レベルの小さな白色雑音状と考えられる．したがって，分析区間を声門閉止区間に設定し，かつ$i_0 > q_0$とすれば，線形システムの系への入力u_nと系の出力s_nの因果律から式(３．４)の近似が成立する．

$$\sum_{n=p+i_0}^{N-1} s_{n-i_0-i} u_{n-j-m} \simeq 0 \tag{3.4}$$

但し，　$i \geqq j$，　$j = 1, 2, \cdots, p - p_0$，　$m = 0, 1, 2, \cdots, q_0$

この場合，式(３．３)，(３．４)より，$\phi_{ij}^{(i_0)}$は$\phi_{i,j+k}^{(i_0)}$の線形結合で近似できることになり，行列$\Phi^{(i_0)}$の階数はほぼp_0となる．したがって，行列$\Phi^{(i_0)}$の固有値をλ_i(但し，$|\lambda_i| \geqq |\lambda_{i+1}|$，　$i = 1, 2, 3, \cdots, p-1$)とすれば，$|\lambda_i / \lambda_{i+1}|$は$i = p_0$において最大となり，ＡＲ部の次数$p_0$が推定できると言える．

ところで,声門閉止区間のような分析区間の短い区間における式(3.4)の妥当性ならびに変形共分散行列の固有値比におよぼすその効果を解析的に明らかにするのは困難であるので，次節において本手法を合成ならびに自然音声の次数推定に適用しその有効性を示す.

3.2.2 合成音による検証

合成音作成のブロック図を図3.1に示す．ARMAシステムの励振源 u_n は声帯波[40] u_n^0 と白色雑音 w_n の和で，u_n の信号対雑音比 S/N を式(3.5)のように定義する.

$$S/N = 10\log_{10}\left(\sum_{n=0}^{N_0-1} u_n^{0^2}/N_0\sigma^2\right) \tag{3.5}$$

但し，$N_0 = T_0/T_s$, T_0: 声帯波のピッチ周期，T_s: 標本化周期，σ^2: w_n の分散である.

声道伝達特性として5個の極と一個の零（いずれも共役複素根）を持ち，さらに実軸に二つの零点を与えた．一つは放射特性，他はスペクトルの傾斜補正である．したがって，本合成音のAR部の次数 p_0 およびMA部の次数 q_0 はそれぞれ $p_0 = 10$, $q_0 = 4$ となる．シミュレーションに用いた極及び零点の周波数ならびに帯域幅の典型例を表3.1に示す．なお，合成音として零点のある音声を用いたのは，音声がたとえARMAモデルで生成されていたとしても本手法によればAR部の次数がMA部とは独立に推定できることを示すためである.

行列 $\Phi^{(i_0)}$ の固有値比 $|\lambda_i/\lambda_{i+1}|$ の例を図3.2に示す．但し，励振源の S/N=40dB，分析区間を声門閉止区間，分析次数 $p = 14$ とし，図3.2(a),(b)は i_0 をそれぞれ0

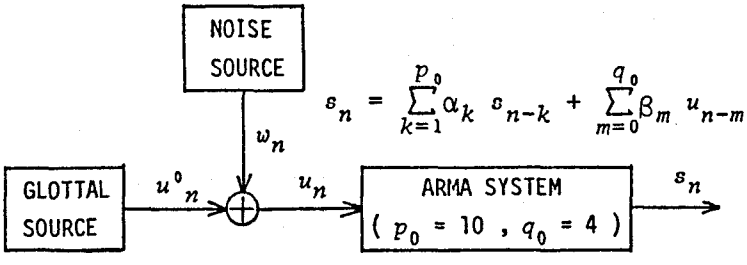


図3.1 合成音作成ブロック図

および 7 とした場合の結果である.

図3．2より，$i_0 = 0$ の場合には$\Phi^{(i_0)}$の固有値比は i=12 において最大となるが，$i_0 = 7(> q_0)$ とすれば $i = 10(= p_0)$ において最大となることが分かる．すなわち，i_0 を適切に設定すれば声門閉止区間のような比較的短い分析区間においても，行列 $\Phi^{(i_0)}$ の階数はほぼ p_0 となると言える.

表3．1　合成音の極および零点

	周波数 (Hz)	帯域幅 (Hz)
極	700.0	54.1
	1300.0	64.1
	2500.0	102.1
	3500.0	152.1
	4500.0	218.1
零点	1750.0	100.0

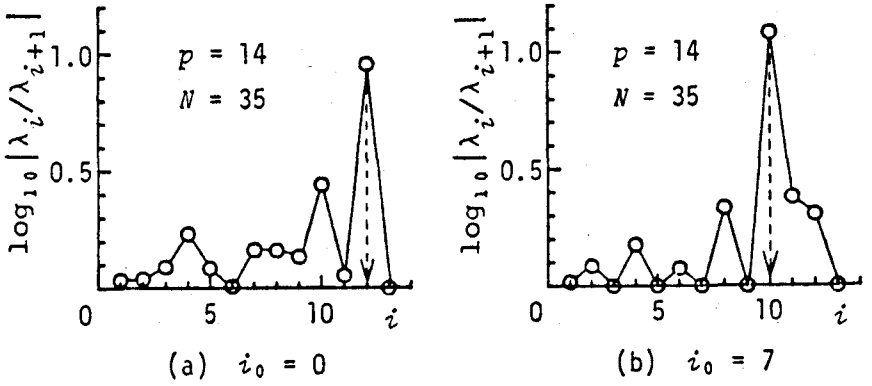


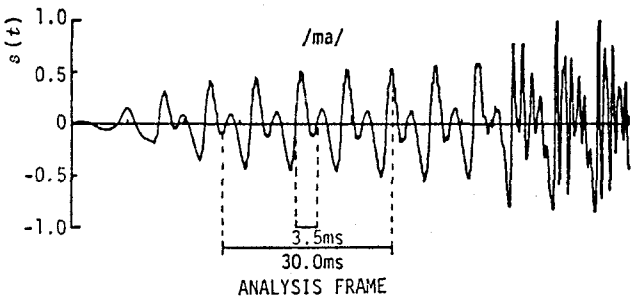
図3．2　行列 $\Phi^{(i_0)}$ の固有値比 $|\lambda_i/\lambda_{i+1}|$ の例
（但し，励振源の S/N = 40dB）

3.2.3　自然音声への適用結果

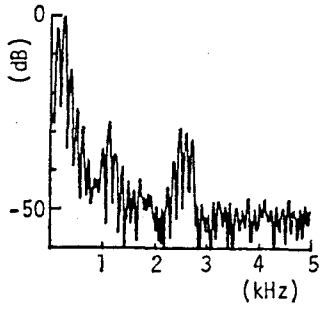
自然音声における本手法の有効性を確かめるために，鼻子音/m/，/n/と5母音の組み合わせによるＣＶ型単音節を成人男性2名が各一回発声した計 20 個の音声資料に本手法を適用した．

ところで，本手法を適用するためには分析区間を声門閉止区間に設定する必要がある．しかしながら，自然音声においてそのような区間を正確に検出することは一般に困難である．ここでは，この分析区間を比較的簡便な手順で設定する一つの方法として，分析フレームの始点を各ピッチ周期ごとの最大振幅値手前の零交叉付近に設定し，分析フレーム長を 1/2 ピッチ周期程度に設定する手法を用いた．

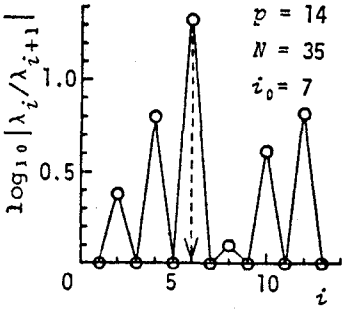
このようにして鼻子音/m/の部分に本手法を適用した一例を図3．3に示す．但し，図3．3（a）は適用した音声波形ならびに分析位置，（b）はその区間の周波数ペクトル，（c）は行列$\Phi^{(i_0)}$の固有値比をそれぞれ示す．本手法によるＡＲ部の次数の推定値は 6（図3．3（c）より $i=6$ において固有値比が最大）となる．ところで，図3．3（b）の周波数スペクトルより，本分析区間におけるスペクトル包絡上の優勢なピークの個数は3個であると言える．したがって，本鼻子音のＡＲ部の次数は 6 と推察されるが，本手法によるＡＲ部の次数の推定値は周波数スペクトルから推察される値と一致する．



(a)音声波形ならびに分析位置



(b)周波数スペクトル
(分析窓長 T_a = 30ms)



(c)行列 $\Phi^{(i_0)}$ の固有値比 $|\lambda_i/\lambda_{i+1}|$
(分析窓長 T_a = 3.5ms)

図3．3　自然音声におけるＡＲ部の次数推定例
（単音節/ma/の鼻子音/m/の部分）

3.3 自己相関行列の近似再構成によるホルマント周波数推定[41]

第2章で述べたように，音声生成系を全極型と仮定する通常の線形予測分析では鼻子音等のように声道伝達特性に零点が存在する場合にはホルマント周波数推定精度に限界がある．このような音声に対する分析方法としては，零点も考慮した極零型モデルに基づく方法[42]~[46]が有用であると言える．しかしながら，極零型モデルに基づく方法を用いる場合には，入力の推定ならびに極及び零部の適切な次数設定が必要であり，入力の推定ならびに次数設定が不適当な場合にはホルマント周波数推定精度がむしろ悪くなるといった問題がある[47],[48]．

ところで，極零型モデル分析の特長は極のみでなく零点の推定ができる点にあると言えるが，音声認識あるいは音声分析・合成系において零点パラメータの有用性に関しては若干の研究[49]はあるものの，まだ一般に認められるには到っていない．一方，最近全極型モデルによる分析で得られたホルマント周波数の時間的変化に注目した知識工学的手法に基づく連続音声認識が試みられており[50]，かなりの認識率が得られている．また，ホルマント構造に注目した鼻音性の検出[51]，あるいは全極型モデル分析による鼻子音の識別[52]などの検討も行なわれており，いずれも良い結果が報告されいる．これらの方法では零点に関する情報は用いられていない．これらのことより，零点が存在する場合でも，極零型モデル分析を行うことなく正確な極情報が推定できる分析方法は実用的には重要であると言える．

本節では，全極型モデル分析における自己相関行列に巡回性を導入し，さらに極情報の選択的利用を行なうことにより，声道伝達特性に零点が存在する場合でもホルマント周波数を正確に推定できる方法を示す．以下，3．3．1において，本方法の基本的な考え方を示し，3．3．2において，その基礎となる自己相関行列の近似再構成を導入し，3．3．3でＦＦＴを用いた実際の処理法を述べる．そして，3．3．4において，合成音のシミュレーションにより本方法のホルマント周波数推定精度の改善を示し，3．3．5では，本方法を実際に鼻音化母音の分析に適用した例を示すと共に鼻音化母音の認識実験によって本方法の有効性を示す．

3.3.1 極情報を強調した線形予測分析

通常の全極型モデルに基づく線形予測分析では，第2章で示したように，音声波の自己相関係数に基づく正規方程式の解を係数とする高次方程式の根からホルマント周波数が推定される．したがって，線形予測分析によるホルマント周波数推定において推定精度を左右するのは主として推定に用いる自己相関係数の性格であると言える．すなわち，声道伝達特性に零点が存在する場合でも，極の位置情報を強調した自己相関係数が得られれば，通常の全極型モデルに基づく手法でもより正確なホルマント周波数推定が可能であると言える．ここでは，極の位置情報を強調した自己相関係数を得るために，自己相関行列の固有値展開による近似再構成を導入し，さらに固有値を周波数スペクトル成分と対応づけるために，近似再構成の対象となる自己相関行列に巡回性を導入する．

本手法は巡回行列に関する次の定理[53]をその理論的根拠としている．

〔定理〕M次巡回行列

$$\boldsymbol{A}_M = \begin{bmatrix} 1 & a_1 & a_2 & \cdots & a_{M-1} \\ a_{M-1} & 1 & a_1 & \cdots & a_{M-2} \\ a_{M-2} & a_{M-1} & 1 & \cdots & a_{M-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & a_3 & \cdots & 1 \end{bmatrix} \tag{3.6}$$

の固有値λ_k及びλ_kに対応する固有ベクトルV_kは，Mが偶数の場合，

$$\lambda_k = 1 + 2\sum_{i=1}^{M/2-1} a_i \cos\frac{2\pi ik}{M} + a_{M/2}\cos\pi k \tag{3.7}$$

$$\left.\begin{aligned} \boldsymbol{V}_k &= \left(\cos\tfrac{2\pi k}{M}, \cos\tfrac{4\pi k}{M}, \cdots, \cos 2\pi k\right)^T \\ \boldsymbol{V}_{M-k} &= \left(\sin\tfrac{2\pi k}{M}, \sin\tfrac{4\pi k}{M}, \cdots, \sin 2\pi k\right)^T \end{aligned}\right\} \tag{3.8}$$

となる．

一方，音声波の第n標本値をs_nとし，任意の時点から相続くN個の標本値からなる音声標本値列を$\{s_n\} = \{s_1, s_2, s_3, \cdots, s_N\}$とすると，$\{s_n\}$の自己相関係数$r_i$を要素とする$M$次の自己相関行列$\boldsymbol{R}_M$は次の形となる．

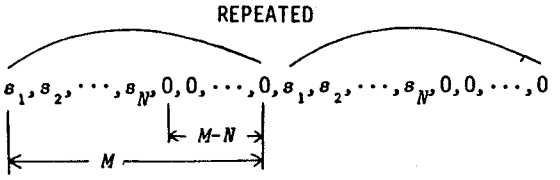


図3．4　自己相関係数算出の信号系列

$$\boldsymbol{R}_M = \begin{bmatrix} r_0 & r_1 & r_2 & \cdots & r_{M-1} \\ r_1 & r_0 & r_1 & \cdots & r_{M-2} \\ r_2 & r_1 & r_0 & \cdots & r_{M-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{M-1} & r_{M-2} & r_{M-3} & \cdots & r_0 \end{bmatrix} \tag{3.9}$$

ここで，r_i を

$$r_i = \sum_{n=1}^{M} s_n s_{n+i} \quad / \quad \sum_{n=1}^{M} s_n^2 \tag{3.10}$$

但し，$M \geqq N$，　$s_n = s_{n+M}$，　$s_{N+1}, s_{N+2}, \cdots, s_M = 0$

と定義すれば，すなわち，適当な窓関数により切り出したN個の音声標本値に$M-N$個の零を付加した時系列を考え，この時系列が周期Mで繰り返しているもの（図3．4参照）とすれば，$r_i = r_{M-i}$ となるので$\boldsymbol{R}_M$は巡回行列となり，前述の定理が適用できる．したがって，式(3．7)より，$\boldsymbol{R}_M$の固有値λ_kはr_iの第kフーリェ展開係数すなわち$\{s_n\}$のパワースペクトルにおける周波数k/MT(T: 標本化周期）に相当する成分に対応することが分る．

3.3.2 自己相関行列の近似再構成

一般に行列のスペクトル分解より，式(3．9)の $\boldsymbol{R}_M$ は

$$\boldsymbol{R}_M = \sum_{k=1}^{M} \lambda_k \boldsymbol{V}_k \boldsymbol{V}_k^T \tag{3.11}$$

但し，　λ_k: $\boldsymbol{R}_M$ の固有値
　　$\boldsymbol{V}_k$: λ_k に対応する固有ベクトル

と記述できる．すなわち，全ての λ_k を使うと $\boldsymbol{R}_M$ が完全に表現されるのであるが，λ_k をある基準によって選択的に使用することにより，スペクトルのローカルピークを強調することが可能になる．すなわち，前節の定理より $\boldsymbol{R}_M$ の固有値 λ_k が $\{s_n\}$ のパワースペクトルの周波数 k/MT に相当する成分に対応するので，λ_k の包絡上で極大となる付近の固有値及びそれに対応する固有ベクトルのみから式(3．11)に基づいて $\boldsymbol{R}_M$ を再計算すれば，その行として極情報のみを担った自己相関係数が得られると言える．したがって，この自己相関係数を用いた正規方程式を解けば，声道伝達特性に零点が存在する場合でも，零点の存在に影響を受けずにより正確なホルマント周波数推定が可能であると期待される．

固有値の選択手法として，一般に固有値の大きいものから順に使用するという方法がよく用いられているが，この方法では小さい固有値に対応するレベルの低いローカルピークを拾うことができないので，ここではこの選択手法は適さない．また通常の線形予測分析ではローカルピークを整合させると言うよりも全体の形を整合させるように予測係数が決まるが，ここで提案する方法によると，スペクトル上でのローカルピークに整合させるような予測係数が求まる．したがって，零点によってスペクトルに谷が生じても，ローカルピーク付近の包絡が歪まなければ，この方法ではそれには不感であるという利点がある．

【処理手順】

本方法は，行列の固有値展開にその基礎を置いてはいるが，式(3．7)より $\boldsymbol{R}_M$ の固有値が $r_i (i = 0, 1, 2, \cdots, M-1)$ の離散的フーリエ展開係数となっているので，実際の処理としては，M を N 以上の2のベキ乗とすることによりＦＦＴを利用することができる．

音声の分析では，一般に分析窓長として20～30msが用いられているが，この場合，標本化周波数を10kHzとすれば，音声標本点数Nは200～300となるため$\boldsymbol{R}_M$の次元Mは300程度となる．したがって，本手法では300×300程度の行列の固有値を求めることになるが，上記のようにＦＦＴを用いることができる．具体的にはN個の音声標本値に適当に0を付加して標本数を256又は512として一回のＦＦＴで全固有値が求まるので，計算上の問題はない．また，$\boldsymbol{R}_M$の近似再構成においては，$\boldsymbol{R}_M$の巡回性より任意の行または列のみを再計算すればよい．すなわち，$\boldsymbol{R}_M$の第1行目に注目すれば式(３．１１)より，

$$r_i = \sum_{k=1}^{M} \lambda_k v_{k1} v_{k,i+1} \tag{3.12}$$

但し，　v_{ki} ：固有ベクトル$\boldsymbol{V}_k$の第i成分

となる．さらに，線形予測分析において必要となる自己相関係数は分析次数をpとすればr_0～r_pであるので，式(３．１２)において$i=p$まで計算すればよいと言える．

これらのことより，本手法の実際の処理手順をまとめると以下のようになる．

【１】音声標本点数N及び$\boldsymbol{R}_M$の次元Mの設定．

但し，Mは２のベキ乗で$M \geqq N$とする．

【２】適当な窓関数によりN個の音声標本点を切り出し，$M-N$点の零を付加して，M点ＦＦＴを用いて$\boldsymbol{R}_M$の固有値λ_kを求める．

【３】λ_kの包絡上で極大となる付近の固有値及びそれに対応する固有ベクトルのみから式(３．１２)に基づいて自己相関係数$r_i(i=0,1,2,\cdots,p)$を近似再計算する．

【４】この自己相関係数を用いた正規方程式の解として得られる予測係数を係数とする高次方程式の根よりホルマント周波数推定を行なう．

3.3.3　合成音による検証

(a)　零点がある場合

標本化周波数10kHz，励振源：ピッチ周期8.0msのRosenberg波[40]，ホルマント周波数：$F_1=844$Hz，$F_2=1344$Hz，$F_3=2469$Hz，$F_4=3469$Hz，$F_5=4469$Hz，零点周波数：$F_Z=1750$Hz，放射特性：6dB/octとして作成した合成音における自己相

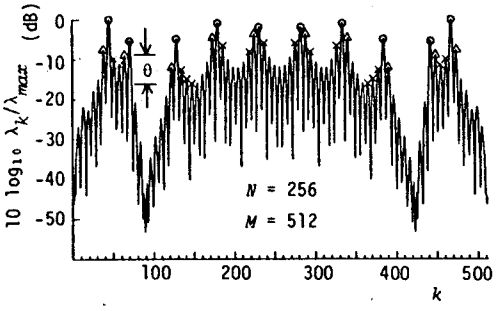


図3．5　自己相関行列 $\boldsymbol{R}_M$ の固有値パターンの例
○：固有値パターンの包絡上の極大固有値
△：○印に隣接する第1位の極大固有値
×：△印の固有値から閾値 θ 以内となる極大固有値

関行列 $\boldsymbol{R}_M$ の固有値パターンの例を図3．5に示す．但し，前処理として一階差分後，窓長 25.6ms の Hamming 窓 ($N = 256$) を用い $\boldsymbol{R}_M$ の次元 $M = 512$ とし，縦軸は $\boldsymbol{R}_M$ の固有値 λ_k をその最大値 λ_{max} で正規化した相対値(dB)，横軸は式(3．7)の形の固有値番号 k である．　図3．5より，$\boldsymbol{R}_M$ の固有値の中で主要な固有値（図中の○印）がホルマント情報を担っていることが分る．主要な固有値の選定方法として，ここでは λ_k の包絡上で極大となる固有値（図3．5の○印，但し，その位置が 300Hz 以下であって，かつそれよりレベルの大きい包絡上での極大固有値が 300Hz ～ 1kHz に存在する場合には除く）及び，その各固有値に隣接する第1位の極大固有値（図3．5の△印）とそれから閾値 θ (dB) 以内となる極大固有値（図3．5で $\theta = 5$dB とすると×印がこれに該当する）を主要な固有値とする方法を用いた．

本手法の精度を示すために，正確なホルマント周波数推定が一般に困難とされている場合，つまり，各ホルマント周波数が隣接する高調波の間の 1/4 あるいは 3/4 の位置にくるように設定[54]した表3．2に示す5種類の声道伝達特性を持つ合成音を基準として，本手法によるホルマント周波数推定精度の閾値 θ 依存性を調べた．

図3．6に閾値 θ と式(3．13)で定義するホルマント周波数推定誤差 E_i の関係を示す．但し，合成音として，表3．2の各 F_1～F_3及び F_Z を平均零の正規乱数（但し，標準偏差：それぞれ 50，50，100，100Hz）により 50 通りに変化させた計 250 個による結果である．

表3．2　基準とした合成音の声道伝達特性

(Hz)

	極周波数			零点周波数
	F_1	F_2	F_3	F_z
合成音1	844	1344	2469	1750
合成音2	344	2219	2844	1750
合成音3	344	1219	2219	1750
合成音4	594	1844	2469	1500
合成音5	594	1031	2469	1750

但し，$F_4 = 3469$ Hz，$F_5 = 4469$ Hz

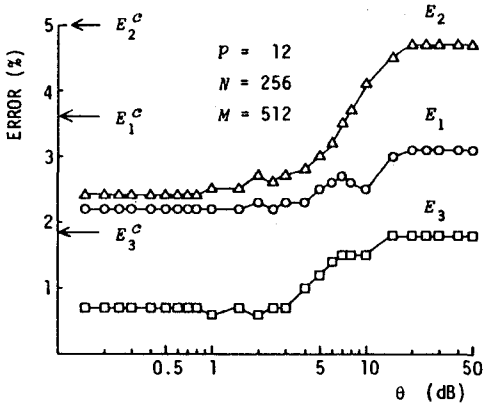


図3．6　零点のある場合のホルマント周波数推定誤差の閾値θ依存性

E_i: 本方法による第iホルマント周波数推定誤差

E_i^c: 従来の線形予測による第iホルマント周波数推定誤差

$$E_i = \overline{|\hat{F}_i - F_i| / F_i} \tag{3.13}$$

但し，　F_i ：第 i ホルマント周波数
　　$\hat{F}_i$ ：第 i ホルマント周波数の推定値

図3．6は，前述の方法で選定した主要な固有値及びそれに対応する固有ベクトルから式(3．12)に基づいて算出した自己相関関数を用いて，通常の線形予測分析（分析次数 p =12）を行った結果である．なお，図3．6において，比較のために通常の全極型モデルに基づく線形予測分析による第 i ホルマント周波数推定誤差 E_i^c を←印にて示す．本シミュレーションに使用した合成音は第1と第2あるいは第2と第3ホルマントの間に零点が存在（表3．2参照）するため，零点の影響を大きく受けていると考えられる第2ホルマントの推定誤差が大きくなっている．本合成音に対して適正な次数で極零型分析を行なえば正確な極情報が推定できることは自明であるが，本節の趣旨は零点が存在する場合でも極零型分析を導入しないで，通常の線形予測法よりも正確に極周波数を推定する分析手法を提案する点にあるため，通常の全極型モデルに基づく手法との比較を行なっている．

図3．6より，本方法において，θ ＞ 15dB の場合には，ホルマント周波数推定誤差は通常の全極型分析とほぼ同じとなるが，θ ≦ 15dB とすることによって零点の影響を受けた固有値が除去でき，ホルマント周波数推定誤差がθと共に減少し，θ ≦ 3dB ではホルマント周波数推定誤差は閾値θにはほとんど依存しなくなり，θ = 1dB では，第1～第3ホルマント周波数推定誤差の平均値が 3.5% から 1.8% に改善されていることが分る．

ところで，本方法は基本的にはスペクトル上でのピークピッキングと同種の処理を行っているとも言えるが，本方法において θ = 1dB とした場合の結果と 512 点のＦＦＴを用いたケプストラム上で式(3．14)の低域ろ波により得られたスペクトル包絡上でのピークピッキング[55]による結果の比較を表3．3に示す．

$$w(t) = \begin{cases} 1 & t < \tau \\ 0.5\{1 + \cos(\pi(t-\tau)/\tau\} & \tau \leq t \leq 2\tau \\ 0 & t > 2\tau \end{cases} \tag{3.14}$$

但し，　$\tau = 1.5$ ms

表3．3より，第1～第3ホルマント周波数推定誤差がいずれもスペクトル包絡上でのピークピッキングよりそれぞれ1.4%，2.5%，及び0.4%改善され，本方法の方がスペクトル包絡上でのピークピッキングよりホルマント周波数推定精度が良いことが分る．

表3．3　ホルマント周波数推定誤差（%）

	E_1	E_2	E_3
本　方　法	2.2	2.5	0.6
ピークピッキング法	3.6	5.0	1.0

（b）　零点のない場合

以上の合成音によるシミュレーション結果より，全極型モデルに基づく線形予測分析を基礎とした本方法は，声道伝達特性に零点が存在する場合でも正確な極情報の推定が可能であることが明らかとなったが，極付近の情報しか使用していないため，零点のない音声すなわち全極型モデルで本質的に記述ができる音声の分析には若干の問題が生じると思われる．この点を明らかにするために，ここでは零点のない合成音に対する検討を行った結果について述べる．

図3．7に，零点のない場合のホルマント周波数推定誤差の閾値θ依存性を示す．但し，声道伝達特性に零点がないことを除けばシミュレーションに用いた合成音ならびに分析条件等は零点ののある場合とまったく同じである．

図3．7より，3dB$<\theta\leq$15dBではホルマント周波数推定誤差が多少変動しているが，$\theta\leq$3dBではホルマント周波数推定誤差は閾値θにほとんど依存せず，$\theta=$1dBの時，第2ホルマント周波数の推定精度が通常の全極型モデルに基づく線形予測分析より0.2%劣るが，第1ホルマント周波数の推定精度は通常の方法より0.1%改善されており，第1～第3ホルマント周波数推定誤差の平均値としてはどちらも1.0%であり，本方法は零点のない音声に対しても通常の全極型モデルに基づく線形予測分析と同定度の精度でホルマント周波数推定が可能であることが分る．

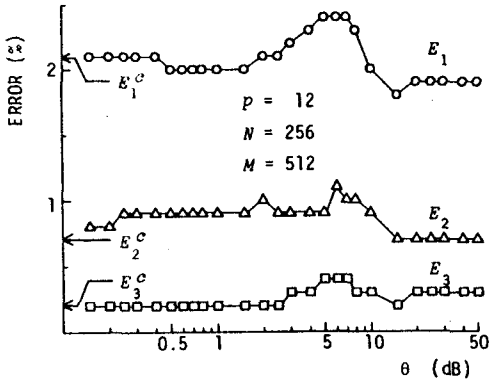


図3．7　零点のない場合のホルマント周波数推定誤差の閾値θ依存性
E_i: 本方法による第iホルマント周波数推定誤差
E_i^c: 従来の線形予測による第iホルマント周波数推定誤差

（c）　分析次数の検討

極情報のより正確な推定を目的とするならば，通常の全極型モデルに基づく線形予測分析においても零点が存在する場合には分析次数を大きくすることによりある程度対処できると言える．

図3．8に，零点のある場合と同様のシミュレーションにおいて分析次数のみを変化した場合のホルマント周波数推定誤差の分析次数p依存性を示す．但し，○印：本方法（$\theta = 1$dB），△印：通常の全極型モデルに基づく線形予測分析の結果である．

図3．8より，零点が存在する場合には通常の全極型モデルに基づく線形予測分析で

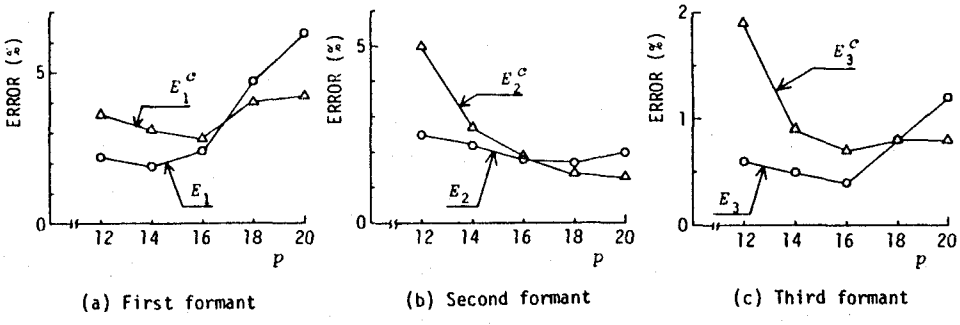


図3．8　ホルマント周波数推定誤差の分析次数p依存性
○：本方法（$\theta = 1$ dB，$N = 256$，$M = 512$）
△：従来の線形予測

も分析次数を大きくすれば，零点に近接することになる第2ホルマント周波数の推定精度の改善が特に著しいが，$p \geqq 18$ とあまり大きくすると第1及び第3ホルマント周波数の推定精度がかえって悪くなる傾向がみられ，分析次数は単に大きくすればよいとは一概に言えず，ある程度適切な分析次数の選定が必要となる．また適切な分析次数を選定できたとしても，分析次数が大きい場合にはホルマントに対応する極の選定問題が生じる（これに関しては3．3．4において具体的に述べる）．

一方，本方法は零点の存在如何にかかわらず分析次数を特に大きくする必要はないが，分析次数を大きくすれば第2ホルマント周波数の推定精度が $p = 12$ の場合よりも良くなっている．しかし $p \geqq 16$ において第1ホルマント周波数の推定精度が悪くなり，分析次数はあまり大きくしないほうが良いと言える．

3.3.4　自然音声への適用結果

本方法を実際に鼻音化母音の分析及び認識に適用し，通常の方法との比較検討を行なって本方法の有効性を明らかにする．

（a）　鼻音化母音の分析例

図3．9に成人男性の鼻音化母音/ĩ/及び/ũ/の定常部における周波数スペクトル包絡

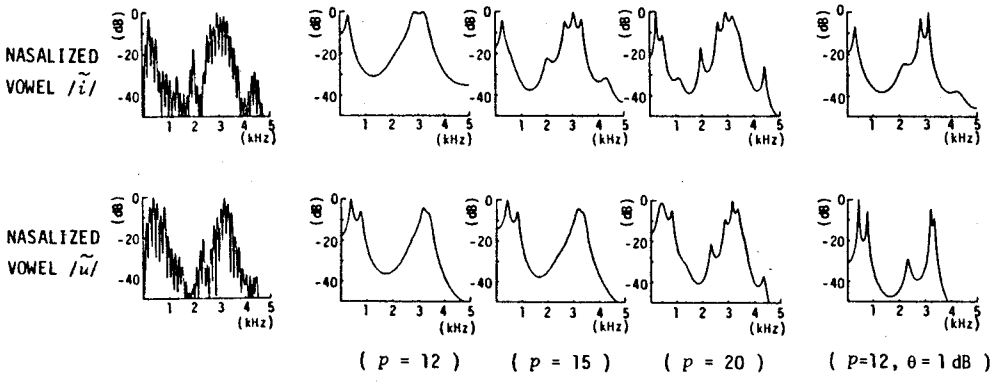


図3．9　鼻音化母音/ĩ/及び/ũ/の分析例

の比較例を示す．但し，前処理として一階差分の後，窓長 25.6ms の Hamming 窓を用い，同図(a)は 512 点のＦＦＴによる周波数スペクトル，同図(b)は通常の全極型モデルに基づく線形予測分析（分析次数 $p = 12$，15 及び 20）による周波数スペクトル包絡，同図(c)は本方法（分析次数 $p = 12$，閾値 $\theta = 1$ dB，$\boldsymbol{R}_M$ の次数 M =512）による周波数スペクトル包絡である．

図３．９の(a)，(b)を比較すると次のことが言える．

(1)通常の線形予測分析（分析次数 $p = 12$）では，ＦＦＴによる周波数スペクトルから推察されるホルマントの位置の極が鼻音化による零点の影響のために抑制されて明確なピークになっていない（/ĩ/の 2kHz 付近の第２ホルマント，/ũ/の 2.5kHz 付近の第３ホルマント）．

(2)零点が存在することを考慮して，通常の線形予測分析において分析次数を大きくした場合，

$p = 15$ の場合：鼻音化母音/ĩ/において，$p = 12$ では不明確であった第２ホルマントが明確に現れているが，3kHz 付近にいずれがホルマントに対応する極であるか判断し難い近接した明確な３個の極が生じる．一方，鼻音化母音/ũ/では $p = 12$ の場合とほとんど差がなくやはり第３ホルマントが不明確である．

$p = 20$ の場合：鼻音化母音/ũ/において，$p = 12$，15 では不明確であった第３ホルマントが明確に現れているが，第１ホルマント付近に近接した２個の極が生じ（見掛け上，第１ホルマントのバンド幅が大きく見える），かつ，3kHz 付近に近接する明確な３個の極が生じる．さらに，鼻音化母音/ĩ/においては，第１と第２ホルマントの間に，ホルマントに対応する極と同定されかねない２個の極が生じる．

このように，通常の線形予測分析で分析次数を大きくすることによって零点の存在に対処しようとした場合，音韻によって適切な分析次数が異なり，またバンド幅の情報のみでは個々のローカルピークについて，それがホルマントか否かを正しく判断するのは困難となる．これに対して，本方法では図３．９(c)のようにＦＦＴによる周波数スペクトルから推察されるホルマントの位置にいずれも明確なピークが存在している．

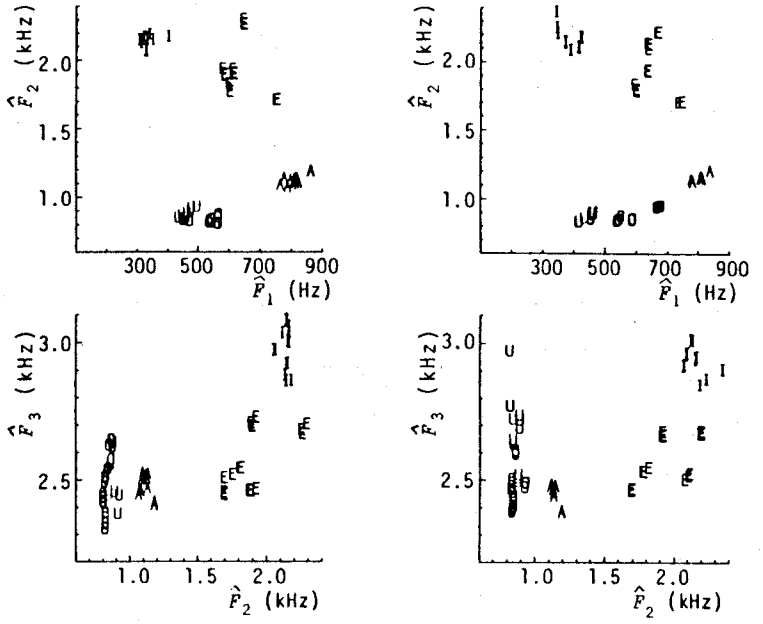
（b） 鼻音化母音の認識

図3．11に，第1，第2及び第3ホルマント周波数（$F_1-F_2-F_3$）空間における鼻音化5母音の分布の一例を通常の線形予測分析と比較して示す．但し，音声試料は成人男性1名が鼻音化5母音を各5回発声した定常部各3フレームについて，前処理として一階差分の後，窓長 25.6ms の Hamming 窓を用い，本方法は分析次数 $p=12$，閾値 $\theta=1$dB，R_M の次数 $M=512$，通常の線形予測分析は分析次数 $p=12$ とした．

図3．10より，次のことが言える．

(1) F_1-F_2 平面上において鼻音化母音/õ/が通常の線形予測分析では2つのクラスタに分かれているのに対して本方法では一つのクラスタを形成している．

(2) F_2-F_3 平面上の鼻音化母音/ũ/の分布より，通常の線形予測分析では第3ホルマント周波数の推定が不安定であるのに対して，本方法はより安定している．



(a) Proposed method
(p = 12, N = 256
M = 512, θ = 1 dB)

(b) Conventional method
(p = 12, N = 256)

図3．10　ホルマント空間における鼻音化5母音の分布

表3．4に成人男性4名が鼻音化5母音を各5回発声した定常部各3フレームについて，鼻音化母音の認識を話者別に第1，第2及び第3ホルマント周波数空間での各重心からの距離により行なった場合の認識結果を示す．

表3．4より，通常の線形予測分析でも鼻音化母音/ĩ/及び/ẽ/はいずれも認識率が100% となるが鼻音化母音/ũ/及び/õ/の認識率が悪いため平均認識率が 94% となるのに対し，本方法ではこれらの認識率の改善が著しく平均認識率は 99% となり，本方法の有効性が確認される．

表3．4　鼻音化5母音の認識結果

(a)本方法

($p=12, N=256, M=512$, $\theta = 1$ dB)

	/ã/	/ĩ/	/ũ/	/ẽ/	/õ/	認識率(%)
/ã/	58		2			97
/ĩ/		60				100
/ũ/			60			100
/ẽ/				60		100
/õ/			1		59	98

(b)従来の線形予測法

($p=12, N=256$)

	/ã/	/ĩ/	/ũ/	/ẽ/	/õ/	認識率(%)
/ã/	56		3		1	93
/ĩ/		60				100
/ũ/			55		5	92
/ẽ/				60		100
/õ/			10		50	83

3.4 結　　言

本章では，音声波の変形共分散行列の固有値の比の大きさを用いた音声生成系のＡＲ部の次数の一推定法，ならびに自己相関行列のスペクトル分解を利用し，極情報を担った自己相関係数を近似再構成することにより，零点のある音声においても極零型モデルに基づく分析を行なうことなく，ホルマント周波数を正確に推定する手法について述べた．

変形共分散行列の固有値に基づくＡＲ部の次数推定法の特長は音声がたとえＡＲＭＡモデルで生成されていたとしてもＭＡ部には影響されずにＡＲ部の次数が推定できる点にある．このことを明らかにするために，零点のある合成音ならびに零点のある代表的な音声と言える実際の鼻子音のＡＲ部の次数推定に適用し，本手法の有効性を示した．零点のある音声における本手法の満足すべき結果を考慮すれば，本手法は零点のない有声音には十分適用可能であることは明らかであると言える．

自己相関行列の近似再構成によるホルマント周波数推定法の特長は声道伝達特性の零点の存在如何にかかわらず正確なホルマント周波数推定ができる点にある．このことを明らかにするために，合成音のホルマント周波数推定精度ならびに自然音声の鼻音化母音の分析例及び認識率に関して，本方法と通常の全極型モデルに基づく線形予測分析との比較を行なった．その結果，250通りの合成音において第１～第３ホルマント周波数推定誤差の平均値が零点のない場合はいずれも 1.0% であるが，零点のある場合は 3.5% から 1.8% に改善され，自然音声の鼻音化母音においては周波数スペクトルからから推察されるホルマントの位置に明確なローカルピークを持ったスペクトル包絡が得られ，第１～第３ホルマント周波数推定値を特徴パラメータとした場合，鼻音化母音の認識率が 94% から 99% に改善されるとの結果が得られ，本方法の有効性が明らかになった．なお，本方法を自然音声に適用するにあたり本方法のパラメータである閾値θを合成音のシミュレーション結果に基づいて固定としたが，自然音声の多様性を考慮すれば閾値θは適応的に変化させる必要があると言えるが，この問題は今後の課題である．

第4章　高ピッチ音声の分析[56]~[58]

4.1　緒　　言

音声生成系が全極型モデルで記述でき，かつ定常的であると考えられる音声区間においても，第2章で述べたように通常の線形予測分析を用いて正確な声道伝達特性が得られるためには，励振源が白色ガウス過程でなければならないが，現実にはその仮定は近似的にしか満たされていないので，ホルマント周波数推定に励振源の影響が生じる．特に，女性あるいは子供が発声した有声音のように基本周期の短いいわゆる高ピッチ音声の場合，ピッチの影響を大きく受けて正確なホルマント周波数推定が困難となる状況がしばしば生じる．この有声音における励振源の影響を軽減する方法としては，有声音源のより実際的なモデル化[59]，あるいは分析窓長を1ピッチ周期以下と短くして声門閉止区間即ち自由振動区間のみを分析対象とする方法[60]~[62]などがある．有声音源のモデル化は，声帯波形関数の推定及び位相の問題等まだ未解決の重要な問題があり，今後の研究課題であると言える．一方，自由振動区間内分析では，声門閉止区間を正確に推定しておく必要があり，種々の方法[63],[64]が検討されてはいるが，自然音声の声門閉止区間を正確に推定するのは一般に困難で，特に女声のような高ピッチ音声の場合には，声門閉止区間の推定はより困難となる．本章では，励振源の影響を受けない正確な分析には，基本的には自由振動区間を対象とした処理が現時点では最良であるとの立場から，従来の自由振動区間内分析における上記の難点のない分析法としての標本選択線形予測分析の改良を示す．

従来の標本選択線形予測分析[65]は線形予測分析に一般逆行列を導入し，Givens変換に基づく逐次計算法を用いることによって各標本値を処理するごとにその選択的利用が効率良く行える利点がある．しかしながら，それは処理時間節減の目的から，各標本値を処理するごとにその標本値を使用するか否かを決定しているため，予測残差の大局的な特徴に基づく選択処理が行えないといった欠点がある．本章では標本の選択処理をフレーム単位で行うことにより，予測残差の大局的な特徴をも考慮して標本の選択を行い，かつこの処理を2段階行うことにより，従来の方式よりも被予測標本と

してより妥当な標本の選択が行える2段標本選択線形予測分析の有効性を示す．以下，4．2において，本方法の基本的な考え方を示し，4．3で合成音のシミュレーションにより本方法のホルマント周波数推定精度の改善を示す．そして，4．4では，本方法を実際に成人女性が発声した単音節の母音部の分析例ならびに連続音声のホルマント周波数抽出に適用した例を示し，通常の線形予測分析では正確な分析が比較的困難であった女声のような高ピッチ音声の分析に対して本手法が特に有効であることを示す．

4.2 標本選択線形予測分析による高ピッチ音声のホルマント周波数推定

4.2.1 標本選択線形予測分析

第2章で述べたように，通常の線形予測分析では予測係数 $\{\alpha_k, k=1,2,3,\cdots,p\}$ の推定値 $\hat{\alpha}_k$ を式(2．8)の正規方程式の解として求めている．ところで，共分散法の場合，式(2．10)により c_{ij} を算出していることより，式(2．8)の正規方程式は，

$$\boldsymbol{S}^T\boldsymbol{S}\hat{\alpha} = \boldsymbol{S}^T\boldsymbol{s} \tag{4.1}$$

但し，

$$\boldsymbol{S} = \begin{bmatrix} s_p & s_{p-1} & s_{p-2} & \cdots & s_1 \\ s_{p+1} & s_p & s_{p-1} & \cdots & s_2 \\ s_{p+2} & s_{p+1} & s_p & \cdots & s_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{N-1} & s_{N-2} & s_{N-3} & \cdots & s_{N-p} \end{bmatrix}$$

$$\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \cdots, \hat{\alpha}_p)^T$$

$$\boldsymbol{s} = (s_{p+1}, s_{p+2}, s_{p+3}, \cdots, s_N)^T$$

と書ける．一方，式(2．6)を $n = p+1, p+2, \cdots, N$ について一括して行列の形式で表現すると，

$$\boldsymbol{S}\alpha + u = s \tag{4.2}$$

但し，

$$\alpha = (\alpha_1, \alpha_2, \alpha_3, \cdots, \alpha_p)^T$$

$$u = (u_{p+1}, u_{p+2}, u_{p+3}, \cdots, u_N)^T$$

と記述できる．式(4．2)の両辺に左から $\boldsymbol{S}^T$ を掛けると，

$$\boldsymbol{S}^T\boldsymbol{S}\alpha + \boldsymbol{S}^T\boldsymbol{u} = \boldsymbol{S}^T\boldsymbol{s} \tag{4.3}$$

となるので，式(4．1)，(4．3)より，

$$\boldsymbol{S}^T\boldsymbol{S}(\hat{\alpha} - \alpha) = \boldsymbol{S}^T\boldsymbol{u} \tag{4.4}$$

となる．すなわち，

$$\boldsymbol{S}^T\boldsymbol{u} = \boldsymbol{0} \tag{4.5}$$

となれば $\hat{\alpha} = \alpha$ となり正確な予測係数が得られることになる．有声音の場合，観測された音声波 s に適当な前処理を施すことによって，励振源 u は概周期的なパルス列とみなすことができるので，成人男性の有声音のようにピッチ周期が比較的長ければ式(4．5)が近似的に成立するため，通常の線形予測分析により予測係数を精度よく推定できるが，成人女性あるいは子供の有声音のようにピッチ周期が短いいわゆる高ピッチ音声の場合には式(4．5)が近似的にも成立しなくなり，推定精度が悪くなる危険性がある．

ところで，有声音の場合，分析窓長を1ピッチ周期以下とし，いわゆる声門閉止区間のみを分析対象とすれば $u = 0$ となるため，予測係数を精度よく推定できると言えるが，自然音声の声門閉止区間を正確に推定するのは一般に困難であり，特に女声のような高ピッチ音声の場合には，それはより困難となるだけでなく，声門閉止区間が推定できたとしてもその区間長が極端に短くなるため，予測係数の個数と予測式（式(4．6)参照）の個数が同程度にしかならず分析フレーム内にわずかでも励振があることによる影響ならびに雑音の影響を過敏に受け分析結果のフレーム間連続性に問題が生じやすいと言える．

ところで，式(4．1)より，通常の線形予測分析は予測係数の推定値 $\hat{\alpha}_k$ を，

$$\underset{(N-p)\times p}{\begin{bmatrix} s_p & s_{p-1} & s_{p-2} & \cdots & s_1 \\ s_{p+1} & s_p & s_{p-1} & \cdots & s_2 \\ s_{p+2} & s_{p+1} & s_p & \cdots & s_3 \\ \vdots & \vdots & \vdots & & \vdots \\ s_{2p-1} & s_{2p-2} & s_{2p-3} & \cdots & s_p \\ \vdots & \vdots & \vdots & & \vdots \\ s_{N-1} & s_{N-2} & s_{N-3} & \cdots & s_{N-p} \end{bmatrix}} \underset{p\times 1}{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_p \end{bmatrix}} \approx \underset{(N-p)\times 1}{\begin{bmatrix} s_{p+1} \\ s_{p+2} \\ s_{p+3} \\ \vdots \\ s_{2p} \\ \vdots \\ s_N \end{bmatrix}} \tag{4.6}$$

なる優決定性（Over−determined）の連立一次方程式の最小2乗解として求めていると言える．すなわち，通常の線形予測分析は励振源の如何にかかわらず分析フレーム内の全音声標本（但し，s_1〜s_p は除く）に予測関係をあてはめて，2乗誤差最小の規準の下に予測係数を推定していることになる．しかしながら，正確な予測係数が推定できるかどうかが式(4．5)の成立度合に依存することを考慮し，分析フレーム内の全音声標本に予測関係を一律には適用せずに，励振源が実効的に零とみなせる時点の音声標本のみに予測関係を適用すれば（すなわち，式(4．6)に示されている $N-p$ 個の各予測式の中から予測誤差の小さい予測式だけを用いれば），励振源の影響を軽減したより正確な分析が可能となると考えられる．これが標本選択線形予測分析の基本的な考え方である．なお，励振音源が実効的に零とみなせる時点の音声標本値の選定方法として，ここでは通常の線形予測分析による残差信号の絶対値の小さいものを選定するものとする．

4.2.2　2段標本選択線形予測分析

以上の標本選択の基本的な手法に関してはすでに溝口らにより提案されているが[65]，この手法を女声のような高ピッチ音声の分析にそのまま適用すると必ずしも満足な結果が得られない場合が生じた．その原因を検討した結果，以下に述べる2種の現象に対する対策が必要であることが判明した．

現象1：高ピッチ音声の場合には，前述の理由により通常の線形予測分析の分析精度が極端に悪くなる場合があり，その残差に基づいて音声標本を選択していたのでは被予測標本として妥当な音声標本の選択ができないことが生じる．

対策：標本選択処理を2段階に拡張する．すなわち，通常の線形予測分析による残差に基づいて一度標本選択線形予測分析を行って得られる残差に基づいて再度被予測標本としてより妥当な音声標本を選択する．

現象2：従来の標本選択線形予測分析では処理時間節減の目的から，線形予測分析に一般逆行列を導入し，Givens 変換に基づく逐次計算法を用いることによって各標本値を処理するごとにその予測残差の大きさを評価し，その標本値を被予測標本として使用するか否かを決定していた．しかし，高ピッチ音声の場合には，この方法で予測残差の絶対値が閾値以上となる音声標本を被予測標本から除いただけでは不十分である．

対策：標本の選択処理をフレーム単位で行うことにより，予測残差の大局的な特徴も考慮して標本の選択を行う．すなわち，有声音の場合，一般に，声門閉止時点において予測残差がパルス状に大きくなるので，大きな予測残差を示す点の手前は声門開口区間と言える．この事実を利用し，予測残差を極性を含めて（絶対値最大が正側に現れるように）正規化し，それが閾値以上となる音声標本ならびにその手前 N_0 個の音声標本を被予測標本から除く．すなわち予備分析における予測残差がいくら小さくとも声門開口部に対応する音声標本は強制的に被予測標本から除く．なお，従来の標本選択線形予測分析においても，閾値設定が適当でなかった場合を考えて，予測残差の絶対値が閾値以上となる音声標本の前後数点を経験的に除いていたが（文献(65)p.61の脚注），ここでは閾値設定上の観点からではなく声門開口区間の音声標本を被予測標本から積極的に除く意味で導入したものである．

以上のように，本手法は，従来の標本選択線形予測分析を2段階に拡張すると共に，予測残差の大局的な特徴を考慮することにより，被予測標本（式(4．6)の右辺）としてより妥当な標本の選択を行い，高ピッチ音声に対しても励振源の影響を軽減したより正確な分析を可能とするものである．従って，本手法は基本的には自由振動区間を対象とした処理法と言えるが，分析窓長は一般に用いられている 20 ～ 30ms でよく，かつ声門閉止区間を推定する必要がないため，従来の自由振動区間内分析に生じる諸々の難点のない手法であると言える．なお，本手法も従来の標本選択線形予測分析と同様，予測残差に基づいて被予測標本として妥当な音声標本を選択しているのであって，選択されなかった標本は分析処理に一切使用されないと言うのではない．例

えば，音声標本値 s_{p+1} の残差が閾値以上である場合，s_{p+1} は被予測標本値としては使用されないが（式(4．6)の第一行目を除去），予測式の要素（式(4．6)の左辺の要素，今の場合，第2行目から第 $p+1$ 行目の左辺の要素）としては使用され得る．

本方法の手順をまとめると以下の様になる．

【1】標本選択のパラメータである閾値 θ と強制的除去標本点数 N_0 の設定（図4．1参照）．

【2】分析窓長 $T_a = 20 \sim 30$ms の通常の線形予測分析を行ない予測係数を求める．

【3】得られた予測係数に基づき残差信号 e_n を計算する．但し，分析フレーム内での残差信号の絶対値の最大値を与える値（符号を含む）で正規化する．すなわち，残差信号は本質的に双極性であるため，この正規化により予測信号の全体的な極性を正極性に正規化することになる．

【4】残差が閾値 θ 以下となる音声標本 $\{s_{n_1}, s_{n_2}, s_{n_3}, \cdots, s_{n_M},\}$ を選定．但し，2回目の標本選択処理では残差が閾値 θ 以上となる音声標本の手前 N_0 個を除く（図4．1参照）．

【5】選定された音声標本を被予測標本とする予測式を連立させ，その最小2乗解（式(4．7)の解）として，予測係数を求める．

$$\boldsymbol{S}_M^T \boldsymbol{S}_M \hat{\alpha} = \boldsymbol{S}_M^T \boldsymbol{s}_M \tag{4.7}$$

但し，

$$\boldsymbol{S}_M = \begin{bmatrix} s_{n_1-1} & s_{n_1-2} & s_{n_1-3} & \cdots & s_{n_1-p} \\ s_{n_2-1} & s_{n_2-2} & s_{n_2-3} & \cdots & s_{n_2-p} \\ s_{n_3-1} & s_{n_3-2} & s_{n_3-3} & \cdots & s_{n_3-p} \\ \vdots & \vdots & \vdots & & \vdots \\ s_{n_M-1} & s_{n_M-2} & s_{n_M-3} & \cdots & s_{n_M-p} \end{bmatrix}$$

$$\boldsymbol{s}_M = \left(s_{n_1}, s_{n_2}, s_{n_3}, \cdots, s_{n_M}\right)^T$$

【6】ステップ【3】に戻り【3】～【5】の処理を再度行う．

4.3 合成音による分析精度の検討

標本化周波数 10kHz，励振源：ピッチ周期 3.8ms の Rosenberg 波[40]（図4．1（a）参照），ホルマント周波数：表4．1，放射特性：6dB/oct として作成した合成5母音を用いて，本方法によるホルマント周波数推定精度の改善度を明らかにする．

合成母音/o/における各部の波形ならびに標本選択の例（図4．1（c）及び（d）の下段｜印）を図4．1に示す．但し，前処理として一階差分，分析次数 $p=12$，分析窓長 $T_a=25.6$ms とした．

図4．1（c）は通常の線形予測分析による残差波形，図4．1（d）は図4．1（c）の残差波形に基づいて前章の手順により一度標本選択線形予測分析（閾値 $\theta=0.2$）を行った場合の残差波形である．両者の残差波形を比較すると，標本選択線形予測分析による残差波形の方が通常の線形予測分析による残差波形よりパルス列状に近くなっていると言える．本方法はこの特徴を積極的に利用したものである．すなわち，従来の標本選択処理による標本選択の例（図4．1（c）の下段｜印，但し，閾値 $\theta=0.2$ とし，残差の絶対値が θ 以上となる点及びその手前3点，後1点を除去）から明らかなように，従来の標本選択処理でも各声門閉止点及びその手前の声門開口部付近に対応する音声標本が被予測標本から除かれてはいるが，正確な声道伝達特性推定に必要な声門閉止部付近の音声標本も被予測標本から除かれる（今の場合，t=2.8ms 及び 6.6ms 付近）ことがあると言える．それに対して本方法では，一度標本選択線形予測分析した後のよりパルス列状に近くなっている図4．1（d）の残差波形に基づいて標本を選択しているため（閾値 $\theta=0.2$，予測残差が閾値 θ 以上となる手前の除去標本点数 $N_0=10$ と

表4．1　合成音のホルマント周波数

(Hz)

	F_1	F_2	F_3	F_4	F_5
/a/	813	1313	2688	3438	4438
/i/	375	2188	2938	3438	4438
/u/	375	1063	2188	3438	4438
/e/	438	1813	2688	3438	4438
/o/	438	1063	2688	3438	4438

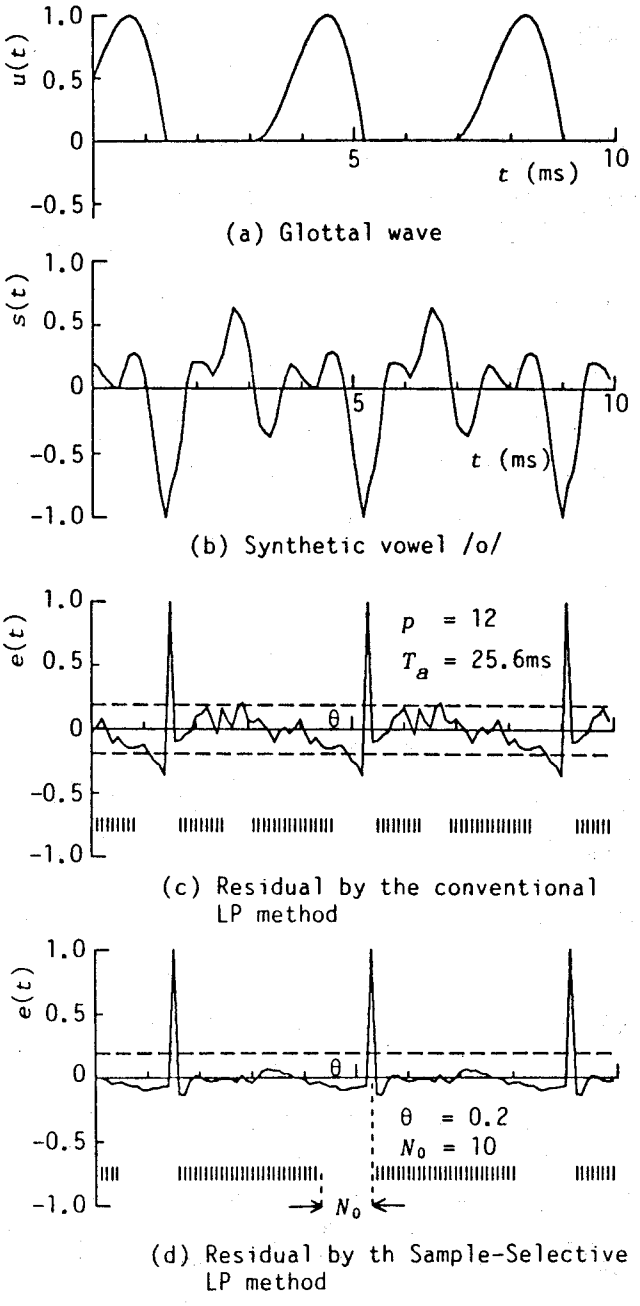


図4．1　合成母音/o/における残差波形及び標本点選択の例
（各残差波形の下段｜印は選択された標本点を示す）

した場合，図4．1（d）の下段｜印となる），各声門閉止点及びその手前の声門開口部付近に対応する音声標本のみが被予測標本から除かれ，より適切な標本が選択されていると言える．なお今の場合，閾値 $\theta \geq 0.3$ とすれば，従来の標本選択処理でも本方法とほぼ同等の標本選択が行えると言えるが，自然音声の場合，通常の線形予測分析による残差が図4．1（c）程度のパルス列状にはならず，また母音定常部でも分析フレーム内の各ピッチごとの残差のピークレベルにかなりの差が生じることがあるため，従来の標本選択処理では適切な閾値θを設定することが困難となる．この点に関しては，4．4で具体的に述べる．

なお，ここでは標本の選択処理を2段階で留めているが，標本選択処理により予測残差が大きくなる音声標本すなわち線形予測モデルに適合しない音声標本が被予測標本から除かれていくので，原理的にはもっと多段階行ってもなんら問題はないと言える．しかしながら，3段階以上行っても合成音ならびに自然音声とも顕著な改善がみられなかったので，標本の選択処理の簡素化を考慮して，2段階に固定した．

4.3.1 閾値 θ の検討

式（4．8）で定義される合成5母音の第1～第3ホルマント周波数推定誤差の平均値 E の閾値θ依存性を図4．2に示す．但し，前処理として一階差分後，分析次数 $p=12$，分析窓長 T_a =25.6ms とし，フレームシフト 0.2ms で一周期に渡って分析した場合の平均値である．そして○印：本方法で N_0=10 とした場合の結果，△印：従来の標本選択線形予測分析の結果である．また，通常の線形予測分析の誤差を図中－－－にて示す．

$$E = \frac{1}{15}\sum_{j=1}^{5}\sum_{i=1}^{3} |\hat{F}_{ij} - F_{ij}| / F_{ij} \tag{4.8}$$

但し，　F_{ij}：第 j 母音の第 i ホルマント周波数
　　$\hat{F}_{ij}$：第 j 母音の第 i ホルマント周波数推定値

図4．2より，ホルマント周波数推定誤差が通常の線形予測分析ではピッチの影響により 5.3% と大きかったものが，従来の標本選択線形予測分析により 2.4% 程度に改善し，さらに本方法により 0.9% と大幅に改善されていることが分る．そして，従来の標

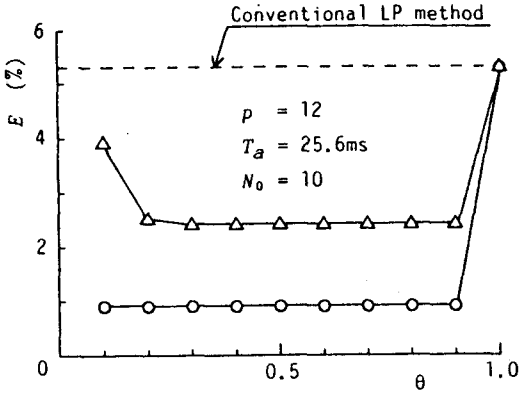


図4．2　ホルマント周波数推定誤差の閾値θ依存性
○：2段標本選択線形予測法
△：従来の標本選択線形予測法

本選択線形予測分析の誤差は $0.3 \leqq \theta < 1.0$ においては閾値θに依存せず一定であるが，$\theta < 0.3$ において閾値θにより変動しているのに対し，本方法の誤差は $0.1 \leqq \theta < 1.0$ において閾値θに依存しないことが分る．ここで用いた合成音では，$\theta \geqq 0.3$ とすれば，従来の標本選択線形予測分析においても残差信号が閾値θ以上となるのは各ピッチごとの実効的な励振点のみとなるため（図4．1（c）参照），誤差は $0.3 \leqq \theta < 1.0$ において閾値θに依存せず一定となる．したがって，$\theta < 0.3$ においてはじめて標本選択処理を2回行う効果が得られていると言える．なお，$\theta \geqq 0.3$ において本方法の誤差が従来の標本選択線形予測分析の誤差より改善しているのは本方法における除去標本点数 N_0 の効果によるものである．すなわち，各ピッチごとの実効的な励振点付近のみを除去するよりも，残差レベルのいかんに拘らず実効的な励振点とその手前10点程度を除去すれば，ホルマント周波数推定誤差が大幅に改善すると言える．

4.3.2　除去標本点数 N_0 の効果

本方法におけるホルマント周波数推定誤差 E の N_0 依存性を図4．3に示す．但し，閾値 θ=0.5 とし，○印：$p=10$，△印：$p=12$，□印：$p=14$ とした場合の結果である．

図4．3より，$N_0>0$ すなわち残差が閾値θ以上となる音声標本値の手前 N_0 個を残差レベルのいかんに拘らず被予測値から除外することにより，ホルマント周波数推定誤差が減少し，$0<N_0\leqq 8$ においては推定誤差の改善度合が分析次数により異なるが，$N_0\geqq 22$ で分析次数に拘らず推定誤差が零となることが分る．

有声音の場合，音声波に適当な前処理を施せば，励振源は概周期的なパルス列とみなすことができると言えるが，これは近似的に言えることであり実際には完全なパルス列とはならない．本シミュレーションでは励振源として図4．1（a）に示すRosenberg波を用いた．したがって，音声波を二階差分すれば，励振源は各声門閉止時点でほぼパルス状とはなるが（今の場合，放射特性として一階差分を用い，分析の前処理として一階差分を行っているので，実質的に励振源を二階差分したことになる），各声門開口区間では零とはならない．ところで，前章で明らかにしたようにホルマント周波数推定精度は式(4．5)の成立度合に依存している．したがって，$N_0>0$ とすることにより各声門開口区間に対応する音声標本が残差レベルの大きさのいかんに拘らず被予測標本から除外されるためホルマント周波数推定精度が改善されることになる．本

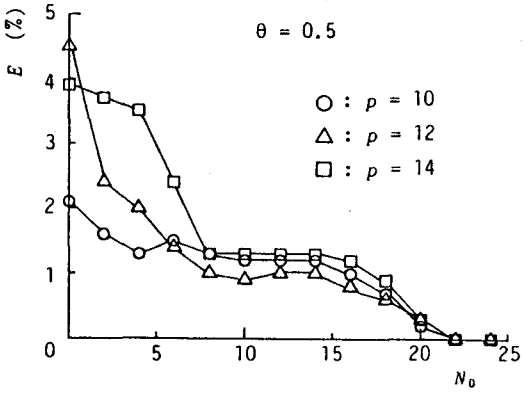


図4．3　ホルマント周波数推定誤差の N_0 依存性

シミュレーションに用いた励振波形の声門開口区間は 2.2ms であるので，$N_0 \geqq 22$ とすれば，各声門開口区間に対応する音声標本は被予測標本から完全に除外されることになり，いわゆる声門閉止区間内分析（式(4．5)が成立）となるため，ホルマント周波数推定誤差は零となる．したがって，合成音の分析に際しては N_0 をできるだけ大きくして，声門閉止区間内分析に近づければ良いと言えるが，自然音声，特に高ピッチ音声の場合，N_0 をあまり大きくすると予測式の個数が少なくなり分析結果の安定性に問題が生じるので N_0 をあまり大きくとることはできない．このことを考慮すれば，図4．3の結果より N_0=10 程度が適当と思われる．

合成5母音について従来の方法と本方法（$\theta = 0.5$，$N_0 = 10$）のホルマント周波数推定誤差の比較を表4．2に示す．

表4．2より，母音/a/において本方法の誤差が従来の標本選択線形予測分析の誤差より若干悪くなっているが，他の母音に関してはいずれも誤差がさらに改善し，特に高ピッチにおいてピッチ周波数と第1ホルマント周波数が接近し，ピッチの影響を大きく受けると思われる母音/i/及び/u/の改善が著しいと言える．

表4．2　ホルマント周波数推定誤差の比較

(%)

母　　音	/a/	/i/	/u/	/e/	/o/
通常の線形予測法	1.5	6.7	5.2	6.8	6.1
従来の標本選択線形予測法	0.3	3.6	3.5	2.1	2.3
本　方　法	0.6	0.4	0.4	1.6	1.7

4.3.3　分析次数の検討

ホルマント周波数推定誤差Eの分析次数 p 依存性を図4．4に示す．但し，○印：本方法（$\theta = 0.5$，$N_0 = 10$），△印：従来の標本選択線形予測分析（$\theta = 0.5$），×印：通常の線形予測分析（共分散法）で，分析次数 $p = 12$，分析窓長 $T_a = 25.6$ms とした場合の結果である．

図4．4より，従来の標本選択線形予測分析の誤差は分析次数と共に若干単調に増大していたのが，本方法により分析次数依存性が改善し，かつ，ホルマント周波数推定

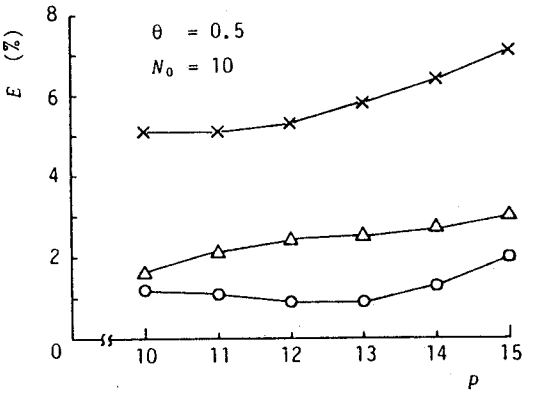


図4．4　ホルマント周波数推定誤差の分析次数 p 依存性
○：2段標本選択線形予測法
△：従来の標本選択線形予測法
×：通常の線形予測法

誤差はいずれの分析次数においても通常の線形予測分析による誤差よりも大幅に小さいと言える．

4.3.4　ピッチ周期に関する頑健性の検討

ホルマント周波数推定誤差 E のピッチ周期 T_s 依存性を図4．5に示す．但し，合声音は励振源のピッチ周期のみを 3.0ms から 5.0ms まで変化させたものであり，他の合成条件（励振源の声門開口比等）および分析条件は図4．4と同じである．なお各印の意味も図4．4と同じである．

図4．5より，ピッチ周期が 4.2ms 以上では従来の標本選択線形予測分析による誤差も本方法による誤差もほぼ同じであるが，従来の標本選択線形予測分析はピッチ周期が 4.2ms 以下になるとホルマント周波数推定精度が徐々に悪くなっていたのが，本方法によりピッチ周期が 4.0ms 以下の誤差が大幅に改善されていると言える．なお，通常の線形予測分析による誤差はピッチ周期が 4.6ms 以下になると急激に大きくなっていたのが，ピッチ周期が 4.0ms 以下になると誤差は平均的には減少する傾向がみられる．これはピッチ周期とホルマント周波数の相対位置関係により，誤差が小さくなったものと考えられる．ホルマント周波数がピッチ周波数の高調波間の 1/4 あるいは 3/4

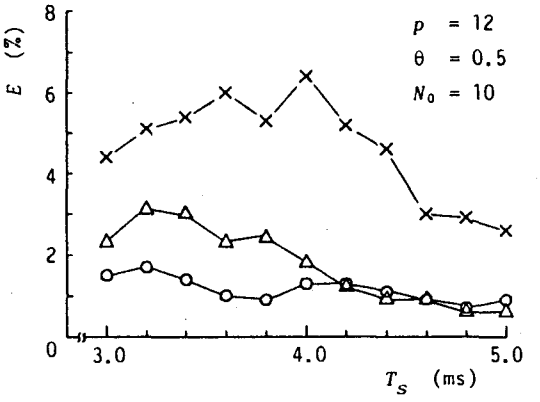


図4．5　ホルマント周波数推定誤差のピッチ周期 T_s 依存性
○：2段標本選択線形予測法
△：従来の標本選択線形予測法
×：通常の線形予測法

付近に位置する場合，通常の線形予測分析では正確なホルマント周波数推定が一般に困難となるが[54]，ここで用いた合声音の各ホルマント周波数はピッチ周期が 4.0ms 程度の時にその状態となるため，それよりもピッチ周期が短い場合にはかえって推定誤差が小さくなるという傾向になっている．

4.4　自然音声への適用結果

標本選択処理を成人女性が発声した単音節/bo/の母音部の/o/（ピッチ周期:約 3.2ms）に適用した場合の残差波形及び標本点選択の例を図4．6 に示す．但し，標本化周波数 10kHz，前処理として一階差分後，分析次数 $p = 12$，分析窓長 $T_a = 25.6$ms，$\theta = 0.5$，$N_0 = 10$ とした場合の例である．

図4．6（b）は通常の線形予測分析による残差波形，図4．6（c）は図4．6（b）の残差波形に基づいて4．2．2の手順により一度標本選択線形予測分析（閾値 $\theta = 0.5$）を行った場合の残差波形である．なお，従来の標本選択法及び本方法により被予測標本点として選択された標本時点をそれぞれ図4．6（b）及び（c）の下段｜印にて示す．図4．6（b）の通常の線形予測分析による残差波形より明らかなように，自然音声の

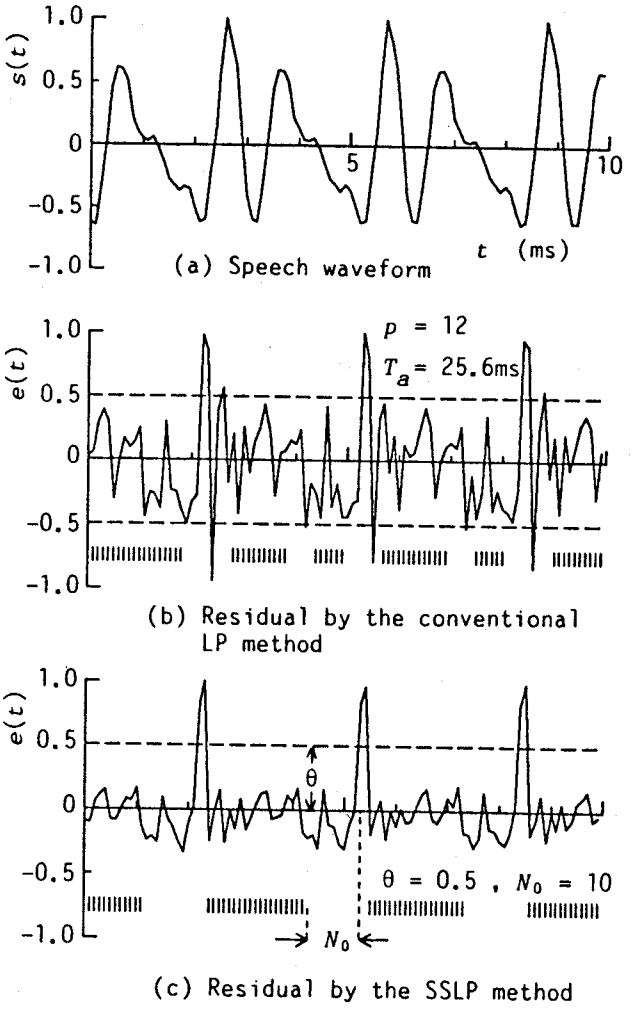
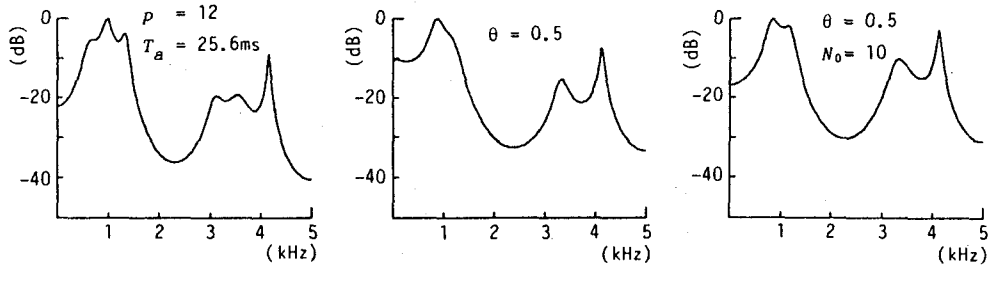


図4．6　自然音声（女声：母音/o/）における残差波形および標本点選択の例
（各残差波形の下段｜印は選択された標本点を示す）

場合，通常の線形予測分析による残差は合成音のようなパルス列状（図4．1（c）参照）とはならないため，従来の標本選択法において閾値$\theta = 0.5$としても適切な標本の選択が行えていない．それに対して，4．2．2の手順により一度標本選択線形予測分析（閾値$\theta = 0.5$）を行った場合の残差（図4．6（c））は通常の線形予測分析による残差（図4．6（b））よりもよりパルス列状となっており，この残差に基づいて再度標本選択処理（$\theta = 0.5$，$N_0 = 10$）を行えばより適切な標本点選択ができることが分る．今の場合，従来の標本選択法でも閾値$\theta \geq 0.6$とすれば良いと言えるが，自然音声では母音定常部でも分析フレーム内の各ピッチごとの残差のピークレベルにかなりの差が生じることがあるため，θをあまり大きく設定できない場合がある．すなわち，従来の標本選択法では実際に女声のような高ピッチ音声を分析する場合，閾値θを入力音声に応じていかに適応的に設定するかを解決する必要がある．これに対して本手法では適切な閾値θの許容範囲が広いため（図4．6の場合，$0.2 \leq \theta < 1.0$），この閾値設定問題は回避できたと言える．

図4．6の場合のスペクトル包絡の比較を図4．7に示す．但し，（a）：通常の線形予測分析によるスペクトル包絡，（b）：従来の標本選択線形予測分析（$\theta = 0.5$）によるスペクトル包絡，（c）：本方法（$\theta = 0.5, N_0 = 10$）によるスペクトル包絡である．

図4．7より，次のことが言える．通常の線形予測分析では1kHz付近及び3.2kHz付近に近接する3個及び2個の明確な極が存在しいずれがホルマントか判断しがたい．従来の標本選択線形予測分析により3.2kHz付近のスペクトル包絡が改善され第3ホ



(a) Conventional LP method　(b) Prototype SSLP method　(c) Two-stage SSLP method

図4．7　スペクトル包絡の比較（女声：母音/o/）

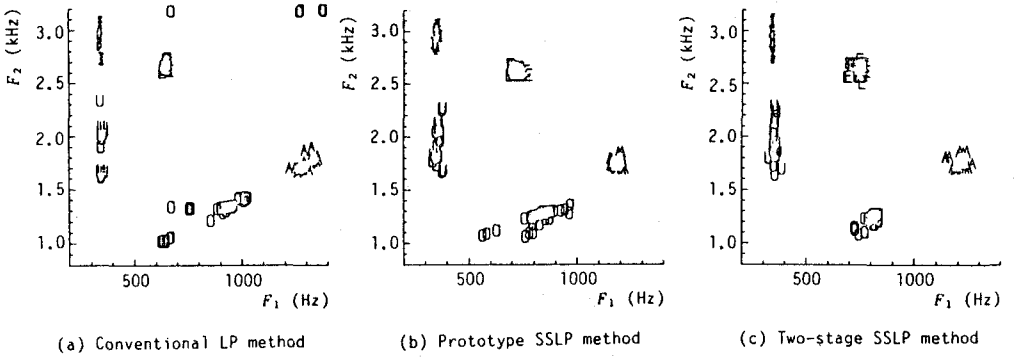


図4．8　ホルマント空間における女声の5母音分布

ルマントが明確となっているが，第2ホルマントが不明確である．本方法ではこれらの点がすべて改善されており第1～第3ホルマントが明確となっている．

F_1-F_2 平面上の5母音の分布の比較を図4．8に示す．但し，音声資料は成人女性1名が発声した単音節（70種）の母音定常部各3フレーム，分析条件は図4．7の場合と同じであり，得られた極のうちバンド幅の小さいものをホルマントとみなした．

図4．8より，次のことが言える．通常の線形予測分析では母音/o/の分布のバラツキが大きく，かつ単なるバンド幅の情報のみではホルマントを誤推定する場合があり（(F_1,F_2)=(600Hz,3.2kHz) 及び (1300Hz,3.2kHz) 付近の計6フレーム），また母音/u/の分布が2クラスに分かれている．従来の標本選択線形予測分析ではバンド幅の情報のみでもホルマントを誤推定することがなく，また母音/u/の分布が改善しているが，母音/o/の分布のバラツキがあまり改善されていない．本方法ではこれらの点がいずれも改善されており，通常の線形予測分析と比較し特に母音/o/の分布の改善が著しいと言える．

ホルマント空間上における分布の良さを評価するために，F_1-F_2 平面上における5母音の分布の類内分散と類間分散に基づいた分離度 D を式(4．9)で定義し，その閾値 θ 依存性を図4．9に示す．

$$D = \sqrt{\frac{\sum_{k=1}^{5}(m_k - m)^T(m_k - m)}{\frac{1}{N}\sum_{k=1}^{5}\sum_{x \in x_k}(x - m_k)^T(x - m_k)}} \tag{4.9}$$

但し，

$x = (F_1, F_2)^T$

$m_k = \frac{1}{N}\sum_{x \in x_k} x$

$m = \frac{1}{5}\sum_{k=1}^{5} m_k$

N:資料数／クラス（今の場合，$N = 42$）

図4．9より，従来の標本選択線形予測分析においても，閾値θが$0.5 \leqq \theta \leqq 0.6$の範囲であれば分離度$D$は大きくなるが，最適な閾値$\theta$の範囲が狭いと言える．これに対して，本方法の分離度は$0.4 \leqq \theta \leqq 0.7$において閾値$\theta$にほとんど依存せず大きな分離度が得られており，本方法の有効性が示されていると言える．

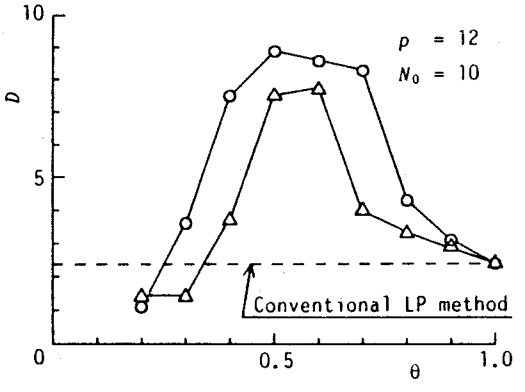


図4．9　ホルマント空間における分離度Dの閾値θ依存性
○：2段標本選択線形予測法
△：従来の標本選択線形予測法

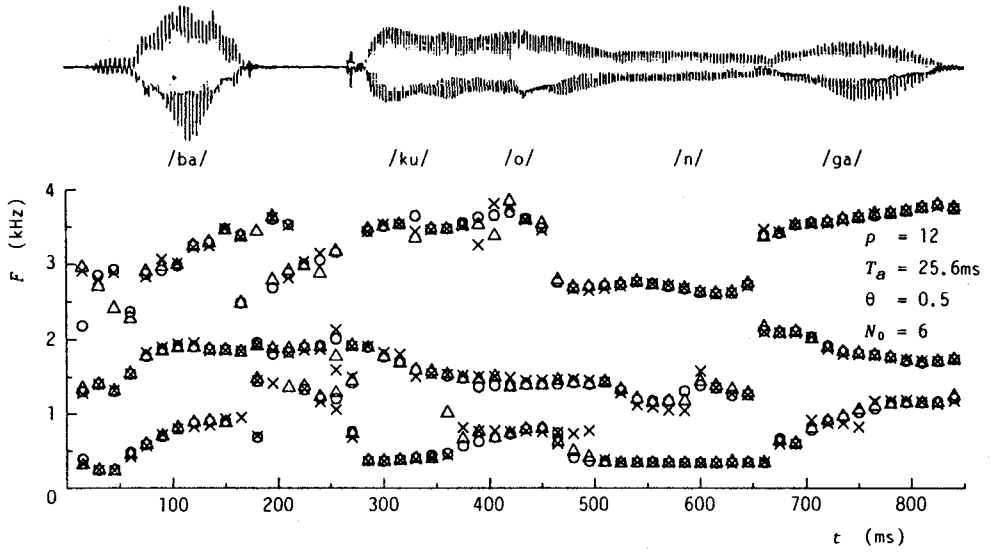


図4．10　連続音声（女声：/bakuonga/）のホルマント周波数抽出例
○：2段標本選択線形予測法
△：従来の標本選択線形予測法
×：通常の線形予測法

成人女性が発声した連続音声「爆音が」のホルマント周波数抽出に適用した例を図4．10に示す．但し，前処理として一階差分後，分析次数 $p = 12$，分析窓長 $T_a = 25.6$ms，フレーム間隔 15ms とし，○印：本方法（$\theta = 0.5$，$N_0 = 6$[1]），△印：従来の標本選択線形予測分析（$\theta = 0.5$），×印：通常の線形予測分析による結果である．なお，ホルマント周波数は各フレームごとに 200Hz ～ 4000Hz 内に得られた極のうちバンド幅の小さいものを第1～第3ホルマントとして抽出した．

図4．10より，次のことが言える．通常の線形予測分析ではホルマント周波数の時間的変化に不自然な不連続が生じている（$t = 330$ms 付近の F_2，$t = 380$ms 付近の F_1 及び F_3，$t = 500$ms 付近の F_1，$t = 600$ms 付近の F_2，$t = 750$ms 付近の F_1）．また，$t = 465$ms の所では 700Hz 付近に近接する明確な極が存在するため第3ホルマントが抽出できていない．従来の標本選択線形予測分析により，これらの不連続性がかなり改善されてはいるが，$t = 380$ms 付近の第3ホルマントの時間的変化がまだ不連続で

[1]本連続音声はピッチ周期が 2.6 ～ 3.5ms に渡って変化しているため，ここではこの値を用いた．

あり，また $t = 360$ms の所で第1ホルマントに誤抽出が生じている．これに対して，本方法ではこれらの不連続性がさらに改善されており，本方法の有効性が示されていると言える．

4.5 結　　言

残差情報を参照することによって線形予測モデルに適合する音声標本点を選択する標本選択線形予測分析において，標本の選択処理を予測残差の大局的な特徴を考慮して行い，かつこの処理を2段階行う2段標本選択線形予測分析の有効性を検討した．その結果，本方法は従来の標本選択線形予測分析よりも被予測標本としてより妥当な標本の選択が行えることが明らかとなった．そして，本方法は通常の線形予測分析ではピッチの影響により，正確な分析がしばしば困難であった高ピッチ音声の分析に特に有効であることが，合成音によるホルマント周波数推定精度の改善，自然音声のスペクトル包絡の改善と抽出したホルマント周波数分布の改善ならびに連続音声のホルマント周波数抽出の改善により明らかとなった．

なお，本方法を自然音声に適用するにあたり，本方法のパラメータである除去標本点数 N_0 を固定としたが，4．2．2で述べたように N_0 は声門開口区間の音声標本をできるだけ被予測標本から除くために導入したパラメータであるので N_0 の最適値は声門開口区間すなわちピッチ周期に依存する量であると言える．したがって，特に連続音声に適用する場合には，N_0 はピッチ周期に応じて適応的に変化させることが望ましいと言えるが，この点に関しては今後の課題である．

第5章　音声の過渡部の分析

5.1　緒　　言

近年，ホルマント周波数の時間的変化に注目した知識工学的手法に基づく連続音声認識システム[50]が開発され，かなり高い認識率が得られている．これらのシステムでは，ホルマント周波数の追尾とその記述法が音韻識別のキーポイントとなっているように，ホルマント周波数の時間的変化を正確に追尾することは音声分析の重要な課題の一つである．本章では，音声の過渡部のホルマント周波数を正確に推定する分析手法として，分析の時間窓の長さを有声音の1ピッチ周期未満に短縮した1ピッチ周期内周波数分析，ならびに分析窓の任意の点を固定して窓長を漸減させた一連の分析の結果から，窓長が零になる場合の値を外挿する窓長漸減型線形予測分析について述べる．

音声の過渡部のホルマント周波数を正確に推定するためには基本的には分析窓長を十分短くすれば良いとの観点から，まず5．2で，音声の周波数分析における分析窓の位置および窓長が周波数スペクトルに及ぼす影響を詳細に検討する．その結果，分析窓長を1ピッチ周期未満とし，声道が声帯波により実効的に励振された時点を包含しないように分析窓の位置を設定すれば，声道伝達特性を保存した平滑な周波数スペクトル包絡を持つ周波数スペクトルが得られ，そのピーク周波数（周波数スペクトル上で極大となる周波数）から精度よくホルマント周波数を推定できることを示す．そして，この周波数分析法を合成および自然有声破裂音のホルマント周波数追尾に適用し，その有効性を示す．

一方，通常の線形予測分析は第2章で述べたように，分析窓内での定常性が仮定されているため，ホルマント周波数の急激な時間的変化を正確に追尾する場合には，分析窓長を数ms程度に短くする必要がある．しかしながら，有声音の場合，分析窓長を1ピッチ周期程度以下に短くすると，通常の線形予測分析は分析窓と励振点との相対位置の影響を大きく受けるといった問題が生じる．5．3では，通常の線形予測分析による極周波数推定値の分析窓長依存性の解析結果に基づき，分析窓の任意の点を固定

して窓長を漸減させた一連の分析の結果から，窓長が零になる場合の値を外挿することにより，分析窓長を短かくすることによる悪影響を受けることなく，音声の過渡部の任意の時点のホルマント周波数を精度よく推定できる窓長漸減型線形予測分析について述べる．そして，本手法を合成および自然有声破裂音のホルマント周波数追尾に適用することにより，その有効性を示す．

5.2 1ピッチ周期内周波数分析によるホルマント周波数推定[66]

ホルマント周波数の時間的変化の追尾を目的として音声波の周波数分析を行なう場合，一般には，高速フーリエ変換（ＦＦＴ）[20]を利用し，その分析窓の長さを 数十 ms 程度とし，分析窓の位置を 10ms 程度に順次推移させた短時間周波数分析を行なうのが通常である[67]．しかしながら，音韻識別を行なう上で重要となる子音部などにおけるホルマント周波数の急激な時間的変化を追尾するには，調音器官の運動を記述するパラメータの時間的変化率を考慮すれば，有声子音の場合，分析窓の長さを1ピッチ周期以下に短くする必要があると言える．分析窓の長さを1ピッチ周期程度とした周波数分析方法としては Mathews らによりピッチ同期スペクトル分析[8] が提案されているが，この分析法は1ピッチ区間の波形が周期的に繰り返すものとしてフーリエ解析を行なっているため，周波数スペクトルが調波構造となりホルマント周波数を推定する場合に推定精度が問題となる．これに対して，1ピッチ区間の波形を切り出し，短時間周波数スペクトル分析を行なう方法が報告されているが[68],[69]，分析窓の長さを1ピッチ周期程度と短くした場合には，分析窓の位置の影響を受けるため声道部の伝達特性が都合よく現われるようにする必要がある．この問題を解決するためは分析窓の位置および窓長に関する詳細な検討が必要と考えられるが，この種の検討は十分に行なわれていないと言える．

本節では，分析窓の長さを1ピッチ周期未満とした有声音の短時間周波数スペクトルの特性を解明することにより，声道が声帯波により実効的に励振された時点を包含しないように分析窓の位置を設定するならば，声道特性を保存した平滑な周波数スペクトル包絡が得られ，そのピーク周波数（周波数スペクトル上で極大となる周波数）

からホルマント周波数を精度よく推定できることを示す．以下，5．2．1において，短時間周波数スペクトル上のピーク周波数の分析窓の位置および窓長依存性を解析的に考察し，5．2．2で，合成音のシミュレーションにより解析結果の検証を行なう．そして，5．2．3では，急激な過渡部を持つ合成有声破裂音ならびに自然有声破裂音のホルマント周波数追尾に適用し，本手法の有効性を示す．

5.2.1　有声音の短時間周波数スペクトル特性

周知のように音声波の短時間周波数スペクトルは音声波を適当な分析窓により切出した波形をフーリエ変換して得られる．本節では，この分析窓の長さを有声音の1ピッチ周期未満と短くした短時間周波数スペクトル上のピーク周波数が分析窓の位置および窓長によりどのように変化するかを明らかにする．そのため，音声波 $s(t)$ の短時間周波数スペクトル $F(\omega, t_s, T_a)$ を分析窓 $w(t)$ の始点 t_s，窓長 T_a の関数として，

$$F(\omega, t_s, T_a) \triangleq \int_{t_s}^{t_s+T_a} w(t-t_s)s(t)e^{-j\omega t}dt \tag{5.1}$$

但し，　$w(t)=0$,　　$t<0$　or　$t>T_a$

と定義し，その特性を解析する．

(a)　ピーク周波数の分析窓の始点および窓長依存性

音声波 $s(t)$ が有声音である場合，

$$f(t)=\int_0^{\infty} h(\tau)u(t-\tau)d\tau \tag{5.2}$$

但し，　$h(t)$：声道インパルス応答
　　$u(t)$：声帯音源波形

と記述できる（但し，音の放射特性は声道特性の中に含める）．今，簡単のために分析窓 $w(t)$ を

$$w(t)=\begin{cases} 1 & 0 \leqq t \leqq T_a \\ 0 & \text{その他} \end{cases} \tag{5.3}$$

なる方形窓とすれば，式(5．1)～(5．3)より

$$F(\omega, t_s, T_a)=\int_0^{\infty} h(\tau)e^{-j\omega\tau}\int_{t_s-\tau}^{t_s+T_a-\tau} u(t)e^{-j\omega t}dtd\tau \tag{5.4}$$

となる．ここで，声帯音源波形 $u(t)$ を

$$u(t)=\sum_{n=-\infty}^{\infty}\delta(t-nT_0) \tag{5.5}$$

なる周期 T_0 のインパルス列とし，分析窓長 T_a を1ピッチ周期未満（$T_a < T_0$）とした場合，$F(\omega, t_s, T_a)$ は

$$F(\omega, t_s, T_a) = \sum_{n=-\infty}^{n_0} e^{-j\omega nT_0} \int_{t_s-nT_0}^{t_s+T_a-nT_0} h(\tau) e^{-j\omega\tau} d\tau \qquad (5.6)$$

但し，$n_0 = \left[\frac{t_s + T_a}{T_0}\right]$，[ ]:ガウス記号

となる．ここで，分析窓の始点 t_s を

$$t_s = mT_0 + t_{s0} \qquad (5.7)$$

但し，　m：整数，$|t_{s0}| \leq T_0/2$

と記述し，$t_{s0} + T_a < T_0$ とすれば，$F(\omega, t_s, T_a)$ は

$$F(\omega, t_s, T_a) = e^{-j\omega mT_0} \sum_{k=0}^{\infty} \int_{t_{s0}}^{t_{s0}+T_a} h(t + kT_0) e^{-j\omega t} dt \qquad (5.8)$$

となる．すなわち，声帯音源が周期 T_0 のインパルス列とみなせる場合，分析窓長を1ピッチ周期未満（$T_a < T_0$）とし，かつ分析窓の位置を次のインパルス印加時点を包含しない（$t_{s0} + T_a < T_0$）ように設定した短時間周波数スペクトルは声道インパルス応答 $h(t)$ を T_0 ずつ負方向にシフトさせ積分範囲を［　t_{s0}，　$t_{s0} + T_a$　］とした短区間フーリエ変換の和に分析窓の位置による位相回転が加わったものとなる．ところで，$h(t)$ は t が増大すれば大局的にみて減衰するので，式(5.8)の右辺の和の中では $k = 0$ の項が主要な成分となる．そして，$t < 0$ では $h(t) = 0$ であるため $t_{s0} < 0$ に設定すれば，この主要な成分の積分範囲が実効的に減衰するので $t_{s0} \geqq 0$ とするのが適切であるといえる．

今，簡単のために声道伝達特性が単一の極である場合を考える．この場合には，

$$h(t) = \begin{cases} e^{-\alpha t} \sin \omega_0 t & t \geqq 0 \\ 0 & t < 0 \end{cases} \qquad (5.9)$$

但し，　$\alpha > 0$

である．ここで，$t_{s0} \geqq 0$ とし，式(5.9)を式(5.8)に代入すれば，

$$F(\omega, t_s, T_a) = e^{-j\omega m T_0}\left[\frac{e^{-\{\alpha+j(\omega-\omega_0)\}t_{s0}}\left[1-e^{-\{\alpha+j(\omega-\omega_0)\}T_a}\right]}{2j\{\alpha+j(\omega-\omega_0)\}\{1-e^{-(\alpha-j\omega_0)T_0}\}}\right.$$

$$\left.-\frac{e^{-\{\alpha+j(\omega+\omega_0)\}t_{s0}}\left[1-e^{-\{\alpha+j(\omega+\omega_0)\}T_a}\right]}{2j\{\alpha+j(\omega+\omega_0)\}\{1-e^{-(\alpha+j\omega_0)T_0}\}}\right] \tag{5.10}$$

となる．ところで，$\omega = \omega_0$ 付近の ω に対しては $|\omega+\omega_0| \gg |\omega-\omega_0|$ となるので式(5．10)の右辺の第2項を省略することができ，

$$|F(\omega, t_s, T_a)| \approx \frac{1}{2}\left|\frac{e^{-\{\alpha+j(\omega-\omega_0)\}t_{s0}}\left[1-e^{-\{\alpha+j(\omega-\omega_0)\}T_a}\right]}{\{\alpha+j(\omega-\omega_0)\}\{1-e^{-(\alpha-j\omega_0)T_0}\}}\right|$$

$$= \frac{e^{-\alpha t_{s0}}}{2}\sqrt{\frac{1+e^{-2\alpha T_a}-2e^{-\alpha T_a}\cos(\omega-\omega_0)T_a}{\{\alpha^2+(\omega-\omega_0)^2\}\{1+e^{-2\alpha T_0}-2e^{-\alpha T_0}\cos\omega_0 T_0\}}} \tag{5.11}$$

となる．従って，$\partial|F(\omega, t_s, T_a)|/\partial\omega = 0$，すなわち，

$$\begin{aligned} D(\omega) &= \left\{\alpha^2+(\omega-\omega_0)^2\right\}T_a e^{-\alpha T_a}\sin(\omega-\omega_0)T_a \\ &\quad -(\omega-\omega_0)\left\{1+e^{-2\alpha T_a}-2e^{-\alpha T_a}\cos(\omega-\omega_0)T_a\right\} \\ &= 0 \end{aligned} \tag{5.12}$$

を満足する ω において $|F(\omega, t_s, T_a)|$ は極値または変曲点となる．この式は t_s，T_0 のいかんにかかわらず成立する．これから，分析窓内にインパルス印加時点が存在しなければ短時間周波数スペクトル上のピーク周波数の位置は分析窓の始点ならびにインパルス列の周期に依存せず窓長にのみ依存することが分かる．更に，$\omega = \omega_0$ は常に式(5．12)を満たすので，これに対応するピークの位置は窓長にも依存しないことが分かる．一例として，$\alpha = 0.3\,\omega_0/2\pi$ の場合の $D(\omega)$ の例を図5．1に示す．

図5．1において，$D(\omega)$ が零交差する ω において $|F(\omega, t_s, T_a)|$ は極値となる．この極値前後における $\partial|F(\omega, t_s, T_a)|/\partial\omega$ の符号より，$D(\omega)$ が正から負に零交差する ω において $|F(\omega, t_s, T_a)|$ は極大となる．又，$D(\omega)$ が ω/ω_0 軸に接する ω において $|F(\omega, t_s, T_a)|$ は変曲点となる．図5．1より，今の場合，$|F(\omega, t_s, T_a)|$ 上には変曲点は存在しないが，$\omega = \omega_0$ の周波数を含めて多くのピークが存在すること

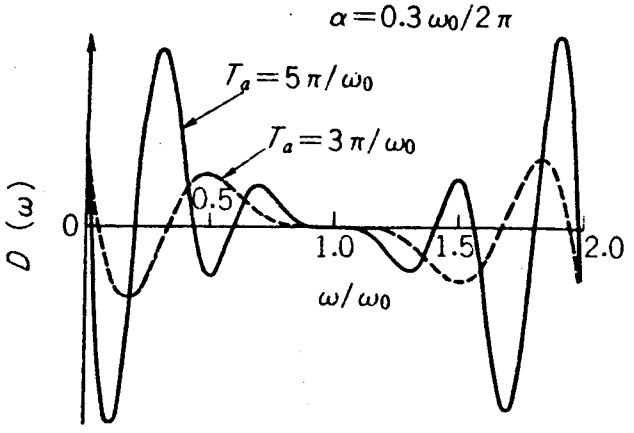


図5．1　$D(\omega)$ の例

が分かる．そして$\omega = \omega_0$ の位置には T_a の値にかかわらずピークが存在し，ほかのピークは T_a の値によりその位置が変動すると言える．

(b)　分析窓長の下限

ホルマント周波数の急激な時間的変化を正確に追尾するには，分析窓長はできるだけ短い方が望ましいが，一方，ホルマント周波数に対応する位置にピーク周波数が存在するためには，周波数分解能を考慮すると窓長をある値以下にすることはできない．この窓長の下限について考慮する．

今，声道伝達特性 $h(t)$ が

$$h(t) = \begin{cases} e^{-\alpha t}\sin\omega_1 t + e^{-\alpha t}\sin\omega_2 t & t \geqq 0 \\ 0 & t < 0 \end{cases} \tag{5.13}$$

但し，　$\alpha > 0$

なる2つの極である場合，$|F(\omega, t_s, T_a)|$ は単一極の場合の式(5．11)を参照し，簡単のために $t_{s0} = 0$，$T_0 = \infty$ とすれば，

$$|F(\omega, t_s, T_a)| \approx \frac{1}{2}\left|\frac{1-e^{-\{\alpha+j(\omega-\omega_1)\}T_a}}{\alpha+j(\omega-\omega_1)} + \frac{1-e^{-\{\alpha+j(\omega-\omega_2)\}T_a}}{\alpha+j(\omega-\omega_2)}\right| \tag{5.14}$$

となる．この周波数スペクトル上に $\omega = \omega_1$ 及び $\omega = \omega_2$ に対応するピークが存在するための窓長の下限は $\omega = (\omega_1+\omega_2)/2$ において周波数スペクトルが極小となる最小の窓長である考えられる．従って，$\omega = (\omega_1+\omega_2)/2$ において $\partial^2|F(\omega, t_s$

$, T_a) |/\partial\omega^2 > 0$，すなわち，

$$-ST_a(T_a+2)e^{-T_a}\Delta\omega^5 + 2\left\{(S^2+1)e^{-2T_a} + C(T_a^2-2)e^{-T_a}+1\right\}\Delta\omega^4$$
$$-8\alpha Se^{-T_a}(2Ce^{-T_a}+T_a^2+2T_a-2)\Delta\omega^3$$
$$+16\alpha^2\left\{(3C^2-1)e^{-2T_a}+C(T_a^2-4)e^{-T_a}+2\right\}\Delta\omega^2$$
$$+16\alpha^3 Se^{-T_a}(4Ce^{-T_a}-T_a^2-2T_a-4)\Delta\omega$$
$$-32\alpha^4\left\{C^2e^{-2T_a}-C(T_a^2+2)e^{-T_a}+1\right\} > 0 \qquad (5.15)$$

但し，　$\Delta\omega = |\omega_1 - \omega_2|$，　$T_a = \alpha T_a$

$S = \sin(\Delta\omega\, T_a/2)$，　$C = \cos(\Delta\omega\, T_a/2)$

を満たす最小の T_a が窓長の下限となる．式(5．15)より，$T_a = \infty$ としても $\Delta\omega > 0.972\,\alpha$ でなければ周波数スペクトル上に $\omega = \omega_1$ 及び $\omega = \omega_2$ に対応するピークがそれぞれ存在しないことになる．図5．2に $\Delta\omega$ と式(5．15)を満たす最小の T_a の関係を示す．

図5．2より，$\Delta\omega$ が $0.972\,\alpha$ に近づく（例えば，α=300 π のとき $\Delta\omega/2\pi$ = 145.8Hz）と T_a は急激に増大するが，$\Delta\omega/2\pi \geqq$ 500Hz 程度になると $50\,\pi \leqq \alpha \leqq 300\,\pi$ の範囲では α の値にかかわらず $\Delta\omega$ と T_a の関係はおおむね $\Delta\omega\, T_a = 1.3\,\pi$ の双曲線となることが分かる．

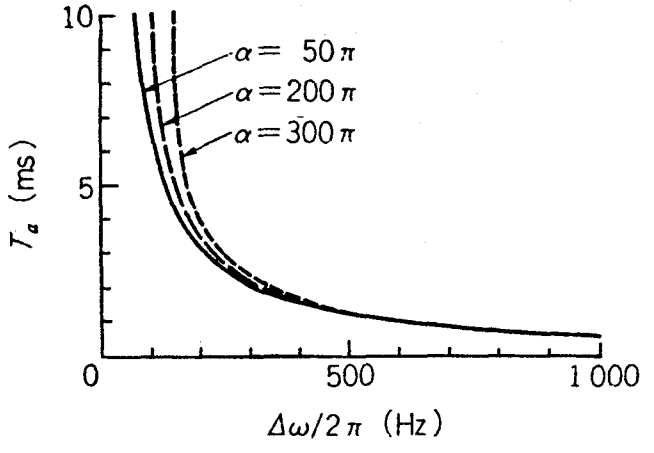


図5．2　$\Delta\omega$ と式(5.15)を満たす最小の分析窓長 T_a の関係

5.2.2　合成音による検証

5．2．1では声帯音源がインパルス列で，声道伝達特性の極数が少ない場合について考察したが，実際の有声音では声帯音源はインパルス列ではなく，声道伝達特性にも数個の極が存在する．このような場合にも同様のことが結論できるかどうかを理論的に導出するのは困難であるため，以下，合成母音を用いてシミュレーションすることによりその検証を行なう．ターミナルアナログ形のディジタルシミュレーション（音源：ピッチ周期 $T_0 = 8$ms の図5．3（a）に示す Rosenberg 波[40]，ホルマント周波数：$F_1 = 700$Hz，$F_2 = 1{,}300$Hz，$F_3 = 2{,}500$Hz，ホルマントの帯域幅 $B_i = 50\{1 + F_i^2/(6 \times 10^6)\}$Hz，高次極補正および放射特性として 12dB/oct の高域強調）により作成した合成母音/a/を用いピーク周波数に及ぼす分析窓の始点，窓長およびピッチ周期の影響を調べる．

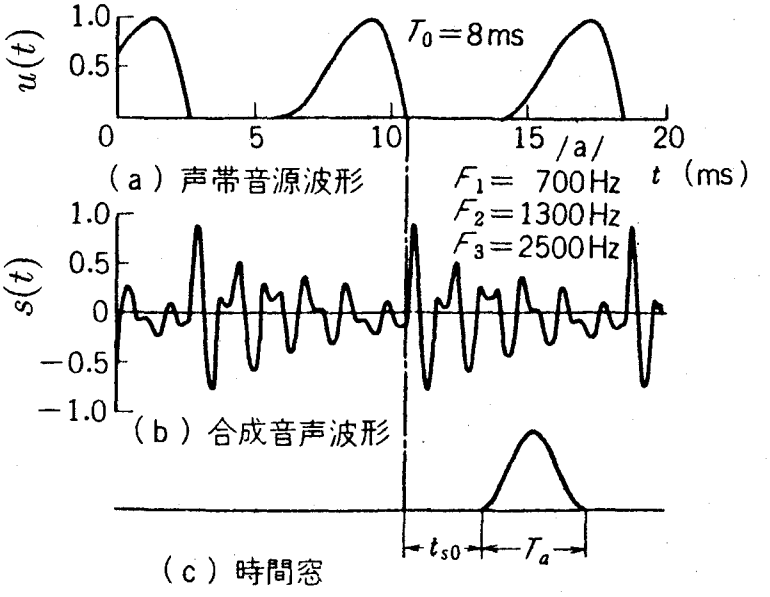


図5．3　シミュレーションに用いた声帯音源波形と
合成音声波形および t_{s0} と T_a の定義

(a)　ピーク周波数の分析窓の始点依存性

分析窓の始点の基準点（$t_{s0}=0$）を声帯音源波形の1ピッチ周期中の実効的な励振点である声門閉止時点にとり（図5.3 (c)参照），分析窓長$T_a=T_0/2(=4\text{ms})$一定とした場合のＦＦＴによる短時間周波数スペクトル上のピーク周波数$\boldsymbol{F}^p=\{F_1^p,F_2^p,F_3^p,\cdots\}$の分析窓の始点$t_{s0}$依存性を図5.4に，又，周波数スペクトル例を図5.5に示す．

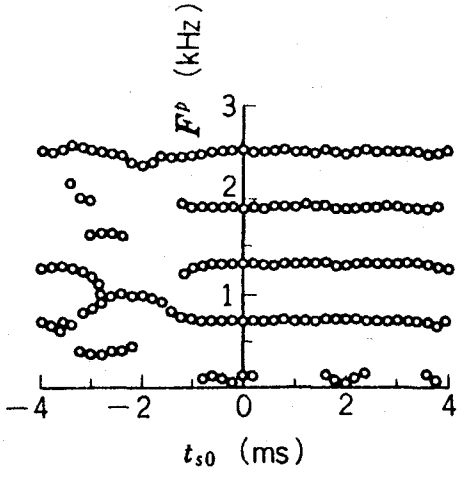
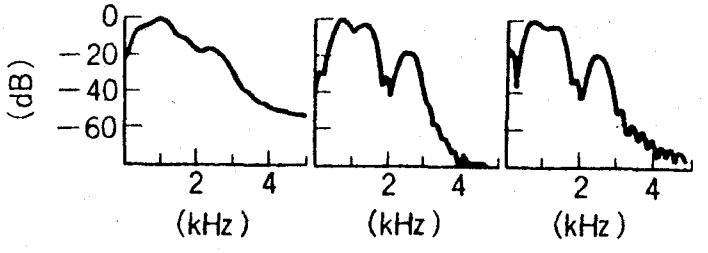


図5.4　ピーク周波数$\boldsymbol{F}^p$の分析窓の始点t_{s0}依存性
（$T_a=4\text{ms}$，合成母音/a/）



（a）$t_{s0}=-2\text{ms}$（b）$t_{s0}=0\text{ms}$（c）$t_{s0}=2\text{ms}$

図5.5　短時間周波数スペクトル例
（$T_a=4\text{ms}$，合成母音/a/）

図5．4より $-0.6\text{ms} \leqq t_{s0} \leqq 4\text{ms}$ の範囲，すなわち，今の場合，図5．3 (c) に示すHanning 窓を用いているため声門閉止時点が分析窓内に実質的に包含されていない範囲においては 200Hz 程度以下のピーク周波数を除き，ピーク周波数の周波数軸上の位置に及ぼす分析窓の始点の影響はほとんどないといえる．そして，図5．5に示すように $t_{s0} \geqq 0$ とすることにより声道伝達特性を保存した平滑な周波数スペクトル包絡が得られることがわかる．

この周波数スペクトル上でのピークピッキングによるホルマント周波数推定誤差E_iを

$$E_i = \frac{F_{ci}^p - F_i}{F_i} \tag{5.16}$$

但し，　F_{ci}^p：F_iに対応するピーク周波数

とし，$t_{s0} = 0\text{ms}$ 及び 2ms の場合の E_i を表5．1に示す．　表5．1より，簡単なピークピッキングの手法で精度よくホルマント周波数推定が可能であると言える．

表5．1　ホルマント周波数推定誤差 E_i
（分析窓長 $T_a = 4\text{ms}$）

t_{s0}(ms)	E_1(%)	E_2(%)	E_3(%)
0.0	0.4	0.7	0.8
2.0	3.2	−0.8	−0.8

(b)　ピーク周波数の分析窓長依存性

図5．3の合成母音において，t_{s0}=0ms の条件の下でピーク周波数 F^p の分析窓長 T_a 依存性を図5．6に，又，周波数スペクトル例を図5．7に示す．図5．6より，ホルマント周波数に対応するピーク周波数は分析窓長を増大させる過程で，いったん現われれば $T_a \geqq 3.4\text{ms}$ の範囲では周波数軸上の位置がほとんど変動していないといえる．すなわち，T_a によりその位置が不変なピーク周波数がホルマント周波数に対応するピーク周波数であり，この特性はホルマント周波数の自動抽出を行なう上で重要な特性となる．表5．2に $T_a = 3.0$，3.4，4.0ms 及び 5.0ms の場合の E_i を示す．

図5．7の周波数スペクトル例より明らかなように，T_a が小さければ F_1 と F_2に対応するピークが分離せず一つの大きなピークとなり $T_a \geqq 2.3\text{ms}$ で始めて二つに分離

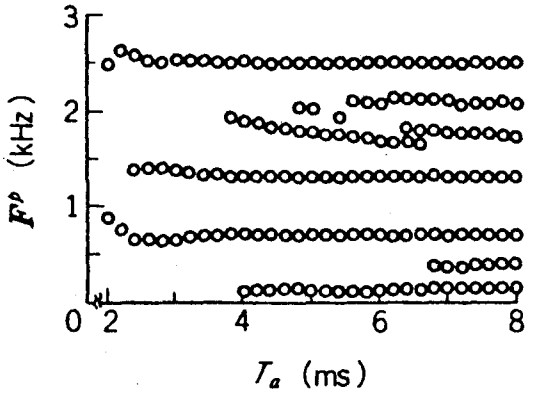


図5．6　ピーク周波数 F^p の分析窓長 T_a 依存性
（t_{s0} = 0ms，合成母音/a/）

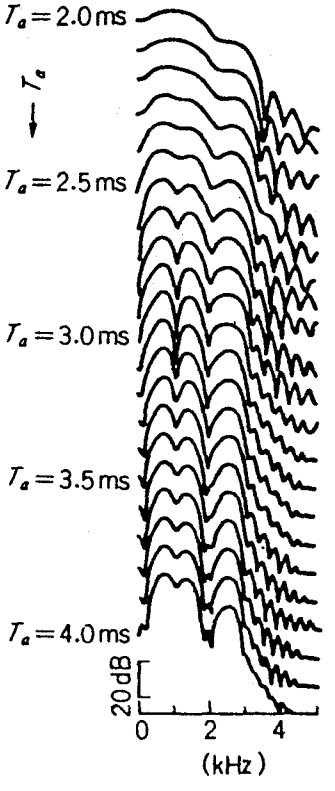


図5．7　短時間周波数スペクトルの分析窓長 T_a 依存性
（t_{s0} = 0ms，合成母音/a/）

表5．2　ホルマント周波数推定誤差 E_i
（分析窓の始点 t_{s0} = 0ms）

T_a(ms)	E_1(%)	E_2(%)	E_3(%)
3.0	−7.9	5.2	0.8
3.4	−2.3	2.2	0.8
4.0	0.4	0.7	0.8
5.0	0.4	0.7	0.0

する．今の場合，第1及び第2ホルマントの周波数ならびに帯域幅はそれぞれ F_1 = 700Hz，F_2=1,300Hz，B_1 = 54.1Hz，B_2 = 64.1Hz であるので，5．2．1（b）の $\Delta\omega$ 及び α がそれぞれ 600Hz 及び 64.1 π に相当するものと考えれば，F_1 と F_2 に対応するピークが存在するための最小分析窓長は図5．2より約 1.1ms となる．ところで，分析窓として5．2．1（b）では方形窓を，本シミュレーションでは Hanning 窓を用いた場合である．一方，分析窓長 T_a の Hanning 窓の実効的な窓長 $T_{ae} = 0.67T_a$ と考えられる[70]．したがって，窓長 2.3ms の Hanning 窓は実効的に 1.54ms の窓長に相当する．更に，合成母音の第1及び第2ホルマントには振幅レベル差がある．これらを考慮すれば $T_a \geqq$ 2.3ms で F_1 と F_2 に対応するピークが存在するというシミュレーション結果は5．2．1（b）の結果とほぼ一致すると言える．

（c）　ピーク周波数のピッチ周期依存性

図5．3の合成母音作成と同様なディジタルシミュレーションにおいて，声帯波の開放係数（声門開口区間と声門閉止区間の比）を一定とし周期 T_0 のみを変化させた合成母音において t_{s0} = 0ms，T_a = 4ms の条件の下でピーク周波数 F^p のピッチ周期 T_0 依存性を図5．8に示す．図5．8より，$T_0 \geqq$ 4.6ms であれば 150Hz 付近のピーク周波数を除き，ピーク周波数の周波数軸上の位置に及ぼすピッチ周期 T_0 の影響はほとんどないと言える．

以上のように単母音の短時間周波数スペクトルにも5．2．1の解析で得られたのと同様な特性があり，分析窓長を1ピッチ周期未満と短くし，かつ1ピッチ周期中の声道が声帯音源により実効的に励振された時点（例えば，図5．3（a）に示す声帯波形の

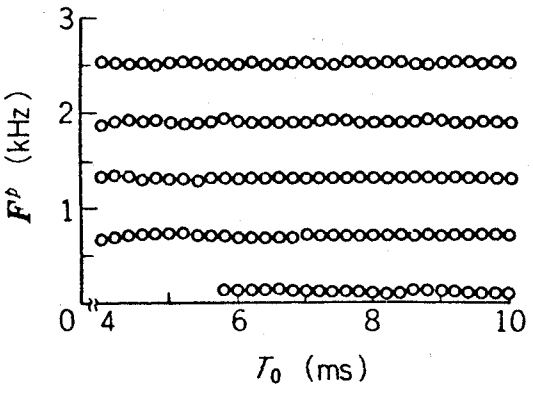


図5．8　ピーク周波数 F^p のピッチ周期 T_0 依存性
（t_{s0} = 0ms，T_a = 4ms，合成母音/a/）

場合には声門閉止時点となる）が実質的に包含されないように分析窓の位置を設定すれば声道伝達特性を保存した平滑な周波数スペクトル包絡が得られ，そのピーク周波数から精度よくホルマント周波数推定が可能であるといえる．以下，この方法をＦＲＡＰＳ（FRActional Period Spectral analysis）法と称することにする．

5.2.3　有声破裂音のホルマント周波数追尾

ＦＲＡＰＳ法は分析窓長が 1/2 ピッチ周期程度と従来の周波数分析の分析窓長（一般に 20～30ms）に比しずっと短いのにもかかわらず声道の伝達特性が保存されている平滑な周波数スペクトル包絡が得られる．このためホルマント周波数が時間的に変化している音声の過渡部のホルマント周波数を追尾する場合この方法は特に有用であると考えられる．その例証として以下，合成および自然有声破裂音のホルマント周波数追尾例を示す．

(a)　合成音への適用結果

5．2．2の合成母音作成と同様なディジタルシミュレーションにおいてホルマント周波数が図5．9の実線で示すように時間的に変化している合成有声破裂音にＦＲＡＰＳ法（t_{s0}=0ms，T_a=T_0/2=4ms）を適用して得られたホルマント周波数に対応するピーク周波数 F^p_{ci} を同図に○印で示す．一方，比較のためケプストラム法による周波数スペクトル包絡上のピークピッキングにより得られた F^p_{ci} を同図にΔ印で示す．な

お，ケプストラム法による周波数スペクトル包絡推定には，窓長 30ms の Hanning 窓を用い，窓の中心位置をＦＲＡＰＳ法の窓の中心位置に設定し，ケプストラム上での低域ろ波として式(５．１７)に示す荷重関数 $w_l(t)$ を用いた．但し，$\tau_1=\Delta\tau=1.5$ms とした．以下，この方法をＳＥ（Spectral Envelope）法と略称する．

$$w_l(t)=\begin{cases}1 & t<\tau_1\\ 0.5\{1+\cos(\pi(t-\tau_1)/\Delta\tau\} & \tau_1\le t<\tau_1+\Delta\tau\\ 0 & t\ge\tau_1+\Delta\tau\end{cases}\qquad(5.17)$$

図５．９より次のことが分かる．(1) ＦＲＡＰＳ法によれば，過渡部においてＳＥ法よりもより正確にホルマント周波数が推定されている．(2) 特に有声破裂音の弁別に重要なホルマントローカス点において表５．３に示すようにＦＲＡＰＳ法によれば推定誤差がＳＥ法よりも１桁程度改善される．又，ピークピッキングを適用すべき周波数スペクトルを求めるためにＳＥ法では３回のＦＦＴを行なう必要があるが，ＦＲＡＰＳ法では１回のＦＦＴでよく処理時間も短縮できる利点がある．

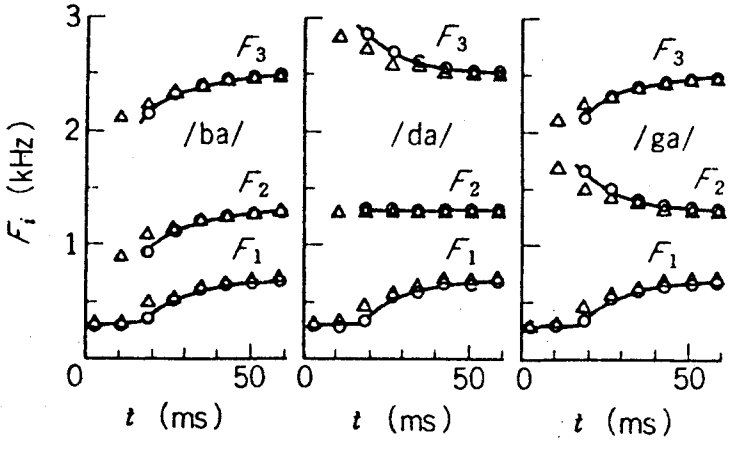


図５．９　合成有声破裂音のホルマント周波数追尾例
－：合成音のホルマント周波数設定値
○：ＦＲＡＰＳ法による推定値
△：ＳＥ法による推定値

表5．3　合成有声破裂音のホルマントローカス点における
ホルマント周波数推定誤差 E_i

有声破裂音	FRAPS法			SE法		
	E_1(%)	E_2(%)	E_3(%)	E_1(%)	E_2(%)	E_3(%)
/ba/	2.2	−0.7	0.2	36.2	13.8	3.8
/da/	2.2	0.7	−0.2	30.5	−2.4	−4.9
/ga/	2.2	0.3	−0.7	30.5	−10.4	3.8

(b)　自然有声音への適用結果

FRAPS法による単音節 /ga/（男声）のホルマント周波数追尾結果をSE法と対比させて図5．10に示す．但し，FRAPS法の適用に際しては付録Cに示すアルゴリスムを用い，分析窓の始点の決定は音声波の1ピッチ周期中の最大振幅点直前の最大振幅値の 20% 交差点とした（詳しくは文献(71)参照）．

図5．10より次のことが言える．(1) SE法によって推定された第1ホルマント周波数には不自然な変化が見られるのに対し，FRAPS法によって推定された第1ホルマント周波数は時間的な連続性がよい．(2) 過渡部においてSE法では図5．10中に示す周波数スペクトル例の (d) 及び (e) から明らかなように 1,600Hz 付近に第3ホルマント周波数と同定される恐れのあるピークが存在するが，ホルマント周波数の時間的連続性から推察してこれを第3ホルマント周波数と同定するのは不自然と思われる．これに対し，FRAPS法ではそのようなピークは存在しない．すなわち，SE法では過渡部においてスプリアスなピークが多く存在するがFRAPS法では少ない．

以上のことから，有声破裂音の過渡部に対してはSE法よりもFRAPS法の方がホルマント周波数の追尾性が良いといえる．

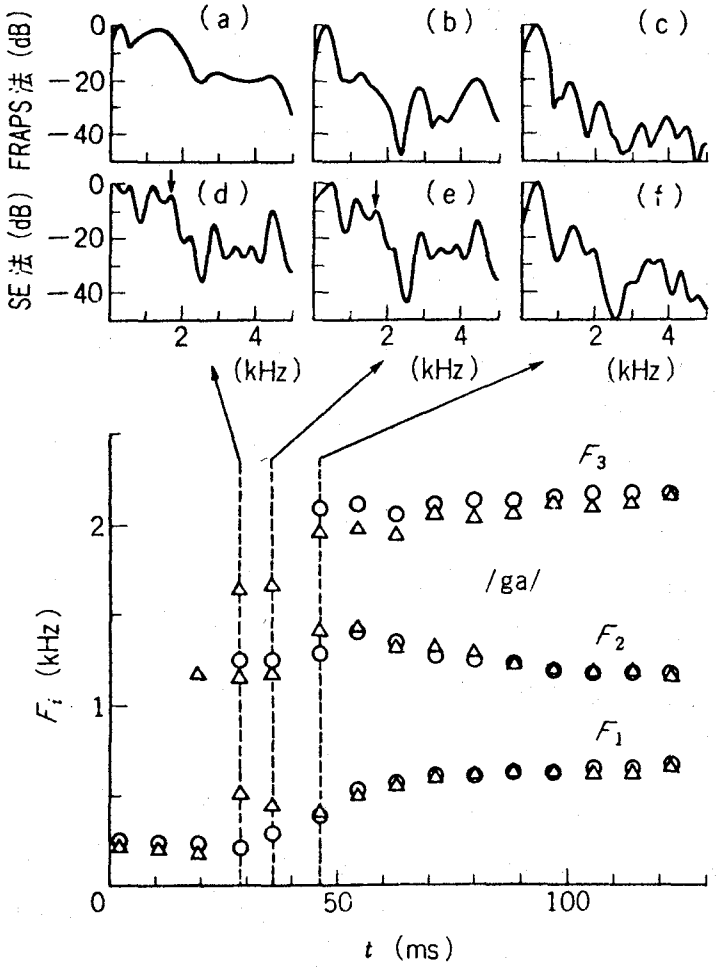


図５．１０　単音節/ga/（男声）のホルマント周波数追尾および周波数スペクトル例

○：ＦＲＡＰＳ法による推定値

△：ＳＥ法による推定値

5.3 窓長漸減型線形予測分析による ホルマント周波数推定[72]

通常の線形予測法は第2章で述べたように，分析窓内での定常性が仮定されているため，ホルマント周波数が急激に変化している音声の過渡部のホルマント周波数を正確に推定するには，分析窓長を数 ms 程度に短くする必要があると言える．しかしながら，有声音の場合，分析窓長を1ピッチ周期程度以下に短くすると，分析窓と励振点との相対位置の影響が生じるため[73]~[75]，声門閉止区間分析[60]~[62]あるいは励振源を考慮した分析[59],[76]等が必要となる．しかしながら，声門閉止区間分析ではホルマント周波数の推定値がピッチ周期毎にしか得られず，ホルマント周波数の急激な時間的変化を追尾するには時間分解能が不十分となり易い．また，励振源を考慮した分析では励振源パラメータの推定が不適切な場合には，ホルマント周波数の推定精度がかえって悪くなると言った問題がある．励振源の影響を軽減するために荷重を導入した線形予測モデル[77]も提案されているが，それを用いた過渡音に対する小区間分析の効果はまだ検討段階である．一方，分析窓内での非定常性を考慮した線形予測法も検討されてはいるが，予測係数の時間的変化の近似が必要であったり[78],[79]，時間的連続性が悪い[44]など，最適な近似空間の適否あるいは時間的追従性の問題がある．推定パラメータの時間的連続性に関しては改善が試みられているが[80]，一般に非定常性を考慮した分析手法は処理手順が複雑になるといった問題もあり，今後の研究課題であると言える．

本節では，通常の線形予測分析を用いて，分析窓の任意の点（始点，中心等）を固定し，窓長を徐々に短くしていった一連の分析結果に基づき，分析窓長が零になる場合の値を外挿すれば，分析窓長を極端に短くすることによる弊害を受けることなく，過渡部の任意の時点のホルマント周波数が安定に精度よく推定できることを示す．以下，5.3.1において，窓長漸減型線形予測分析の概略を示し，5.3.2で，その理論的基礎として線形予測分析による極周波数推定値の分析窓長依存性を解析的に考察する．そして，5.3.3において，合成音のシミュレーションにより本手法のホルマント周波数推定精度の改善度合を示し，5.3.4では，実際に自然有声破裂音のホルマント周波数軌跡推定に適用し，本手法の有効性を示す．

5.3.1　窓長漸減型線形予測分析

今,分析窓長をT_aとした時,分析窓の始端からγT_a（但し，$0 \leqq \gamma \leqq 1$）の分析窓中の時点（端点を含む）を音声の特徴パラメータ（ホルマント周波数等）の瞬時的な値を推定しようとする時点t_0に一致させ，γを一定値に保ったまま，窓長T_aを漸減した一連の分析窓を設定し，それらの窓に対応する各々の線形予測分析の結果から，分析窓長が零になる場合の特徴パラメータの値を外挿推定する手法を窓長漸減型線形予測分析と名付ける（図5．11参照）．勿論，γは$0 \leqq \gamma \leqq 1$の任意の値を取り得るが，窓長漸減型線形予測分析の典型としては，$\gamma = 0$とした始点固定型，$\gamma = 0.5$とした中心固定型，$\gamma = 1$とした終点固定型などが考えられる．

窓長漸減型線形予測分析は分析窓長を極端に短くすることなく，窓長が零になる場合の値を推定する手法であるため，分析窓長を実際に1ピッチ周期程度以下に短くした場合に生じる難点を避けることのできる分析手法と言える．したがって，本手法は音声の過渡部のホルマント周波数推定に，特に有用であると考えられるので，この点について詳細に検討した結果を以下に述べる．なお，始点固定型，中心固定型および終点固定型はそれぞれ語頭，語中および語尾における音声の特徴パラメータ推定に有用と考えられるが，終点固定型は基本的には始点固定型で時間軸を反転したものと言えるので，本論文では，始点固定型と中心固定型に関して検討した結果を述べる．

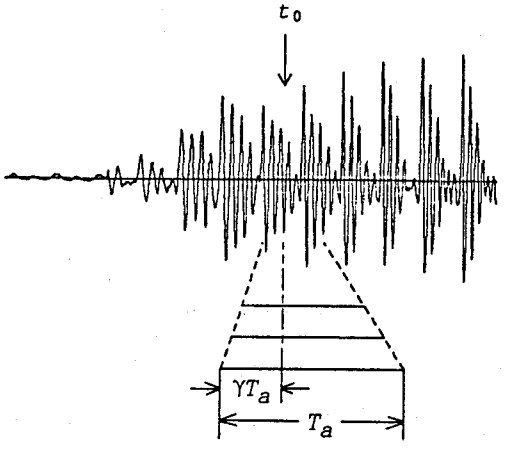


図5．11　窓長漸減型線形予測分析

5.3.2 線形予測分析による極周波数推定値の分析窓長依存性

本手法の基本的な特性を明らかにするため，振幅および周波数が時間と共に線形に変化している式（5．18）の過渡モデル音を用いて，通常の線形予測分析による極周波数推定値の分析窓長依存性の解析的検討を行う．

$$s(t) = \{1 + (t - t_0)\Delta A\} \sin\left\{ \omega_0 t + (t - t_0)^2 \Delta\omega/2 - \omega_0 t_0 \right\} \qquad (5.18)$$

但し，ΔA および $\Delta\omega$ はそれぞれ振幅および周波数の時間的変化率である．

通常の線形予測法では，周知のように，信号波の自己相関係数に基づく正規方程式の解を係数とする高次方程式の根から極周波数が推定される．したがって，線形予測法による極周波数推定値の分析窓長依存性を解析するために，信号 $s(t)$ の任意の時刻 $t = t_s$ から $t = t_s + T_a$ までの区間における自己相関係数 $R(\tau, t_s, T_a)$ を式（5．19）のように定義する．

$$R(\tau, t_s, T_a) = \int_{t_s}^{t_s+T_a} s(t)s(t+\tau)dt \quad / \quad \int_{t_s}^{t_s+T_a} s^2(t)dt \qquad (5.19)$$

今，式（5．18）で表わせる過渡モデル音を窓長 T_a，分析次数 $p = 2$ で線形予測分析する場合を考える．

(a) 始点固定型

始点固定型は図5．11において $\gamma = 0$ として分析窓の始点を特徴パラメータ推定時点 t_0 に固定し，分析窓長 T_a を変化させるので，この場合の自己相関係数は式（5．19）において $t_s = t_0$ とし，これに式（5．18）を代入すれば，

$$R(\tau, t_0, T_a) = \sin\frac{(\Delta\omega\,\tau\,T_a/2)}{\Delta\omega\,\tau\,T_a/2} \cos(\omega_0\tau + \Delta\omega\,\tau\,T_a/2 + \phi_1) \qquad (5.20)$$

但し，$\phi_1 = \tan^{-1}\dfrac{2\Delta A}{(1 + \Delta A T_a)\Delta\omega\,\tau}\{1 - \Delta\omega\,\tau\,T_a/2cot(\Delta\omega\,\tau\,T_a/2)\}$

となる（付録D参照）．

一方，分析次数 $p=2$ の線形予測法による極周波数推定値 $\hat{F}$ は

$$\hat{F}=\frac{1}{2\pi T}\cos^{-1}\frac{-\alpha_1}{2\sqrt{\alpha_2}} \tag{5.21}$$

但し，

$$\alpha_1=\frac{r_1(r^2-r_0)}{r_0^2-r_1^2}\quad,\quad \alpha_2=\frac{r_1^2-r_0r^2}{r_0^2-r_1^2}$$

r_i：遅延 iT の自己相関係数

T：標本化周期

となる．そして，信号 $s(t)$ の振幅の時間的変化率 ΔA が比較的小さい場合には，Z 平面上の根はほぼ単位円上付近にあると言えるので，$\alpha_2\approx 1$ となる．この場合，式 (5.21) より，

$$\hat{F}=\frac{1}{2\pi T}\cos^{-1}(r_1/r_0) \tag{5.22}$$

となる．したがって，式 (5.18) の信号 $s(t)$ を分析次数 $p=2$，分析窓長 T_a の線形予測分析を行って得られる極周波数推定値 $\hat{F}$ は，式 (5.20)，(5.22) より，

$$\hat{F}=\frac{1}{2\pi T}\cos^{-1}\left\{\frac{\sin(\triangle\omega TT_a/2)}{\triangle\omega TT_a/2}\cos(\omega_0T+\triangle\omega TT_a/2+\phi_1)\right\} \tag{5.23}$$

但し，$\phi_1=\tan^{-1}\dfrac{2\Delta A}{(1+\Delta AT_a)\Delta\omega T}\{1-\Delta\omega TT_a/2\cot(\Delta\omega TT_a/2)\}$

となる．ところで，$\triangle F=\triangle\omega/2\pi=10$Hz/ms（音声の過渡部では一般にこのオーダーでホルマント周波数が変化していると言える），$T_a=20$ms，$T=0.1$ms の場合，$\triangle\omega TT_a/2=0.02\pi$ であるので $\sin(\triangle\omega TT_a/2)/(\triangle\omega TT_a/2)\approx 1$ と近似できる．したがって，式 (5.23) より，

$$\hat{F}=F_0+\triangle FT_a/2+\phi_1/2\pi T \tag{5.24}$$

となる．すなわち，式 (5.18) のように振幅および周波数が時間と共に線形に変化している過渡モデル音 $s(t)$ の場合，式 (5.24) より，分析窓の始点を $t=t_0$ に固定した通常の線形予測分析による周波数推定値は『分析窓の中心位置での過渡モデル音の瞬時周波数と $\phi_1/2\pi T$ の和』となり，信号 $s(t)$ の振幅が変化せず（$\Delta A=0$），

周波数のみが線形に変化する場合は，$\phi_1 = 0$ となるので，周波数推定値は分析窓の中心位置での瞬時周波数となることがわかる．

(b) 中心固定型

中心固定型は図5．11において $\gamma = 0.5$ として分析窓の中心を特徴パラメータ推定時点 t_0 に固定し，分析窓長 T_a を変化させるので，この場合の自己相関係数は式(5．19)において $t_s = t_0 - T_a/2$ とし，これに式(5．18)を代入すれば，前節と同様の導出過程より，

$$R(\tau, t_0, T_a) = \frac{\sin(\Delta\omega\tau T_a/2)}{\Delta\omega\tau T_a/2}\cos(\omega_0\tau + \phi_2) \tag{5.25}$$

$$\text{但し，}\quad \phi_2 = \tan^{-1}\frac{2\Delta A}{\Delta\omega\tau}\{1 - \Delta\omega\tau T_a/2\cot(\Delta\omega\tau T_a/2)\}$$

となる．

したがって，式(5．18)の信号 S(t) を分析次数 $p = 2$，分析窓長 T_a の線形予測分析を行って得られる極周波数推定値 $\hat{F}$ は中心固定型の場合，式(5．22)，(5．25)より，

$$\hat{F} = \frac{1}{2\pi T}\cos^{-1}\left\{\frac{\sin(\triangle\omega T T_a/2)}{\triangle\omega T T_a/2}\cos(\omega_0 T + \phi_2)\right\} \tag{5.26}$$

$$\text{但し，}\quad \phi_2 = \tan^{-1}\frac{2\Delta A}{\Delta\omega T}\{1 - \Delta\omega T T_a/2\cot(\Delta\omega T T_a/2)\}$$

となる．ここで，前節の式(5．23)から式(5．24)を導出したのと同様の近似を行えば，式(5．26)より，

$$\hat{F} = F_0 + \phi_2/2\pi T \tag{5.27}$$

となる．すなわち，式(5．18)の過渡モデル音 $s(t)$ の場合，式(5．27)より，分析窓の中心を $t = t_0$ に固定した通常の線形予測分析による周波数推定値は『$t = t_0$での過渡モデル音の瞬時周波数と $\phi_2/2\pi T$ の和』となり，信号 $s(t)$ の振幅が変化せず($\Delta A = 0$)，周波数のみが線形に変化する場合は，$\phi_2 = 0$ となるので，周波数推定

値は周波数の時間的変化率にかかわらず $t = t_0$ での瞬時周波数 F_0 と一致することがわかる.

(c) 数値計算例

式(5.18)に示す過渡モデル音の時刻 $t = t_0$ における周波数推定値の分析窓長 T_a 依存性を図5.12に示す. 但し, 標本化周期 $T = 0.1$ms, $F_0 = \omega_0/2\pi = 1000$Hz, $\Delta F = \Delta\omega/2\pi = 15$Hz/ms, $\Delta A = 0.03$/ms とし, 図中の○, △, □および●印はそれぞれ $\gamma = 0$, 0.2, 0.4 および 0.5 とし分析窓中の γT_a の時点を $t = t_0$ に一致させ通常の線形予測分析 (分析次数 $p = 2$) を行って得られた値, また図中の○印近傍の実線は式(5.23), 破線は式(5.24)による計算値, そして●印近傍の実線は式(5.26), 破線は式(5.27)による計算値である (図中の点線は後述).

図5.12より, 式(5.18)に示す過渡モデル音を通常の線形予測分析して得られる周波数推定値は実線すなわち $\gamma = 0$ (始点固定型) の場合には式(5.23)による計算値, また $\gamma = 0.5$ (中心固定型) の場合には式(5.26)による計算値とほぼ一致し, 本解析の妥当性が示されていると言える. また, 破線 ($\gamma = 0$ の場合, 式(5.24), $\gamma = 0.5$ の場合, 式(5.27)による計算値) とも比較的よく一致していると言え, 式(5.24)および式(5.27)はそれぞれ $\gamma = 0$ および $\gamma = 0.5$ に対する近似式として十分妥当であると言える. さらに重要な特徴は, $\hat{F}/F_0$ は分析窓長 T_a と共に, $\gamma = 0$ の場合にはほぼ直線的に, また, $\gamma = 0.5$ の場合には $T_a = 0$ に対称軸を持つほぼ2

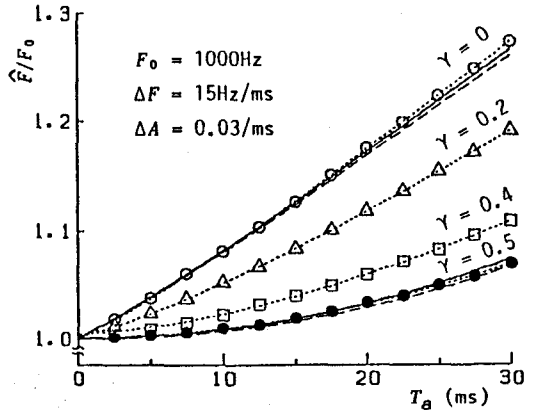


図5.12 周波数推定値 $\hat{F}$ の分析窓長 T_a 依存性
— 過渡モデル音 —

次曲線的に変化し，$0 < \gamma < 0.5$ の場合にはその中間となっていることである．すなわち，分析窓中の γ T_a の時点を固定にした窓長の異なる線形予測分析の結果を，

$$f(T_a) = a(T_a)^{\zeta} + b \tag{5.28}$$

但し，　$\zeta_1 \leqq \zeta \leqq \zeta_2$

なる関数で最小自乗近似し，$T_a \rightarrow$ 0ms の値を外挿すれば，非定常な場合でも分析窓長を極端に短くすることなく，通常の線形予測分析で正確な極周波数推定が可能になると期待できる（この関数において $\zeta_1 = 1$, $\zeta_2 = 2$ とし，ζ を 0.01 の精度で $T_a =$ 10～30ms の分析結果を最小自乗近似した場合，図中の各点線となり，各点線は $T_a =$ 0ms に於いていずれも $\hat{F}/F_0 \approx 1$ となる）．なお，今の場合，分析窓長を十分短く $T_a =$ 2.5ms とすれば，ほぼ正確な推定値が得られているが，これは本過渡モデル音が励振源のないいわゆるＡＭ・ＦＭ音のため励振源の影響がないからである．

【ζ の範囲】

式(5．２８)による最小自乗近似は ζ に関して非線形なためここでは 0.01 の精度で最適な ζ を求めたが，ζ_1 および ζ_2 を γ の関数とすることにより計算量を大幅に軽減することができる．すなわち，前節の結果より，基本的には γ のいかんにかかわらず $\zeta_1 = 1$，$\zeta_2 = 2$ とすれば十分であるが，図５．１２から明らかなように，γ $= 0$ および $\gamma = 0.5$ 付近の最適な ζ はそれぞれ 2 および 1 程度になることはないと言える．図５．１３に，$T_a = 10$～30ms の分析結果を最小自乗近似して得られる $T_a \rightarrow$ 0ms の外挿値が ± 1%以内の誤差となる ζ の範囲を γ の関数として示す．但し，F_0 $=$ 500～3000Hz，$\Delta F = -20$～20Hz/ms，$\Delta A = 0$～0.3/ms の範囲（但し，$F_0 =$ 500Hz の時，ΔFの下限は −10Hz/ms）で変化させた計 312 個の過渡モデル音を用いて ζ のきざみ幅 0.01 の精度で求めた．

図５．１３より，最適な ζ を探索すべき範囲は非常に限られており，特に，$\gamma \geqq 0.45$ の時は最適な ζ を探索する必要がないことがわかる．なお，ΔF およびΔA の変化範囲は実際の音声の過渡部でのホルマント周波数ならびに振幅レベルの時間的変化を考慮したものであり（例えば，30ms で振幅レベルが 1～10 倍，周波数が 0～± 600Hz 変化），以後，式(５．２８)の ζ_1 および ζ_2 として，それぞれ図５．１３に示す領域の下限値および上限値を用いる．

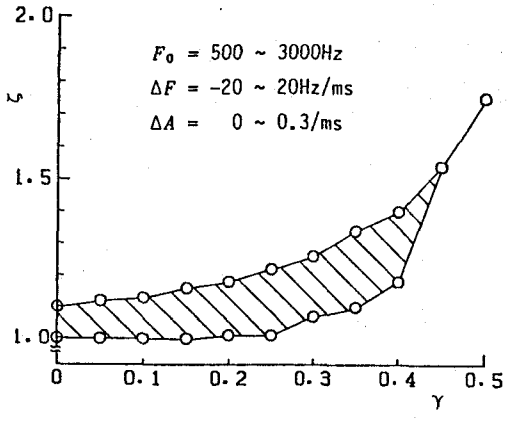


図5．１３　過渡モデル音に対する ζ の適切な範囲
（γ の関数として示す）

以上の結果は正弦的振動波において振幅ならびに周波数が線形に変化している場合の解析結果であるが，実際の音声の過渡部では数個の極が一般には指数関数的に変化していると考えられる．しかし，このような場合にも同様のことが結論できるかどうかを解析的に導出するのは困難であるため，以下，合成音を用いたシミュレーションによりその検証を行った結果について述べる．

5.3.3　合成音による検証

過渡的音声の代表例と言える有声破裂音を用いて前章の検証を行う．図5．１４に合成有声破裂音 /ga/ における極周波数推定値の分析窓長 T_a 依存性を示す．但し，合成条件は標本化周波数 10kHz，励振源：ピッチ周期 8ms の Rosenberg 波[40]（但し，破裂時点から 2ms 長のノイズバースト付加），ホルマント周波数：$F_1 \sim F_3$ は時変（図5．１６の実線参照），$F_4 = 3437.5$Hz 一定（但し，図5．１６で t $\geqq$ 20ms），放射特性：6dB/oct であり，分析は前処理として一階差分後，分析窓の始点から γ T_a の時点を破裂時点から 17ms 後の過渡部に固定し（図5．１４上段参照），分析次数 $p = 10$ で通常の線形予測分析を行ったもので，○および△印はそれぞれ $\gamma = 0$（始点固定型）および $\gamma = 0.5$（中心固定型）とした場合の結果である．なお，縦軸左側の矢印は破裂時点から 17ms 後における合成音のホルマント周波数である．

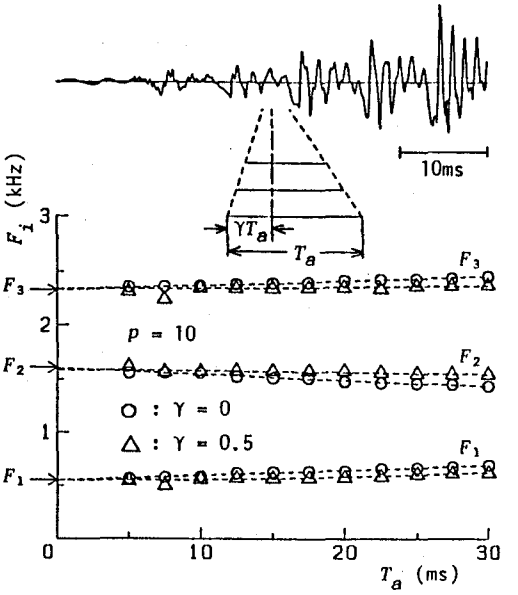


図5．14　極周波数推定値の分析窓長 T_a 依存性
－ 合成有声破裂音 /ga/ －

図5．14より，複数の極が時間と共に指数関数的に変化している場合でも，5．3．2の解析結果と同様，各窓長に対する個々の極周波数推定値は T_a と共に，$\gamma = 0$ の場合には，ほぼ直線的に，また，$\gamma = 0.5$ の場合には，ほぼ2次曲線的に変化していることがわかる．5．3．2と同様，$T_a = 10 \sim 30$ms の分析結果を式(5．28)で最小自乗近似した場合，それぞれ図中の点線となり，これらの点線の $T_a = 0$ms における値はいずれも合成音のホルマント周波数とほぼ等しくなると言える．すなわち，分析窓の始点から γT_a の時点を固定にした分析窓長の異なる通常の線形予測分析結果を式(5．28)で最小自乗近似した時の $T_a \rightarrow 0$ms における値を求めれば，分析窓長を極端に短くすることなく正確なホルマント周波数が推定できると言える．

ところで，どのような窓長の分析結果に基づいて $T_a \rightarrow 0$ms における値を外挿するのが適当であるかが問題になるが，一般に分析窓長の最短値が1ピッチ周期程度以下に短くなると，分析窓と励振点との相対位置の影響ならびに分析データ点数が少なくなることによる弊害が生じる．このことを考慮して，分析窓長の最短値は10msとし，ホルマント周波数推定誤差の外挿データ点数 N 依存性を図5．15に示す．但し，図5．15は前処理として一階差分後，分析次数 $p = 10$ とし，破裂時点から破裂時点後

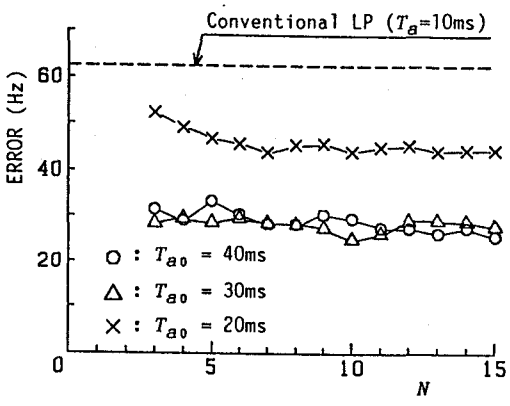


図5．15　ホルマント周波数推定誤差の外挿データ点数 N 依存性
－ 合成有声破裂音 /ga/ －

10ms までの区間について 2ms 間隔毎に外挿して求めた計6フレームの第1～第3ホルマント周波数推定誤差の平均値で，○，Δおよび×印は漸減窓長の初期値（最長値）T_{a0} をそれぞれ 40, 30 および 20ms とし，$T_a \rightarrow$ 0ms の値を式(5．28)に基づき最小自乗近似外挿した場合の結果である．なお，γ は $T_a = T_{a0}$ の時に分析窓の始点が破裂時点以前とならないように式(5．29)により設定した．

$$\gamma = \frac{t_0 - t_b}{T_{a0}} \tag{5.29}$$

但し，t_0: 分析時点，t_b: 破裂時点であり，$\gamma > 0.5$ となる場合には $\gamma = 0.5$ とする．また，通常の線形予測分析（共分散法）において分析窓長を 10ms とした場合の誤差を図中破線にて示す．

図5．15より，外挿データ点数 N が7点以上であればホルマント周波数推定誤差の外挿データ点数依存性はほとんどなく，また，漸減窓長の初期値が 20～40ms のいずれでも，通常の線形予測分析において分析窓長を 10ms（本手法での最短の分析窓長）とした場合よりホルマント周波数推定誤差が改善することがわかる．そして，今の場合，漸減窓長の初期値 T_{a0} = 30ms で外挿データ点数が 10 の時，すなわち，窓長を 30ms から 10.2ms まで 2.2ms ずつ漸減して得られる 10 個の分析結果を最小自乗近似し，$T_a \rightarrow$ 0ms における値を外挿すれば，ホルマント周波数推定誤差の平均値が 62.5Hz から 24.9Hz に大幅に改善することがわかる．なお，合成有声破裂音 /ba/ お

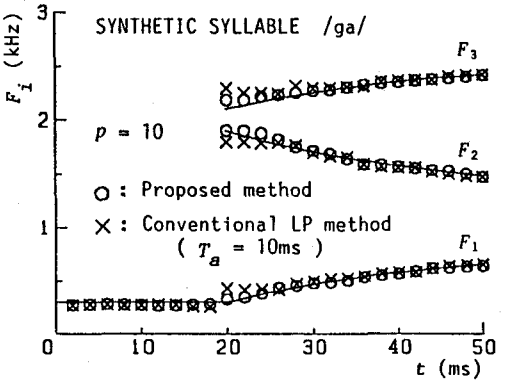


図5．16　ホルマント周波数軌跡推定の比較
－ 合成音有声破裂音 /ga/ －

よび /da/ においても，ホルマント周波数推定誤差の平均値が通常の線形予測分析ではそれぞれ 69.4Hz および 53.2Hz であったのが上記と同じ外挿条件でそれぞれ 24.7Hz および 14.5Hz に改善した．

合成有声破裂音 /ga/ のホルマント周波数軌跡推定例を図5．16 に示す．但し，前処理として一階差分後，分析次数 $p = 10$，フレーム間隔は 2ms とし，○印は本手法による推定値（各分析フレームにおいて，γ を式(5．29)により設定し，窓長を 30ms から 10.2ms まで 2.2ms ずつ減少させた線形予測分析を行って得られる極周波数を式(5．28)で最小自乗近似した時の $T_a \rightarrow 0$ms における値），×印は通常の線形予測分析にによる推定値（分析窓長 10ms，分析窓の中心を分析時点とみなす）であり，実線は合成音のホルマント周波数を示す．

図5．16より，通常の線形予測分析において，分析窓長を 10ms と 1 ピッチ周期以上にすると，分析位置と励振点との相対位置関係が原因でホルマント周波数推定誤差が極端に大きくなるようなことは起こらないが（$t = 28$ms 付近の第3ホルマント周波数推定値に若干の影響がみられる），通常の方法では有声破裂音の相互識別に重要となるホルマントローカス（遷移開始時点）付近の推定誤差が大きいのに対して，本方法での最小の分析窓長は 10.2ms と今の場合の通常の線形予測分析の分析窓長とほぼ同じに設定してあるにもかかわらず，本方法の方がより正確なホルマント軌跡が推

定でき，特に，ホルマントローカス付近の推定誤差が大幅に改善されていると言える．

5.3.4　自然音声への適用結果

成人男性が発声した単音節 /ga/ における極周波数の分析窓長 T_a 依存性を図5．17に示す．但し，前処理として一階差分後，分析窓の始点から $\gamma\ T_a$ の時点を破裂時点から 17ms 後の過渡部に固定し（図5．17上段参照），通常の線形予測分析（分析次数 $p = 12$，窓長：30ms から 10.2ms まで 2.2ms 間隔で減少）を行ったもので，○および△印は γ をそれぞれ 0 および 0.5 とした場合の結果である．図中の各点線は窓長を漸減して得られた極周波数を式(5．28)で最小自乗近似したものである．

図5．17より，今の場合，本分析位置付近では第3ホルマントがほぼ定常状態となっているため（図5．18 /ga/ の例で $t = 27$ms 付近参照），第3ホルマント周波数推定値は γ ならびに T_a にかかわらずほぼ一定値となるが，第1および第2ホルマ

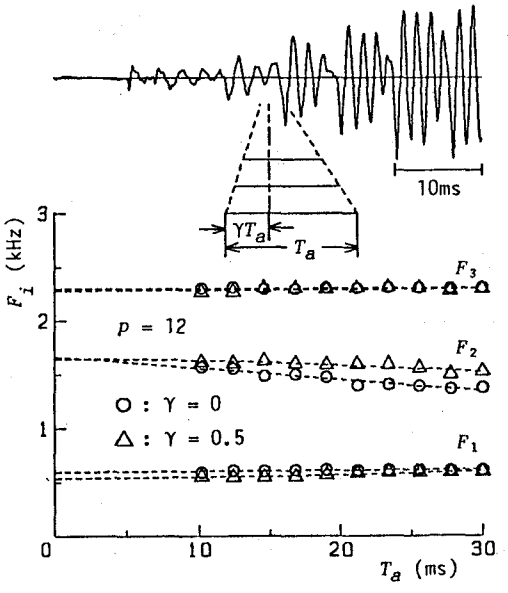


図5．17　極周波数推定値の分析窓長 T_a 依存性
－ 自然有声破裂音 /ga/ －

ント周波数推定値の T_a 依存性には前章の合成音の場合と同様の特性があると言える．但し，$\gamma = 0$ の第1ホルマント周波数推定値の T_a 依存性が合成音の場合と多少異なるため，$T_a \rightarrow$ 0ms の外挿値に $\gamma = 0$ と $\gamma = 0.5$ では若干の差が生じている．これは本分析位置で $\gamma = 0$ とした場合，$T_a \geqq$ 15ms において第1ホルマントがほぼ定常状態となっている区間が主な分析対象区間となるからである．すなわち，始点固定型（$\gamma = 0$）は分析位置以降のホルマント変化のみに基づき分析位置での値を推定するのに対し，中心固定型（$\gamma = 0.5$）の推定値は分析位置前後のホルマント変化に基づいているため，中心固定型による推定値の方がより信頼性があると考えられる．しかしながら，破裂音の破裂時点のように声道特性が急変する付近を中心固定型で分析すると，分析窓内に声道特性が急変する時点を含む窓と含まない窓が混在し推定値が不安定となるので，このような付近では始点固定型の方が良いと言える．以上のような特徴を考慮して γ を設定すれば，自然音声においても，分析窓中の任意の点を固定にした分析窓長の異なる通常の線形予測分析の結果を式(5．2 8)により最小自乗近似した時の $T_a \rightarrow$ 0ms における値を用いれば，より正確なホルマント周波数推定が可能であると推察される．

成人男性が発声した有声破裂音 /ba/，/da/ および/ga/のホルマント周波数軌跡推定例を図5．1 8に示す．但し，前処理として一階差分を行い，分析次数 $p = 12$，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5．1 6と同じである．但し，実際の有声破裂音では，分析窓長を 10ms 程度に短くすると，特に破裂時点付近において第3ホルマント周波数推定値のバラつきが大きくなる場合があるので，本手法では，窓長を漸減して得られる極が窓長を漸減する前の極と比較して，しきい値（今の場合，± 10%）以内になければ最小自乗近似から除くと共に，γ の設定を分析時点が視察により求めた破裂時点から 10ms までは（図5．1 8で 10ms ≦ t ≦ 20ms）零とし，それ以降は式(5．2 9)の分子を $t_0 - t_b -$ 10ms とした．なお，破裂時点以前は $\gamma = 1.0$ とし，第1ホルマント周波数のみを推定した．

図5．1 8より，本手法によれば，/ba/の $t = 20$ms 付近のように通常の線形予測分析では分析窓長を 10ms とすると第3ホルマントが正しく推定できない場合でも妥当な第3ホルマントが推定でき，また /da/ および /ga/ の結果から明らかなように，有

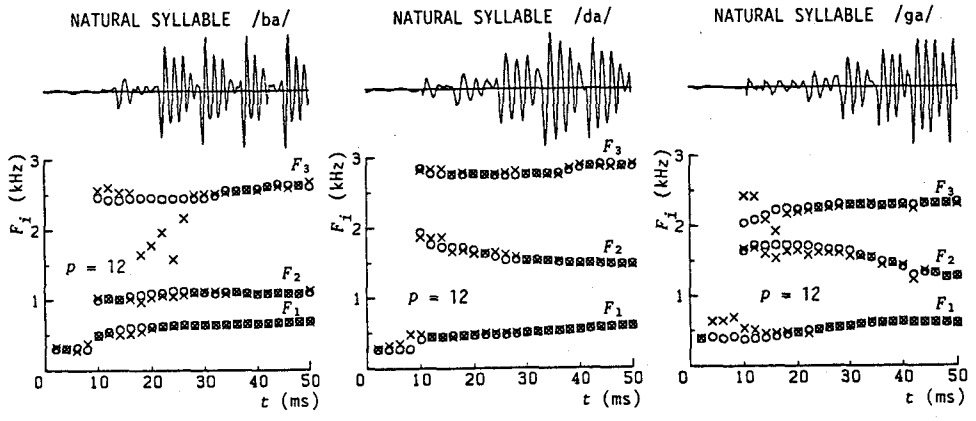


図5．18　ホルマント周波数軌跡推定の比較
－ 自然有声破裂音 /ba/,/da/,/ga/ －
○印：本方法による推定値
×印：通常の線形予測分析 (T_a = 10ms) による推定値

声破裂音の相互識別に重要となるホルマントローカス付近のホルマント軌跡が通常の線形予測分析よりも正確に推定できると言える．

5.4 結　　言

音声の過渡部のホルマント周波数推定手法として，有声音の１ピッチ周期未満の特定区間の音声波形の周波数分析をＦＦＴにより行うＦＲＡＰＳ法，なびに分析窓の任意の点を固定して窓長を漸減させた一連の線形予測分析の結果から，分析窓長が零になる場合の値を外挿する窓長漸減型線形予測分析について述べた．

短時間周波数スペクトルに及ぼす分析窓の始点および窓長の影響を詳細に検討した結果，分析窓長を１ピッチ周波数未満と短くし，かつ，窓の位置を声帯音源の１ピッチ周期中の実効的な励振点を包含しないように設定するならば声道伝達特性を保存した平滑な周波数スペクトル包絡が得られ，そのピーク周波数から精度よくホルマント周波数推定が行えることが明らかとなった．また，窓長漸減型線形予測分析の基本型である分析窓の始点を固定した始点固定型ならびに分析窓の中心を固定した中心固定型について詳細な検討を行った結果，分析窓の始点あるいは中心を固定して窓長を漸

減させた一連の線形予測分析を行って得られる極周波数推定値は音声の過渡部においては分析窓長と共にそれぞれほぼ直線的あるいは2次曲線的に変化することが過渡モデル音による解析ならびに合成音によるシミュレーションにより明らかとなった．そして，ＦＲＡＰＳ法及び窓長漸減型線形予測分析とも，特に，音声の過渡部のホルマント周波数推定に有効な分析手法であることを合成有声破裂音ならびに実際の有声破裂音のホルマント周波数軌跡推定に適用することにより示した．

ＦＲＡＰＳ法を自然音声に適用する際には，声帯音源の1ピッチ周期中の実効的な励振点を検出する必要がある．ここでは音声波の1ピッチ周期中の最大振幅付近としたが，より正確にはエポック抽出法[64]などの適用が考えられるが，その必要性，ならびに窓長漸減型線形予測分析では分析窓長をピッチ周期とは独立に漸減させたが，漸減する各窓長をピッチ周期の整数倍とした場合の検討，また外挿のための特徴量として，線形予測分析により得られる極周波数を用いたが，線形予測係数あるいはＬＰＣケプストラム係数を用いた場合との比較等が今後の課題と言える．なお，ＦＲＡＰＳ法は分析窓の位置を声帯音源の1ピッチ周期中の実効的な励振点を包含しないように設定する必要があるため，音声の過渡部の任意の時点のホルマント周波数推定を行なうことができないが，ＦＦＴをその基本としているため，ノンパラメトリックな分析手法であるのに対し，窓長漸減型線形予測分析は分析窓長を極端に短くすることなく，分析窓長を零にした場合の値が推定できるため，音声の過渡部の任意の時点のホルマント周波数推定ができるが，通常の線形予測分析をその基本としているため，パラメトリックな分析手法であると言え，この両分析手法の特質を考慮して音声の過渡部の分析に適用する必要があるが，この適用基準が今後の課題と言える．

第6章　結　　論

本論文では，音声の伝送・認識において重要となる声道伝達特性の極周波数であるホルマント周波数の精密な推定が通常の分析手法では困難な，(1) 零点のある音声，(2) 高ピッチ音声，および (3) 音声の過渡部に対するそれぞれ有効な分析手法を提案した．これをまとめると次のようになる．

対　象	提案した方法
零点のある音声	変形共分散行列の固有値に基づく次数推定 自己相関行列の近似再構成による分析
高ピッチ音声	2段標本選択線形予測分析
音声の過渡部	1ピッチ周期内周波数分析 窓長漸減型線形予測分析

以下，本研究おいて得られた成果を順に述べる．

零点のある音声の精密な分析手法として，

(1) 変形共分散行列の固有値に基づくＡＲ部の次数推定

(2) 自己相関行列の近似再構成によるホルマント周波数推定

を提案した．

変形共分散行列の固有値に基づくＡＲ部の次数推定では，分析区間を有声音の声門閉止区間にとり，式(3．2)で定義する変形共分散行列 $\Phi^{(i_0)}$ のパラメータ i_0 をＭＡ部の次数より大きく設定すれば，$\Phi^{(i_0)}$ の固有値を大きさの順に並べた場合の隣り合う二つの固有値の比より，零点を考慮した極零型モデルに基づく分析（ＡＲＭＡモデル分析）において重要となるＡＲ部の次数を推定できることを解析的に示した．従来の次数推定法の多くは残差に対する評価基準最適化の条件より，次数推定を行っていたので，予測されるＡＲ部ならびにＭＡ部の全組み合わせに対して残差を評価する必要があった．本次数推定法の特長は残差を求める必要がなく，かつＭＡ部とはほぼ独立にＡＲ部の次数が推定できることである．このことを，零点のある合成音ならびに

零点のある音声の代表例と言える実際の鼻子音のＡＲ部の次数推定に本手法を適用することにより示した．

自己相関行列の近似再構成によるホルマント周波数推定では，音声波の自己相関行列に巡回性を導入すれば，その固有値 λ_k は音声波のパワースペクトルにおける第 k/MT(但し，M：自己相関行列の次数，T：標本化周期）成分に対応することを示し，この固有値の包絡上で極大となる固有値およびそれに対応する固有ベクトルのみから行列のスペクトル分解に基づき自己相関行列を近似再構成すれば，極情報のみを担った自己相関係数が得られ，この自己相関係数を用いることにより，声道伝達特性に零点が存在する音声においても極零型モデルに基づく分析を行なうことなく，全極型モデルに基づく分析でホルマント周波数を精密に推定できることを示した．零点が存在する音声に対する分析方法としては，基本的には零点も考慮した極零型モデルに基づく方法が合理的であると言える．しかしながら，極零型モデルに基づく分析方法をとる場合には，系への入力信号に関する情報が必要であり，音声では系への入力信号である励振波形が観測できないので，入力の推定問題ならびに極及び零部の次数設定が不適当な場合には分析精度がむしろ悪くなるといった問題がある．また，極零型モデル分析の特長は極のみでなく零点の推定ができる点にあると言えるが，音声認識あるいは音声分析・合成系における零点パラメータの有用性に関してはまだ検討段階であり，また聴覚的にも零点の知覚は大変鈍いことを考慮すれば，零点が存在する場合でも極零型モデル分析を行うことなく正確な極情報が推定できる分析方法は実用的には重要であると言え，本分析手法はその一つとなるものである．この手法の有用性を明らかにするために，合成音のホルマント周波数推定精度ならびに自然音声の鼻音化母音の分析例及び認識率に関して，本方法と従来の全極型モデルに基づく線形予測分析との比較を行った．その結果，250 通りの合成音において第１～第３ホルマント周波数推定誤差の平均値が零点のない場合はいずれも 1.0% であるが，零点のある場合は 3.5% から 1.8%に改善され，自然音声の鼻音化母音においては周波数スペクトルからから推察されるホルマントの位置に明確なローカルピークを持ったスペクトル包絡が得られた．また，第１～第３ホルマント周波数推定値を特徴パラメータとした場合，鼻音化母音の認識率が 94% から 99% に改善され，本方法の有効性が明らかになった．

なお，本方法を自然音声に適用するにあたり，本方法のパラメータである閾値θを合成音のシミュレーション結果に基づいて固定としたが，自然音声の多様性を考慮すれば閾値θは適応的に変化させる必要があると言える．この問題は今後の課題である．

高ピッチ音声に対する精密な分析手法として，線形予測分析における残差情報の大局的な特徴を考慮して線形予測モデルに適合する音声標本の選択を行い，かつこの処理を2段階行う2段標本選択線形予測分析について述べた．

通常の線形予測分析を用いて正確なホルマント周波数が推定できるためには，励振源が白色ガウス過程であるという条件が満たされることが必要であり，したがって特に女性あるいは子供が発声したピッチ周期の短い高ピッチ音声の精密な分析が通常の線形予測分析ではしばしば困難となることを示した．そして，本方法と従来の標本選択線形予測分析および通常の線形予測分析のホルマント周波数推定精度の比較を合成音により行った結果，ピッチ周期3.8msの合成5母音の第1～第3ホルマント周波数推定誤差の平均値が通常の線形予測分析では励振の影響により5.3%と大きかったものが，従来の標本選択線形予測分析により2.4%程度に改善し，さらに本方法により0.9%と大幅に改善された．そして，実際に成人女性が発声した単音節の母音定常部の分析に適用した結果，本方法によってより妥当な周波数スペクトル包絡が得られ，第1および第2ホルマント周波数空間における式(4．9)で定義する5母音の分離度が通常の線形予測分析では2.2と小さかったものが，従来の標本選択線形予測分析により7.8に改善し，本方法により8.9とさらに改善されると共に，最適な閾値θの範囲が従来の標本選択線形予測分析では$0.5 \leqq \theta \leqq 0.6$と比較的狭かったのが，本手法では$0.4 \leqq \theta \leqq 0.7$と広範囲に改善された．また成人女性が発声した連続音声のホルマント周波数追尾に適用した結果，通常の線形予測分析ではホルマント周波数の時間的変化に不自然な不連続が生じていたのが，本手法ではこれらの不連続が大幅に改善されるとの結果が得られ，本方法の有効性が明らかになった．なお，本方法を自然音声に適用するにあたり，本方法のパラメータである除去標本点数N_0を固定としたが，N_0は声門開口区間の音声標本をできるだけ被予測標本から除くために導入したパラメータであるので，N_0の最適値は声門開口区間すなわちピッチ周期に依存する量であると言える．したがって，特に連続音声に適用する場合には，N_0はピッチ周期に応

じて適応的に変化させることが望ましいと言えるが，この点に関しては今後の課題である．

音声の過渡部の精密な分析手法として，

(1) 1 ピッチ周期内周波数分析によるホルマント周波数推定

(2) 窓長漸減型線形予測分析によるホルマント周波数推定

を提案した．

1 ピッチ周期内周波数分析によるホルマント周波数推定では，音声の過渡部の精密な分析を行なうためには，基本的には分析窓長を十分短くすれば良いとの観点から，短時間周波数スペクトルに及ぼす分析窓の位置および窓長の影響の詳細な検討により，分析窓長を 1 ピッチ周期未満とし，声道が声帯波により実効的に励振された時点を包含しないように分析窓の位置を設定すれば，声道伝達特性を保存した平滑な周波数スペクトル包絡を持つ周波数スペクトルが得られ，そのピーク周波数（周波数スペクトル上で極大となる周波数）からホルマント周波数が高精度で推定できることを示した．そして音声の過渡部の代表例といえる合成ならびに自然有声破裂音のホルマント周波数追尾に本手法と通常の周波数スペクトル包絡法とを適用して比較した結果，合成有声破裂音 /ba/, /da/, /ga/ のホルマントローカス点における第 1 ～第 3 ホルマント周波数推定誤差の平均値が通常の手法では 15.1% と大きかったものが，本方法により 1.0% と大幅に改善され，自然有声破裂音のホルマント周波数追尾においては，通常の手法ではホルマント周波数の時間的変化に不自然な不連続が生じ，かつ周波数スペクトル上にホルマントと同定されかねない優勢なピークが存在していたのが，本手法ではこれらの不連続が大幅に改善され，かつ声道伝達特性を保存した平滑な周波数スペクトル包絡を持つ周波数スペクトルが得られるとの結果が得られ，本方法の有効性が明らかになった．なお，本方法を自然音声に適用する際には声帯音源の 1 ピッチ周期中の実効的な励振点を検出する必要があるが，ここでは音声波の 1 ピッチ周期中の最大振幅付近とした．より正確にはエポック抽出法などの適用が考えられるが，この問題は今後の課題である．

窓長漸減型線形予測分析によるホルマント周波数推定では，音声の過渡部における通常の線形予測分析によるホルマント周波数推定値の過渡モデル音による解析結果に

基づき，分析窓の任意の点を固定して窓長を漸減させた一連の分析の結果から，窓長が零になる場合の値を外挿することにより，分析窓長を極端に短くすることなく，音声の過渡部の任意の時点のホルマント周波数を精度よく推定できることを示した．音声の過渡部の精密な分析を行なうためには，基本的には分析窓長を十分短くすれば良いと言えるが，通常の線形予測分析において分析窓長を1ピッチ周期程度以下に短くすると分析窓と励振点との相対位置の影響が生じるため，音声の任意の時点を分析することができない．また分析窓内での非定常性を考慮した線形予測分析も検討されてはいるが，一般に非定常性を考慮した分析手法は処理手順が複雑になるといった問題がある．したがって，非定常性を考慮した分析を行なうことなく，音声の過渡部の任意の時点のホルマント周波数を精度よく推定できる分析手法は実用的には重要であると言える．本分析手法はその一つとなるものである．このことを明らかにするために，音声の過渡部の代表例といえる合成ならびに自然有声破裂音のホルマント周波数追尾に適用した結果，合成有声破裂音の破裂時点から破裂時点後10msまでの区間を2ms間隔毎に分析した計6フレームの第1～第3ホルマント周波数推定誤差の平均値が，通常の線形予測分析において分析窓長を10msとした場合，/ba/，/da/，/ga/それぞれにおいて，69.4Hz，53.2Hz，62.5Hzと大きかったものが，本方法により24.7 Hz，14.5Hz，24.9Hzと大幅に改善され，また自然有声破裂音のホルマント周波数追尾においても，通常の線形予測分析ではホルマント周波数の時間的変化に不自然な不連続が生じていたのが，本手法ではこれらの不連続が大幅に改善され，本手法の有効性が示された．なお，窓長はピッチ周期と独立に漸減させたが，漸減する各窓長をピッチ周期の整数倍とした場合の検討，また外挿のための特徴量として，線形予測分析により得られる極周波数を用いたが，線形予測係数あるいはＬＰＣケプストラム係数を用いた場合との比較等が今後の課題と言える．

ところで，音声の過渡部の精密な分析手法として，1ピッチ周期内周波数分析と窓長漸減型線形予測分析の二つの手法を提案したが，1ピッチ周期内周波数分析は分析窓の位置を声帯音源の1ピッチ周期中の実効的な励振点を包含しないように設定する必要があるため，音声の過渡部の任意の時点のホルマント周波数を推定することができないが，ＦＦＴをその基本としているため，ノンパラメトリックな分析手法である

といえる．一方，窓長漸減型線形予測分析は分析窓長を極端に短くすることなく，分析窓長を零にした場合の値が推定できるため，音声の過渡部の任意の時点のホルマント周波数を推定できる特長があるが，通常の線形予測分析をその基本としているため，パラメトリックな分析手法であるといえる．この両分析手法の特質を考慮して音声の過渡部の分析に適用する必要があるが，この適用基準が今後の課題である．

以上，本論文では，従来の分析手法では精密な分析が困難であった種々の音声に対するそれぞれ有効な分析手法を示した．本論文で提案した分析手法はいずれも分析対象となる音声を限定するものではない．すなわち本論文で提案した分析手法を，通常の線形予測分析等で精度よく分析できる音声あるいは音声区間に適用してもなんら問題はない．ところで，音声認識において，音響分析部での特徴パラメータ抽出と言語処理部での意味理解は，相互に相補ってこそ高性能な音声認識システムが実現できるといえる．すなわち分析部における分析精度の向上が言語処理部での処理の負担を軽減するが，一方，高精度の分析にはそれ相応の処理時間を要する．したがって，通常は従来の分析手法で特徴パラメータを抽出し，言語処理部でより高精度の特徴パラメータ抽出の必要性が生じた時点で本論文で提案した分析手法を適用すればよく，今後，音響分析部と言語処理部相互の有機的な結合形態の研究が重要となって来ると思われる．

謝　辞

筆者が姫路工業大学の長期研修制度により，昭和４８～４９年の間大阪大学産業科学研究所にて研修する機会を得て以来，本研究の全過程を通じ，直接懇切なる御指導，御鞭達を賜わった大阪大学産業科学研究所角所収教授に衷心より感謝の意を表する．

筆者の研修当初から昨年まで角所研究室に在籍されていた郵政省通信総合研究所音声研究室柳田益造室長には，本研究の端緒より，終始有益な御助言と御教示を賜わった．ここに衷心より感謝する次第である．

本研究をまとめるに当たり，貴重な御助言と御教示を賜わった大阪大学工学部電子工学科児玉慎三教授，寺田浩詔教授，白川功教授，ならびに貴重な御助言と御指摘を賜わった大阪大学工学部電子工学科西原浩教授，浜口智尋教授，吉野勝美教授，大阪大学電子ビーム研究施設裏克己教授，塙輝雄教授に厚く御礼申し上げる．

本研究の遂行に当たり，終始適切な御助言と御鞭達を頂いた姫路工業大学電子工学科大和一晴教授，いつも有益な御助言を頂いた角所研究室の溝口理一郎助教授，ならびに御討論頂いた山口高平助手，山下洋一技官はじめ角所研究室の諸氏に深く感謝する．そして，角所研究室に在籍されていた広島大学翁長健治教授には有益な御助言と御鞭達を頂き，同じく大阪大学言語文化部平藤暢夫助手，ならびに摂南大学中嶋鴻毅講師には多大な御援助を頂いた．ここに記して深く感謝する次第である．

筆者の研究活動に対し，姫路工業大学電子工学科の諸先生方には多大の御便宜と御支援を賜わった．ここに厚く御礼申し上げる次第である．

文　　献

(1) J.L.Flanagan:"Speech analysis synthesis and perception", 2nd Edition, Springer-Verlag(1972).

(2) H.Dudley:"Remaking speech", J.Acoust.Soc.Am., **11**, pp.169−177(1939).

(3) R.K.Potter, et al.:"Visible speech", D. Van Nostrand Co., Inc., New York(1947).

(4) K.N.Stevens, A.S.House:"Development of a quantitative description of vowel articulation", J.Acoust.Soc.Am., **27**, pp.484−493(1955).

(5) G.Fant:"Acoustic theory of speech production", Mouton(1960).

(6) J.L.Flanagan:"Automatic extraction of formant frequencies from continuous speech", J.Acoust.Soc.Am., **28**, pp.110−118(1956).

(7) C.G.Bell, H.Fujisaki, et al.:"Reduction of speech spectra by Analysis-by-Synthesis techniques", J.Acoust.Soc.Am., **33**, pp.1729−1736(1961).

(8) M.V. Mathews, J.E. Miller, E.E. David:"Pitch synchronous analysis of voiced sounds", J.Acoust.Soc.Am., **33**, pp.179−186(1961).

(9) E.N.Pinson:"Pitch-synchronous time-domain estimation of formant frequencies and bandwidths", J.Acoust.Soc.Am., **35**, pp.1264−1273(1963).

(10) 鈴木, 角川, 中田："モーメント法によるホルマント周波数の抽出", 日本音響学会誌, **19**, pp.106−114(1963).

(11) K.H.Davis, R.Biddulph, S.Balashek:"Automatic recognition of spoken digits", J.Acoust.Soc.Am., **24**, pp.637−642(1952).

(12) F.H.Olson, H.Belar:"Phonetic typewriter", J.Acoust.Soc.Am., **28**, pp.1072−1081(1956).

(13) D.B.Fry:"Theoretical aspects of mechanical speech recognition", J.Brit.IRE, **19**, pp.211−218(1959).

(14) P.Denes, M.V.Mathews:"Spoken digit recognition using time-frequency pattern maching", J.Acoust.Soc.Am., **32**, pp.1450−1455(1960).

(15) 鈴木, 大泉 : "日本語母音の 2 進符号化および認識の学習", 信学誌, **46**, pp.291−299(1963).

(16) 鈴木, 中田 : "数字語識別の実験", 信学誌, **45**, pp.303−309(1963).

(17) 中田 : "音声の認識", 信学誌, **46**, pp.1600−1608(1963).

(18) 坂井, 堂下 : "会話音声識別装置", 信学誌, **46**, pp.1696−1702(1963).

(19) 加藤, 千葉, 永田 : "数字音声識別装置", 信学誌, **47**, pp.1319−1325(1964).

(20) J.W.Cooley, J.W.Tukey:"An algorithm for the machine calculation of complex fourier series", Mathematics of Computation, **19**, pp.297−301(1965).

(21) 板倉, 斉藤 : "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**, pp.35−42(1970).

(22) B.S.Atal, S.L.Hanauer : "Speech analysis and synthesis by linear prediction of the speech", J.Acoust.Soc.Amer., **50**, pp.637−655(1971).

(23) 迫江, 千葉 : "動的計画法を利用した音声の時間正規化に基づく連続単語認識", 日本音響学会誌, **27**, pp.483−500(1971).

(24) T.B.Martin:"Practical application of voice input to machines", Proc. IEEE, **64**, pp.487−501(1976).

(25) H.Sakoe:"Two-level DP-matching − A dynamic programing based pattern matching algorithm for connected word recognition", IEEE Trans., Acoust., Speech & Signal Process., **ASSP-27**, pp.588−595(1979).

(26) C.S.Myers, L.R.Rabiner:"Connected digit recognition using a level-building DTW algorithm", IEEE Trans., Acoust.,Speech & Signal Process., **ASSP-29**, pp.351−363(1981).

(27) 中川:"パターンマッチング法による連続単語および連続音節の音声認識アルゴリズム", 信学論(D), **J66-D**, pp.637−644(1983).

(28) B.Aldefeld, et al.:"Automated directory listing retrieval system based on isolated word recognition", Proc. IEEE, **68**, pp.1364−1379(1980).

(29) 古井:"単音節認識とその大語い単語音声認識への適用", 信学論(A), **J65-A**, pp.175−182(1982).

(30) 管村, 古井:"疑音韻標準パタンによる大語い単語音声認識", 信学論(D), **J65-D**, pp.1041−1048(1982).

(31) 大泉充郎, 藤村 靖:"音声科学", 東京大学出版会(1972).

(32) 中田和男:"音声", コロナ社(1977).

(33) 斉藤収三, 中田和男:"音声情報処理の基礎", オーム社(1981).

(34) 古井貞熙:"ディジタル音声処理", 東海大学出版会(1985).

(35) Y. Miyoshi, K. Yamato and O. Kakusho: "Order setimation of AR model based on eigenvalues of covariance matrix of speech", 10th International Congress on Acoustics, A1-10.5(1980).

(36) Y. Miyoshi, K. Yamato and O. Kakusho : "Order estimation of speech production model based on the eigenvalue ratios of quasi-covariance matrix", J. Acoust. Soc. Jpn.(E), **4**, pp.45−47(1983).

(37) 石崎:"音声分析における極零モデルの次数の同定", 信学論(A), **J60-A**, pp.423−424(1977).

(38) Y.Monden, M.Yoshida, S.Arimoto :"Fast algorithm for identification of an ARX model and its order estimation", IEEE Trans., Acoust.,Speech & Signal Process., **ASSP-30**, pp.390−399(1982).

(39) 森川, 藤崎："ＳＥＡＲＭＡ法による音声分析における次数の適応的・連続的推定", 日本音響学会音声研究会資料, **S82**-50(1982).

(40) A.E.Rosenberg :"Effect of glottal pulse shape on the quality of natural vowels", J.Acoust.Soc.Am., **49**, pp.583−590(1971).

(41) 三好, 大和, 柳田, 角所："自己相関行列の近似再構成による極周波数の精密推定", 信学論（Ａ）, **J68-A**, pp.1389−1397(1985).

(42) 深林, 鈴木："極−零形の線形モデルによる音声分析", 信学論（Ａ）, **J58-A**, pp.270−277(1975).

(43) 森川, 藤崎："ＡＲ・ＭＡパラメータの同時推定法による音声分析", 信学論（Ａ）, **J61-A**, pp.195−202(1978).

(44) 宮永, 三木, 永井, 羽鳥："時変ＡＲＭＡパラメータの適応的同時推定", 信学論（Ｄ）, **J64-D**, pp.308−315(1981).

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

(46) M.Morf, D.T.Lee, J.R.Nickolls, A.Vieira:"A classification of algorithm for ARMA models and ladder realizations", Proc.IEEE Intern. Conf. Acoust, Speech and Signal Processing, Hartford, CT, pp.13−19(1977).

(47) 深林, 鈴木："極−零形モデルと全極形モデルによる音声分析結果の比較", 信学論（Ａ）, **J59-A**, pp.855−862(1976).

(48) 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）, **J65-A**, pp.454−461(1982).

(49) 森川，藤崎："ＳＥＡＲＭＡ法に基づく音声分析合成系"，日本音響学会音声研究会資料，**S79**-47(1979).

(50) 溝口，田中，福田，辻野，角所："連続音声認識エキスパートシステム－ＳＰＲＥＸ－"，信学論（Ｄ），**J70-D**, pp.1189−1198(1987).

(51) 千葉："連続音声中の鼻音区間検出法の検討"，日本音響学会音声研究会資料，**S83**-39(1983).

(52) 北澤，横井，堂下："後続母音に独立な特徴による鼻子音の識別"，日本音響学会春季講演論文集，2−3−15(1984).

(53) G.M.Jenkins, D.G.Watts:"Spectral Analysis and its Applications", Holden-Day(1968).

(54) 藤崎，佐藤："音声のホルマント抽出の諸方式の比較検討"，日本音響学会音声研究会資料，**S74**-1(1974).

(55) R.W.Schafer, L.R.Rabiner :"System for Automatic Formant Analysis of Voiced Speech", J.Acoust.Soc.Am., **47**, pp.634−678(1970).

(56) Y. Miyoshi, K. Yamato, M. Yanagida and O. Kakusho: "Analysis of speech signals of short pitch period by the sample-selective linear prediction", Proceeding of International Conference on Acoustics,Speech, and Signal Processing, pp.1245−1248(1986).

(57) Y. Miyoshi, K. Yamato, R. Mizoguchi, M. Yanagida and O. Kakusho : "Analysis of speech signals of short pitch period by a sample-selective linear prediction", IEEE Trans. Acoust., Speech & Signal Processing, **ASSP-35**, pp.1233−1240(1987).

(58) 三好，大和，柳田，角所："２段標本選択線形予測法による高ピッチ音声の分析"，信学論（Ａ），**J70-A**, pp.1146−1156(1987).

(59) M.Ljungqvist, 藤崎:"線形予測分析にもとづく声帯音源・声道パラメータの同時推定法", 音響学会音声研資, **S85**-21(1985).

(60) S.Chandra, W.C.Lin : "Experimental comparison between stationary and nonstationary formulations of linear prediction applied to voiced speech analysis", IEEE Trans., Acoust.,Speech & Signal Process., **ASSP-22**, pp.403−415(1974).

(61) 河原, 栃内, 永田:"小区間の線形予測分析とその誤差評価", 日本音響学会誌, **33**, pp.470−479(1977).

(62) K.Steiglitz, B.Dickinson : "The use of time-domain selection for improved linear prediction", IEEE Trans., Acoust.,Speech & Signal Process., **ASSP-25**, pp.34−39(1977).

(63) H.W.Strube : "Determination of the instant of glottal closure from the speech wave", J.Acoust.Soc.Am., **56**, pp.1625−1629(1974).

(64) T.V.Ananthapadmanbha, B.Yegnanarayama : "Epoch extraction of voiced speech", IEEE Trans., Acoust.,Speech & Signal Process., **ASSP-23**, pp.562−570(1975).

(65) 溝口, 柳田, 谷口, 角所:"一般逆行列を用いた音声の選択的線形予測分析", 信学論(A), **J66-A**, pp.56−63(1983).

(66) 三好, 大和, 角所:"有声音の1ピッチ周期内周波数分析によるホルマント周波数抽出", 信学論(A), **J61-A**, pp.633−640(1978).

(67) R.W.Schafer, L.R.Rabiner :"System for automatic formant analysis of voiced speech", J.Acoust.Soc.Am., **47**, pp.634−648(1970).

(68) 藤崎, 吉宗:"準周期的波形の短時間周波数スペクトル推定について", 日本音響学会春季講演論文集, 1−1−11(1971).

(69) W.J.Hess : "A pitch-synchronous digital feature extraction system for phonemic recognition of speech", IEEE Trans. Acoust.,Speech & Signal Process., **ASSP-24**, pp.14−25(1976).

(70) R.B.Blackman, J.W.Tukey : "The measurement of power spectra", Dover, New York(1959).

(71) 三好, 大和, 角所:"ＦＲＡＰＳ法によるフォルマント周波数の抽出", 信学技報, **EA74**-47(1975).

(72) 三好, 大和, 柳田, 角所:"窓長漸減型線形予測分析による過渡的音声のホルマント周波数抽出", 信学論（A),（1988 年 10 月掲載予定）.

(73) 藤崎, 佐藤:"各種ホルマン周波数抽出方式における短区間分析の時間窓の影響", 音響学会春季講演論文集, 2−2−2(1974).

(74) 三好, 大和, 角所:"線形予測法による有声音の 1 ピッチ周期内分析", 信学技報, **EA76**-53(1977).

(75) 片桐, 松井, 牧野, 城戸:"高ピッチ音声に対する短区間線形予測分析の検討", 信学技報, **EA80**-31(1980).

(76) 深林:"線形予測法による音声分析の精度向上", 信学論（Ａ）, **J61-A**, pp.1168−1169(1978).

(77) 柳田, 角所:"重み付き線形予測分析の検討", 音響学会音声研資, **S85**-08(1985).

(78) 中島, 鈴木:"非定常態音声分析法", 音響学会春季講演論文集, 2−7−2(1980).

(79) Y.Grenier : "Time-dependent ARMA modeling of nonstationary signals", IEEE Trans. Acoust.,Speech & Signal Process., **ASSP-31**, pp.899−911(1983).

(80) 芹沢, 三木, 宮永, 永井:"時変ＡＲＭＡモデルに基づく適応的音声分析法", 信学論（Ａ）, **J71-A**, pp.434−442(1988).

付　　録

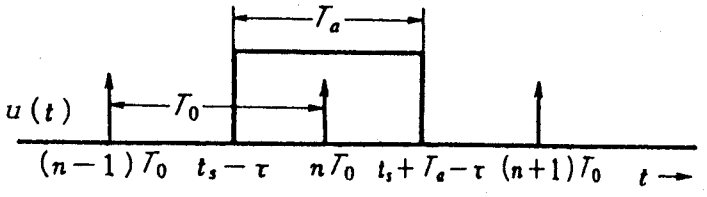
A.　式（5.6）の導出

$u(t)$ が周期 T_0 のインパルス列 $\sum_{n=-\infty}^{\infty}\delta(t-nT_0)$ で分析窓長 T_a が $T_a < T_0$ の場合，式(5．4)の $\int_{t_s-\tau}^{t_s+T_a-\tau} u(t)e^{-j\omega t}dt$ は積分範囲内にインパルス印加時点を含むとき，すなわち $t_s-\tau \leqq nT_0$，かつ $t_s+T_a-\tau \geqq nT_0$ のときにのみ非零となる（図A・1参照）．一方，$\tau < 0$ では $h(\tau)=0$ であるため $F(\omega, t_s, T_a)$ は $t_s - nT_0 \leqq \tau \leqq t_s + T_a - nT_0$ で，かつ $t_s + T_a - nT_0 \geqq 0$ のときに非零となる．したがって式(5．4)，(5．5)より，

$$F(\omega, t_s, T_a) = \sum_{n=-\infty}^{n_0} \int_{t_s-nT_0}^{t_s+T_a-nT_0} h(\tau)e^{-j\omega\tau} \int_{t_s-\tau}^{t_s+T_a-\tau} \delta(t-nT_0)e^{-j\omega t}dtd\tau$$

$$\text{但し，}\quad n_0 = \left[\frac{t_s+T_a}{T_0}\right],\quad [\ ]:\text{ガウス記号}$$

となる．これより式(5．6)が導かれる．



図A・1　$u(t)$ と積分範囲の関係

B. 式（5.8）の導出

$t_{s0}+T_a < T_0$ のとき，$n_0 = \left[\frac{t_s+T_a}{t_0}\right]$ より $n_0 = m$ となり，式(5．6)は，

$$F(\omega, t_s, T_a) = \sum_{n=-\infty}^{m} e^{-j\omega nT_0} \int_{t_{s0}+(m-n)T_0}^{t_{s0}+(m-n)T_0+T_a} h(\tau)e^{-j\omega\tau}d\tau$$

$$= e^{-j\omega mT_0} \int_{t_{s0}}^{t_{s0}+T_a} h(\tau)e^{-j\omega\tau}d\tau$$

$$+e^{-j\omega(m-1)T_0} \int_{t_{s0}+T_0}^{t_{s0}+T_0+T_a} h(\tau)e^{-j\omega\tau}d\tau$$

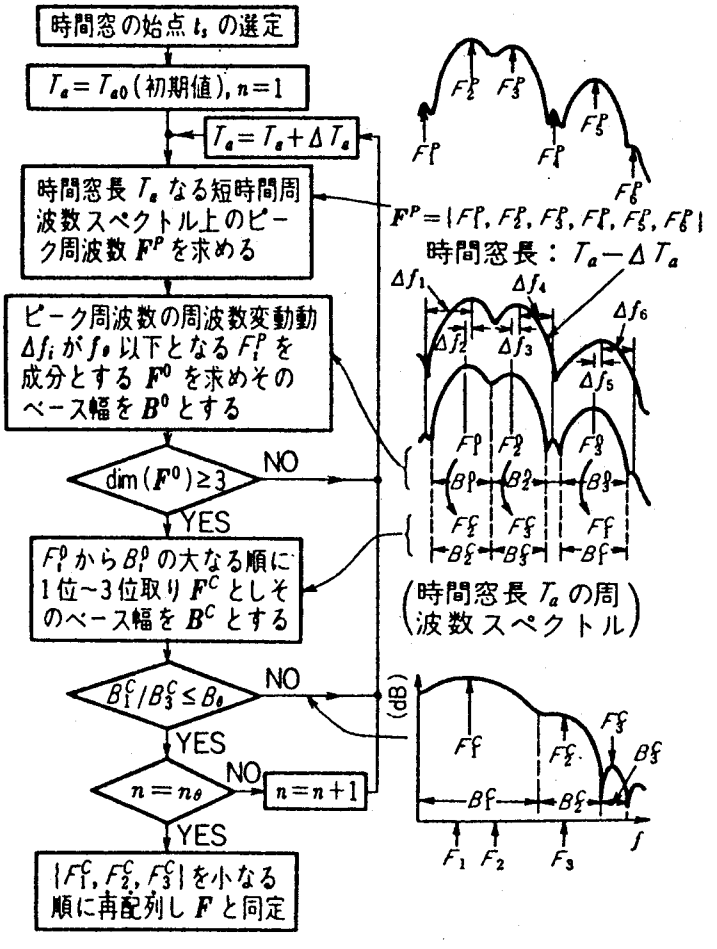
$$+e^{-j\omega(m-2)T_0} \int_{t_{s0}+2T_0}^{t_{s0}+2T_0+T_a} h(\tau)e^{-j\omega\tau}d\tau + \cdots$$

$$= e^{-j\omega mT_0} \sum_{k=0}^{\infty} e^{j\omega kT_0} \int_{t_{s0}+kT_0}^{t_{s0}+kT_0+T_a} h(\tau)e^{-j\omega\tau}d\tau$$

となる．ここで，$t = \tau - kT_0$ と変数変換を行なえば式(5．8)が導かれる．

C. ＦＲＡＰＳ法によるホルマント周波数推定アルゴリズム

本アルゴリズムは５．２．２で述べた「T_a によりその位置が不変なピーク周波数がホルマント周波数に対応するピーク周波数である」という重要な特性に基礎をおいたものである．



f_θ：F_i^P の周波数変動 Δf_i のしきい値
B_θ：ベース幅比 B_1^C/B_3^C のしきい値
n_θ：繰返し回数 n のしきい値

図Ａ・２　ＦＲＡＰＳ法によるホルマント周波数推定アルゴリズム

D. 式（5.20）の導出

$t - t_0 \rightarrow t$ と変数変換を行い $(\Delta A)^2$ の項を省略すれば，

$$\int_{t_0}^{t_0+T_a} s(t)s(t+\tau)dt$$

$$= \frac{1+\Delta A\tau}{2}\left\{\int_0^{T_a}\cos(\Delta\omega t+\Omega)\tau\, dt - \int_0^{T_a}\cos(\Delta\omega t^2+2\Omega t+\Omega\tau)dt\right\}$$

$$+\Delta A\left\{\int_0^{T_a} t\cos((\Delta\omega t+\Omega)\tau\, dt - \int_0^{T_a} t\cos(\Delta\omega t^2+2\Omega t+\Omega\tau)dt\right\}$$

$$= \frac{1}{\Delta\omega\tau}[(1+\Delta A\tau)\sin\theta_1\cos(\theta_1+\Omega\tau)$$

$$+\Delta A\{T_a\sin(2\theta_1+\Omega\tau) - \frac{2}{\Delta\omega\tau}\sin\theta_1\sin(\theta_1+\Omega\tau)$$

$$-\tau\sin(\Delta\omega T_a^2/2+\Omega T_a)\cos(\Delta\omega T_a^2/2+\Omega(T_a+\tau))\}]$$

$$-\frac{1}{2}\sqrt{\frac{\pi}{2\Delta\omega}}(1-\frac{2\Delta A\omega_0}{\Delta\omega})[\{C(x_2)-C(x_1)\}\cos\theta + \{S(x_2)-S(x_1)\}\sin\theta]$$

但し，　$\Omega = \omega_0 + \Delta\omega\tau/2$

$S(\cdot)$ ： 正弦フレネル関数，　$C(\cdot)$ ： 余弦フレネル関数

$$x_1 = \Omega\sqrt{\frac{2}{\Delta\omega\pi}},\quad x_2 = x_1 + \sqrt{\frac{2\Delta\omega}{\pi}}T_a$$

$$\theta_1 = \Delta\omega\tau\, T_a/2$$

$$\theta_2 = \frac{\omega_0^2 - (\Delta\omega\tau/2)^2}{\Delta\omega}$$

となる．ところで，$F_0 = \omega_0/2\pi = 1000$Hz, $\Delta F = \Delta\omega/2\pi = 10$Hz/ms, $T_a = 20$ms, $\tau = 0.1$ms の時，x_1=20.01，x_2=24.01 となるので，フレネル関数の特徴より，$S(x_1) \approx S(x_2) \approx 0.5$，$C(x_1) \approx C(x_2) \approx 0.5$ となり，また，$T_a \gg \tau$ より，上式の下から1行目ならびに2行目を省略し，かつ $1+\Delta A\tau \approx 1$，$\Omega\tau \approx \omega_0\tau$ とすれば，

$$\int_{t_0}^{t_0+T_a} s(t)s(t+\tau)dt$$

$$= \frac{1}{\Delta\omega\tau}\sqrt{(1+\Delta AT_a)^2\sin^2\theta_1 + (\Delta A)^2\{\frac{2}{\Delta\omega\tau}\sin\theta_1 - T_a\cos\theta_1\}^2}$$

$$\cos(\omega_0\tau + \theta_1 + \phi)$$

但し, $\phi = \tan^{-1}\dfrac{2\Delta A}{(1+\Delta AT_a)\Delta\omega\tau}(1-\theta_1\cot\theta_1)$

となる．そして，平方根中の第1項 ≫ 第2項（例，ΔF=10Hz/ms，T_a=20ms，τ=0.1ms，ΔA=0.01/ms の時，第1項 = 5.68×10^{-3}，第2項 = 6.92×10^{-8}）より，

$$\int_{t_0}^{t_0+T_a} s(t)s(t+\tau)dt = \frac{1+\Delta AT_a}{\Delta\omega\tau}\sin(\Delta\omega\tau T_a/2)\cos(\omega_0\tau + \Delta\omega\tau T_a/2 + \phi)$$

となる．同様に

$$\int_{t_0}^{t_0+T_a} s^2(t)dt = \frac{1}{2}(1+\Delta AT_a)T_a$$

となり，これより式(5．2 0)が導かれる．