Vol.15 No.80 電波研究所季報 September 1969
pp. 495—506

研　　究

UDC 534.78

高速ホルマント周波数抽出法と合成音によるその評価

中津井　護*　鈴木誠史*

FAST AND RELIABLE FORMANT FREQUENCY EXTRACTION TECHNIQUE AND ITS EVALUATION BY THE SYNTHESIZED SPEECH

By

Mamoru NAKATSUI and Jouji SUZUKI

New procedure of the formant frequency extraction which utilizes the characteristic features of the vowel-type spectrum is proposed. Constructing this procedure in the form of computer simulated program, the exrtaction experiment was carried out on both artificial (synthesized) and natural speech sounds.

There, in general, are some troublesome situations where the conventional procedure shows many difficulties in the formant frequency extraction, such as, existence of source zeros, spectral variabilities related to the speaker difference, rapid transitions and contiguity of formants, and so on. The artificial speech sounds attempting to realize the various troublesome situations described above and the natural speech sounds composed of the five vowels by five male speakers and of the conversational speech by two male speakers were prepared for the experiment. Using the FFT, they were spectrally analyzed and then fed to the extraction program.

An extraction principle of this procedure is the successive iteration of the following processes. Provided that the vowel-type spectrum is composed of the first three formants and the correcting term, the correcting term is canceled out from the input spectrum. Subtracting two resonant spectra, which are calculated by the two approximated formant frequencies, from it, the remainder is considered as an approximation for the resonant spectrum of a certain formant omitted in above calculation. An approximation for this formant frequency is obtained by calculating the first order moment of the remainder. New approximation is used in the speculation of the other formant frequencies.

It was found that this method shows fairly accurate, high speeded and reliable performance compared with conventional ones, despite of weaving the troublesome situations into the synthetic speech sounds artificially and of the various individualities of total seven different speakers.

* 通信機器部音声研究室

1. はじめに

われわれは，音声を通信の手段として，日常のさまざまな場面で自由に駆使している。この言語情報をになう音声がどのような過程を通じて発声され，聞きとられるのか，また，物理的な現象として，音声波がどのような特性をもっているかなどについては，古くから興味がもたれてきた。1950年代の終りになって，スエーデンの Fant により音声スペクトルの生成機構についての近似理論の基礎が確立され[1]，このころから音声の研究は急に活況を呈してきた。

Fant の理論によれば，音声波のスペクトルは，音源，伝達，および放射の3特性の積で現わされる。さらに，母音型スペクトルの場合には伝達特性の極のうち，低次の数個の極—われわれはこれらをホルマントと呼んでいる—によってほぼ完全にその特質が現わされる。

このホルマント（特に，その周波数）は，音声波のスペクトルの物理的な性質を記述する重要なパラメータとして，分析，合成の両面にかかわらず，音声研究に多く用いられてきた。たとえば，米国の Haskins 研究所で行なわれた，一連の合成音の聞きとり実験[2],[3]は有名であり，また，音声の分析や識別の研究においても，ホルマント周波数は，中心的役割をはたしてきた[4]。

最近，めざましく発展している計算機とその利用技術にささえられて，ホルマント周波数の自動抽出が，多く試みられてきた。その代表例は，peak-picking 法[5]と，Analysis-by-Synthesis (A-b-S) 法[6]であろう。前者は，音声スペクトルの局所ピークの位置からホルマント周波数を求めようとするもので，その抽出過程は簡単であるが，抽出精度は良くない。一方，A-b-S 法は，先にふれた音声スペクトルの生成機構についての理論を，積極的に利用した方法*で，抽出精度が非常に良いものと考えられている。しかし，これは抽出過程が非常に複雑な構成となっており手軽に用いることはできない。

当所の音声研究グループによって開発されたモーメント法[8]は，これら両者の中間的なもので，音声の分析，識別に多くの役割をはたしてきた[4],[9]。しかし，抽出対象を小数の個人に限らずに広げたとき，安定した抽出のできない場合がかなりみられた。

これらの各方式は，音声スペクトル構造の情報およびその特徴の利用の方法が，積極的であるか消極的であるか，また全体的であるか局所的であるか，などの観点で分類できる。さらに，その方法がそのまま各方式の長所とも短所ともなっている。

筆者らの考えでは，いままでに提案されてきたホルマント周波数の自動抽出法を実用的な立場からみると，抽出精度と抽出の速さおよび簡便さとのかねあいで，じゅうぶん実用に耐え得る方式はみあたらないようである。そこで

(1) 母音型スペクトルの特徴的な構造をうまく利用すること。

(2) 抽出過程が簡便かつ高速で，かなりよい抽出精度であること。

(3) 抽出対象の個人差によらずに，安定な抽出ができることを目的として，3*FE* (Triple-*FE*; *F*ast and Reliable *F*ormant *F*requency *E*xtraction) と名づけたホルマント周波数抽出法を提案する。これを FORTRAN 言語によるプログラムとして構成し，抽出実験を行なった。なおこの抽出法の原理は先に報告した "消去法"[10] を発展させたものである。

ここで報告する抽出実験の特色としては，次の2点をあげることができる。

(1) 音声のスペクトル分析に高速フーリェ変換（FFT）[11],[12] を用いたこと。

(2) 諸元の明らかな合成音を用意して本方式の抽出能力の評価を行なったこと。

とくに (2)については，一般にホルマント周波数の抽出が困難と思われる悪条件，たとえばホルマント相互の近接や急激な変化，音源スペクトルの零点の存在などを考慮した合成音を準備した。

これらの合成音と，単独母音および連続音声中の母音部の計7名分（男性）の自然音とを用いて計算機シミュレーションによる抽出実験を行なった。その結果

(1) 抽出精度の目安としては，音源基本周波数 (F_0) の半分 ($F_0/2$) 以内であること。

(2) 抽出に要する時間は，一つの短時間スペクトルあたり 0.25 秒であること（時間率にして25倍）。

(3) 個人差や先にふれた悪条件によらずに安定した抽出ができること

などが確められ，所期の目的をじゅうぶん満足していることがわかった。なお，女声の場合や，び音などの伝達系に零点が含まれている場合などの取扱いは今後の課題である。

2. 3*FE* 法の発想と抽出過程の構成

2.1. 母音型スペクトルの構造とホルマント

Fant の成生理論にしたがえば，母音型音声波のスペ

* 提案者の K. N. Stevens らによれば，この手法はホルマント周波数抽出法としてだけでなく，音声認識の一般的なモデルとしても考えられている[7]。

クトル $P(s)$ はラプラス変換表示で，次のような3要素の積で表わされる。

$$P(s)=Q(s)\cdot T(s)\cdot R(s) \tag{1}$$

ここで，$Q(s)$ と $R(s)$ はそれぞれ声帯振動に基づく音源特性と口からの放射の特性であり，$T(s)$ は声門から口に至る声道の伝達特性である。さらに，これらのうち母音としての韻質を規定するものは $T(s)$ であり，$Q(s)$ や $R(s)$ はそれに比べて比較的一定した項として取扱われる。声道の伝達特性は，母音型の場合には次のように極の無限積で表わすことができる。

$$T(s)=\prod_{i=1}^{\infty}\frac{s_i\ s_i^*}{(s-s_i)(s-s_i^*)} \tag{2}$$

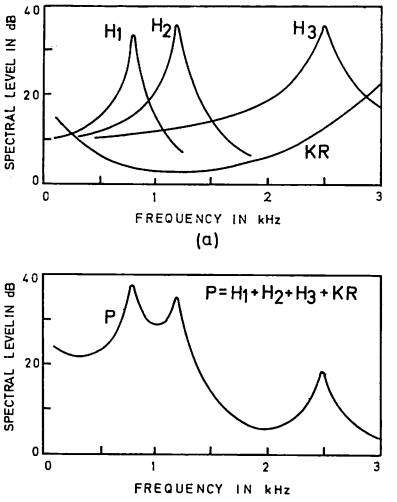
ここで，$s_i=\sigma_i+j\omega_i$ で s_i^* は s_i の複素共役（$s_i^*=\sigma_i-j\omega_i$）である。

いま対象とする周波数範囲を 3kHz 以下に限れば，母音としてのスペクトルの特質は $T(s)$ の低次の数個の極（ホルマント）で決まり，それ以上の高次の極はまとめて補正項として扱われる。そこで，低次の極として3個をとり，われわれの扱いやすい振幅スペクトル表示として(1)式を書きかえると，

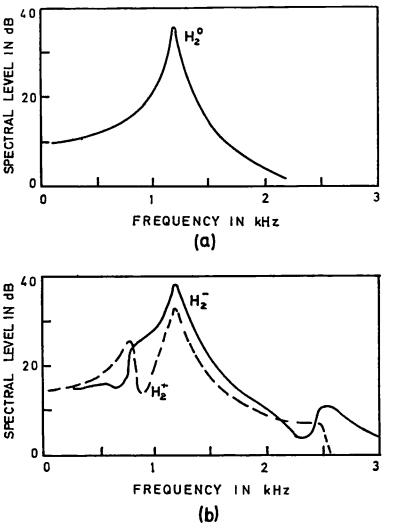
$$P(f)=\prod_{i=1}^{3}Hi(f)\cdot KR(f) \tag{3a}$$

$$KR(f)=Q(f)\cdot HP(f)\cdot R(f) \tag{3b}$$

となる。ここで，$H_i(f)$ は i 番目の極の特性，つまり第 i ホルマントの単共振特性であり，$HP(f)$ は高次ホルマントの補正項である。先に述べたように，母音の種類によらず比較的に一定と考えられる項を $KR(f)$ としてまとめてある*。これらの各項の近似式は付録に示しておく。また，これらの計算式を用いた，典型的な母音型スペクトル包絡の模式図を第1図に示す。



第1図 母音型スペクトル包絡の構成の模式図 (a)低次の3個のホルマント特性(H_1, H_2, H_3)と音源，放射および高次ホルマントを含む補正項(KR)．(b)それらをまとめたスペクトル包絡



第2図 抽出された第2ホルマント特性の模式図 (a)第1，第3ホルマント周波数の推定値が正しい場合。(b)それらの推定値に正(H_2^+)，負(H_2^-)の誤差がある場合

2.2. 3*FE* 法の導入；抽出原理

ここで第1図のスペクトル包絡をもとに考える。ホルマントはスペクトル包絡の特徴的な山（peak）となっていることは明らかである。まず，第2ホルマント周波数（F_2）を除く他のホルマント周波数が，なんらかの方法でわかったとしよう。そこで，第2ホルマントを欠いた擬似音声スペクトルを計算して，P から差し引くと第2図(a)の $H^0{}_2$ となり，当然第1図(a)の H_2 と一致する。そこで，F_1，F_3 の推定値としてある誤差を含ませて，同様の操作を行なうと第2図(b)の H_2^+，H_2^- となる。ここで注目すべきことは，F_1，F_3 の推定が悪くてもこのような操作によって取り出された H_2^+，H_2^- に，F_2 に対応する優勢なピークが残っていること，および，共振特性としての全体的な型がかなりよく保存されていることである。

このようにして取り出された第2ホルマントの共振特性について，そのピークを含む周波数領域で一次モーメント（スペクトルの重心）を計算すれば，F_2 についてのかなりよい値が期待できる。さらに，これを他のホル

* ここではスペクトル包絡と ホルマントの関係を考察しているので，従来の習慣に従って $KR(f)$ を一定な項として扱っている。しかし，このうち音源特性の項は，ホルマント周波数の抽出という立場からはそのような扱いはできない。それは，ひとつには音源スペクトル包絡が個人差により，あるいは発声の状況によりかなり変動しており，とくに零点のふるまいは無視できないこと，また，声帯の準周期的振動により母音型スペクトルが線スペクトル構造をもっていることなどによる。これらの問題については，本方式の抽出能力の評価を行なう立場から，第3章においてやや詳しく考察する。

マントについての同様の操作—たとえば，第1ホルマントを欠く擬似スペクトルの計算—を行なうときに F_2 の推定値として用いる。各ホルマントについて，順次このような操作をくりかえすと，かなり正確なホルマント周波数が得られることが期待される。

2.3. 抽出過程の構成

フーリエ変換によってスペクトル分析された振幅スペクトルは，そのときの時間窓長を D sec とすると，$1/D$ Hz ごとの周波数きざみで与えられる。そこで，これらの周波数点に，順次低いほうから番号をつけ，BPF 群による分析の場合にならって，チャネル番号とよぶことにする。したがって以下のスペクトル処理計算や一次モーメントの計算などは，これらの周波数点でのみ計算される。

ホルマント周波数の抽出においては，与えられた短時間スペクトル（以下フレームとよぶ）が独立したものであるか，あるいは，同一母音型区間内での連続したフレームであるかの二つの場合に分けて考えると都合がよい。ここでは，前者の場合を MODE-Ⅰ，後者の場合を MODE-Ⅱ とよぶことにする。2.2 節でふれた操作で，特定のホルマントについての擬似スペクトルの計算からそのホルマント周波数を得るまでをステップとよび，3個のホルマントについての3個のステップをまとめてサイクルとよぶことにしよう。

そこで，2.2 節の考えを計算機プログラムとして実現するためには，各ホルマント周波数の初期値の決め方，一次モーメントの計算帯域の決め方，ステップの順番やくり返し方，およびその打切りの条件などを明らかにする必要がある。

(1)初期値の設定：まず MODE-Ⅰ の場合，つまり与えられたフレームについてなんの情報も手元にない場合で，このときの初期値としては，声道を均一管とみなしたときの共振モード，F_1=500，F_2=1500，F_3=2500 (Hz) を用いる。次に，MODE-Ⅱ の場合には，同一母音型区間内ではホルマントが連続的な動きをすることに注目して，一つ前のフレームの最終結果を当フレームの初期値として用いることにする。

(2)一次モーメントの計算帯域：まず MODE-Ⅰ では次の方法による。$ch[X]$ で，周波数 X に最も近いチャネル番号を示すことにして

$$(i_l,\ i_h)=(ch[F_{j-1}],\ ch[F_{j+1}]) \tag{4}$$

つまり第 j ホルマントについては，その両隣りのホルマント周波数のそのときの推定値，F_{j-1}，F_{j+1} ではさまれる領域$(i_l,\ i_h)$ で行なう。ただし，第1ホルマントについては $i_l=ch[200]$ を，第3ホルマントについては $i_h=ch[3500]$ をそれぞれ選ぶ。MODE-Ⅱ の場合には，すでにかなりよい各ホルマント周波数の推定値がわかっているので，いま求めようとするホルマント周波数のそのときの推定値，F_j をもとに

$$(i_l, i_h)=(ch[F_j\cdot(1-\alpha)-\beta],\ ch[F_j\cdot(1+\alpha)+\beta]) \tag{5a}$$

$$\alpha=0.15 \tag{5b}$$

$$\beta=200 \tag{5c}$$

とする。これは F_j を中心とする対称な領域であり，そのホルマントのピーク近辺のみを含むように α，β を定めた。

(3)漸近近似の順序とその打切り：(1)項で述べた初期値を用いて，まず第2ホルマントのステップから始め，$F_2\rightarrow F_1\rightarrow F_3$ の順に一つのサイクルを行なう。これは一般に F_2 が最も大きな変化範囲をもっていることによる。そして各ステップで得た新しい推定値を次のステップに取り入れながら，漸近近似をくり返えす。これらの手続きは MODE によらない。くり返えしの打切りは次のように行なわれる。予め閾値として，MODE-Ⅰ の場合に $\{TH1_i\}$，MODE-Ⅱ の場合に $\{TH2_i\}$ をそれぞれ決めておく。ここで i はホルマント番号である。各サイクルの終りで，すべてのホルマント周波数の新旧推定値の差の絶対値がその閾値より小さいとき漸近近似を打切る。この閾値は，要求される抽出精度により任意に選べるが，ここでは $\{TH1_i\}=\{50,50,70\}$，$\{TH2_i\}=\{10,20,20\}$ と定めた。これらの数値は，D, L. による人間のホルマント周波数の弁別閾値 3～5%[13] や A-b-S 法の精度[14] などを参考にして決めた。

3. 抽出実験の準備；その考え方

3.1. ホルマントの自動抽出における一般的な問題点

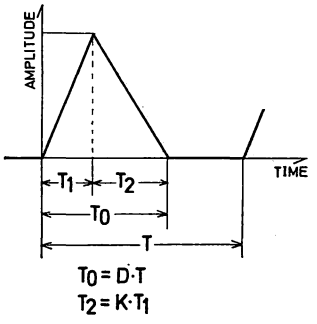
短時間スペクトルから，ホルマント周波数の自動抽出を行なうときに，これが困難となる要因としては，音声波スペクトルそのものの構造によるものと，スペクトル観測機構によるものとに分けることができよう。前者については，その生成過程からみて，音源の特性によるものと，伝達特性によるものに分けられる。とくに，音源の性質がホルマント周波数抽出の立場から考察されたことはほとんどなかったようである*。2.1 節では，母音型スペクトル包絡の構造を，もっぱら伝達特性に注目して話を進めてきた。同節の脚注で少しふれたが，ホルマント周波数の自動抽出の立場から，音源特性として考察しておくべき問題は次の3点であろう。

(1)音源基本周波数 (F_0)：われわれが観測するスペク

* 先にふれた A-b-S 法では，音源スペクトル包絡をも含めた補正項 (KR) として，6種の補正曲線を用意して自動抽出に用いている[15]。

トルは，これを間隔とした高調波構造をもち，またそれらの周波数点でしか伝達特性を知ることができない。この F_0 は話者によって，また発声の場合によってさまざまに変動する。とくに，スペクトル包絡上のホルマントのピークと高調波の同調，離調の関係は大きな問題である。

(2)音源波形と対応したスペクトルの零：声帯振動による音源波形についてはまだ未知のことが多いが，報告はたくさんある[16],[17]。それは第3図に示すような三角波に近いものと考えられており，ほとんどの合成実験はこれを採用している。Dunn らの計算[18]によれば，この三角波のスペクトルの零点は，その非対称係数（K）によって第1表に示すようにさまざまな周波数値をとり，母音型スペクトル包絡に強い影響を与える。たとえば K が1や0.5に近づくと，二重零でしかもその実部が0に近いものが現われ，スペクトル包絡に急峻な谷が生じる。これがホルマントに重なると，ホルマントのピークが消滅してしまい（第11図参照），ホルマント周波数抽出にとっては最悪の事態となる。



第3図　音源波形の近似として用いた三角波

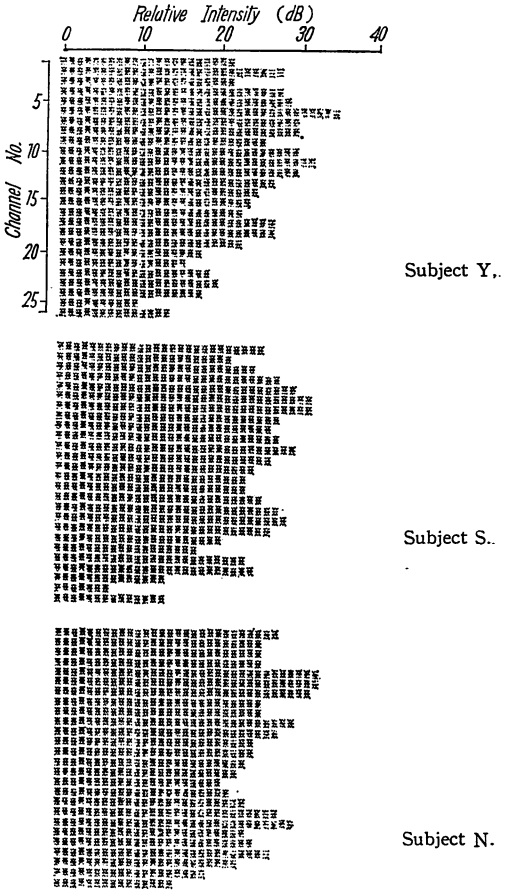
第1表　第3図における非対称係数 K をパラメータとした三角波スペクトルの零点の位置

Frequency Points of Source Zeros ($\omega T_0/\pi$)

	(Asymmetry Factor : K) 0. 5	0. 7	1. 0
1 st	3. 0	3. 3	4. 0*
2 nd	6. 0*	4. 9	〃
3 rd	〃	6. 9	8. 0*
4-th	9. 0	9. 6	〃
5-th	12. 0*	10. 4	12. 0*

* double zeros

(3)巨視的にみた音源スペクトルの個人差：筆者らの経験では，第4図に例を示すように，特定の母音についてもそのスペクトル全体の型に個人差がみられる。これは厳密には伝達特性での個人差も含まれていようが，おもに音源スペクトル全体の型（巨視的にみた）の個人差によるものとわれわれは考えている**。このような場合にモーメント法では安定したホルマント周波数の抽出ができなかった。



第4図　母音の平均スペクトルの個人差の例。
各人20個の母音/a/を抽出し，10ms ごとのサンプル値を累積して平均スペクトルとした。話者は男子（20～30才），これらのスペクトルは第3表にある BPF 群 type-S によってスペクトル分析されたものである

次に伝達特性による要因としては

(4)ホルマント相互の近接，

(5)ホルマントの急激な変化

などがあげられる。

スペクトル観測機構については，従来主として BPF 群が用いられ，その通過帯域の遮断特性や周波数軸上でのそれらの配列と関連して，BPF と高調波の同調，離調の関係，さらにこれに(1)の問題がからみあって複雑な様相を呈する[20]。しかしここでは，より原理的なフー

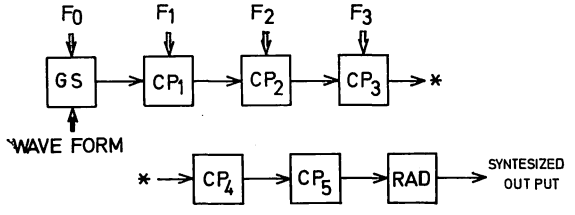
** 最近，当室の高杉らによって進められている声帯音源波の抽出実験[19]においても，声帯波形の個人差が著しいことが明らかになりつつある。

リェ変換によってスペクトル分析を行ない，この問題をさけた。しかし，実用上は無視できない問題なので，BPF 群を用いる場合についての若干の考察を 4.3 節で行なう。

3.2. 資料としての合成音の準備

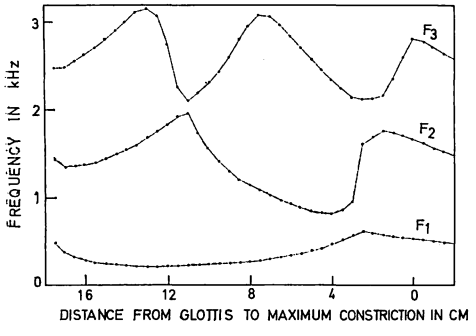
抽出実験にあたって，前節で列挙した問題点を含むように合成音を準備し，*3FE* 法の抽出能力をかなり厳密に評価することにした。

合成方式としては，計算機合成に便利な z 変換法によるターミナルアナログ方式を用いた。合成系のブロックダイアグラムは第5図に示すとおりで，制御可能な合成パラメータは，音源波形とその基本周波数 (F_0) および第1から第3までのホルマント周波数 (F_1, F_2, F_3) である。第6以上の高次ホルマントの補正は，z 変換法固有の折返し特性で代用させてある[21]。なお系のサンプリング周期は 100μsec で，5kHz までの信号帯域を考



GS : GLOTTAL SOURCE WAVE GENERATOR
CP : CONJUGATE POLE CIRCUIT
RAD : RADIATION CIRCUIT
⇨ : CONTROL INPUT

第5図 合成系のブロックダイアグラム



第6図 合成音のホルマント周波数パタン；声道の四管近似で調音点を声門から口に動かしたときの共振モード，横軸 1 cm は合成時に 40 ms に対応させた

えた。また第4，第5ホルマント周波数はそれぞれ 3500 Hz と 4500Hz に固定し，ホルマントの帯域幅は 50 $\{1+F_i^2/6\times10^6\}$ Hz である（F_i は第 i ホルマント周波数[22]。

まず前節の問題点(4)，(5)に関して，合成音のホルマント周波数を第6図のように選んだ。これは Fant によって与えられたもので，声道を4個の断面積の均一な音響管からなるとみて，その調音点（最小の断面積の管の中心）を，声門に相当する位置から口に相当する位置に向って動かしたときの共振モードにあたる[1]。図からもわかるように，第2，第3ホルマントの急激な移行部と，ホルマント相互の近接を含んでいる。また母音型スペクトルとして，いわゆる *F*-パタンをほとんどすべて含んでいる。

次に音源特性の問題点(1)，(2)に関しては，音源波形に着目したものと，その F_0 に着目したものの二つのグループに分けた。グループAは，第3図の三角波の K が 0.5，0.7 および 1.0 のものとインパルスに 6dB/oct の低域強調をほどこしたものの計4種の音源波形を用いた（$D=0.5$，$F_0=140$ Hz 一定）。グループBは，$K=0.7$ の三角波のみで，F_0 として 100 Hz から 200 Hz まで 20 Hz おきの計6種からなっている。これら2グループの音源と，先ほどの *F*-パタンによって合計10種の合成音を準備した。

3.3. その他の資料

人の発声になる自然音（合成音に対比してこうよぶことにした）としては，日本語5母音および天気予報の一節を用意した。前者は高杉らの声帯音源波形の抽出実験[16]に用いられたもので，男子成人5名によって防音室で発声され，コンデンサマイクロホンを通じてFMレコーダに収録されたものである。天気予報は，NHKアナウンサ2名分で，FM放送から通常のテープレコーダに収録されたものである。

なお母音の発声者には第4図の発声者2名を含んでおり，以上7名分の自然音と前節の合成音グループAによって，3.1 節の問題点(3)を確かめてみる意味も含まれている。

3.4. スペクトル分析

従来，音声のスペクトル分析には，主として BPF 群が用いられてきた。しかしこれには BPF 群の構成のしかたによる多くの制約があった[20]。このため Mathews らは精密な母音分析にあたってフーリェ分析を用いた[23]。最近，Cooley と Tukey によって発表された高速フーリェ変換（FFT）の手法[11],[12]は，従来のフーリェ変換の計算法に比べて各段に速く，現有の計算機でじゅうぶん実用になるとして各方面で注目されてきた。当所の音声研究グループでもこれを FORTRAN のサブルーチンとして利用できるようにし，音声研究に役立て

つつある[19]。そこで，本実験でも FFT によるスペクトル分析を用いることにした。

母音と天気予報の自然音はまず 3.4kHz の LPF を通したのち，10kHz のサンプリング周波数で10進3桁に数値化され，計算機用磁気テープに書き込まれる。合成音もやはり同じ形式で磁気テープに書き込まれる。

これらは 12.8ms の時間長の humming window を用いてフーリェ変換される。したがって振幅スペクトルの周波数きざみは約 79Hz となる。この時間窓を 10ms ごとにずらし，10ms ごとの短時間振幅スペクトルが抽出プログラムに供給される。

4. 分析結果と検討；抽出能力の評価

4.1. 自然音の分析結果と検討

人の発声による自然音は，その物理的構成に関するパラメータを正確に知ることができないので，抽出結果についての厳密な評価を行なうことはできない。しかし，抽出プログラムは最終的に自然音を扱うのであるから，経験的事実を結集してその当否を確かめておく必要があろう。

5母音の定常部（持続時間の中心）の抽出結果は第2表に示すとおりである。スペクトログラムや FFT によるスペクトルの視察から，これらは各母音のホルマント周波数としてもっともな数値と思われる。さらに第2表の結果は同一資料を用いた A-b-S 法の結果ともほぼ一致している。

天気予報の母音型部の抽出結果の例を第7図および第8図にスペクトログラムとともに示す。各図のスペクトログラム（wideband）の写真とホルマント周波数のトレースは同一縮尺としてあり，これらを重ねることによって，ホルマント周波数の抽出が非常にうまくいっていることがわかる。以上の単独母音と連続音声の母音型部の抽出結果は 3.1 節の問題点(3)に対する解答の一部ともなり，合計7名分の話者の個人差によらず安定な抽出ができることを示している。

第7図(a)の横太線で示した母音型部の継続時間は 280 ms であり，これの抽出に要した時間は約 42sec であった。したがって1フレームあたり 1.5sec で，時間率にして 1.5×10^2 となる。

4.2. 合成音の分析結果と検討

合成音については，そのグループごとに抽出結果を示しながら検討する。なお抽出例はすべてを掲載することができないので，原則として最悪のもは必ず示すことにする。

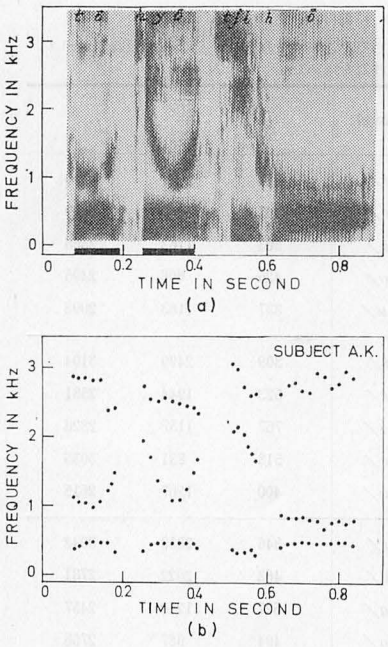
第2表 単独母音のホルマント周波数

各母音の持続時間の中心部を視察により定め，この点の短時間スペクトルからホルマント周波数を求めた。

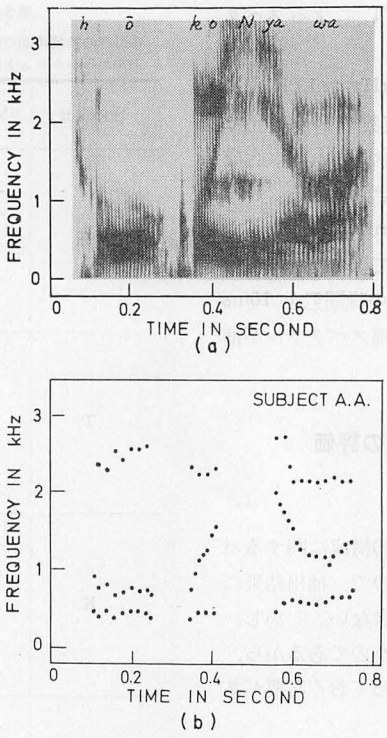
Speaker	Vowel	F_1	F_2 (Hz)	F_3
S.	/i/	320	2081	3190
	/e/	447	1961	2409
	/a/	563	1162	2368
	/o/	425	606	2495
	/u/	337	1103	2098
T.	/i/	309	2429	3104
	/e/	523	1944	2581
	/a/	767	1157	2826
	/o/	518	831	3035
	/u/	400	1267	2615
K.	/i/	346	2350	3042
	/e/	463	2022	2731
	/a/	527	1240	2457
	/o/	494	987	2758
	/u/	391	1246	2954
N.	/i/	273	2108	3147
	/e/	500	1904	2608
	/a/	543	1151	2938
	/o/	486	970	2944
	/u/	355	1164	2250
SM.	/i/	312	2526	3263
	/e/	412	2161	2560
	/a/	806	1216	2531
	/o/	382	774	2379
	/u/	327	879	2517

まずグループBについては，F_0=200 Hz の場合の抽出結果をスペクトログラム（narrow）とともに第9図に示す（以下，合成音のスペクトログラムは，高調波構造との関係に注目するのですべて narrow による）。これはグループBの中で最も高調波間隔が広く，したがってホルマント周波数の抽出が困難と思われるものであるが，かなりよい結果を示している。合成音グループBの合成時に与えたホルマント周波数（第6図参照）と抽出された値との差の絶対値を累積分布としてみた。これらのうち，F_0 が 120 Hz，160 Hz および 200 Hz の場合のものを第11図に示す。これから，一般に F_0 が高くなるほど抽出精度が悪くなることがわかる。

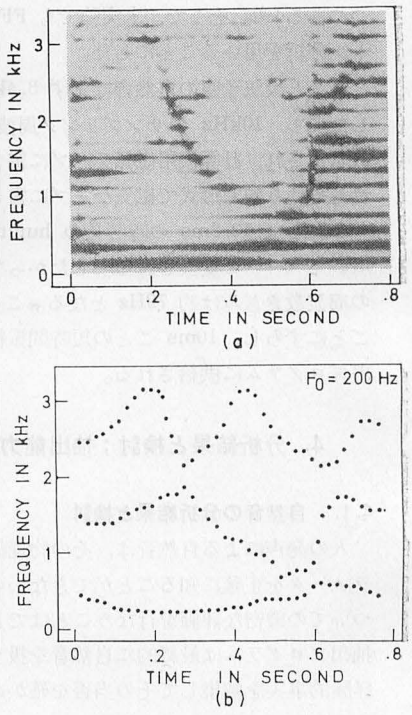
次にグループAについては，音源三角波の非対称係数 K が 0.5 のものを第11図に，同じく 0.7 のものを第12図に，さらにインパルスに 6dB/oct の低域強調をほどこしたものを第13図にそれぞれ示す。これらのうち，と



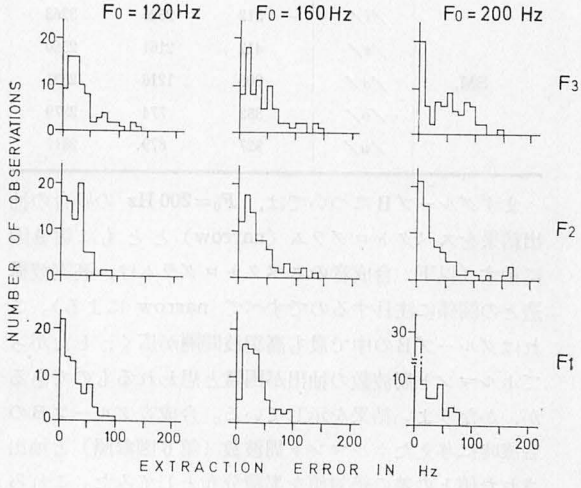
第7図 連続音声の母音型部のホルマント周波数抽出例-Ⅰ。
(a)スペクトログラム(wide)。
(b)母音型部のホルマント周波数のトレース



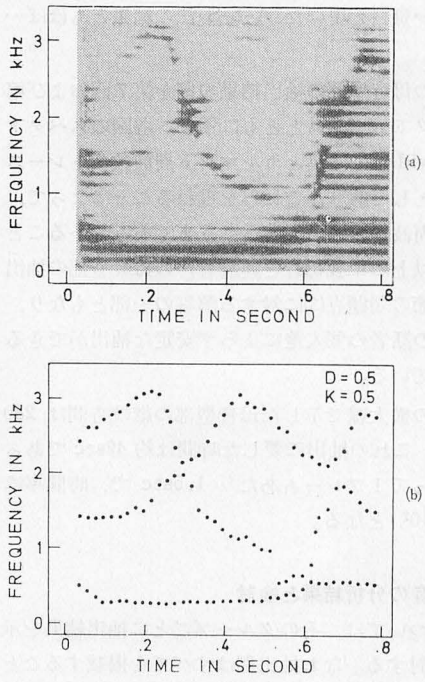
第8図 連続音声の母音型部のホルマント周波数抽出例-Ⅱ。
(a)スペクトログラム(wide)。
(b)母音型部のホルマント周波数トレース



第9図 合成音グループBの F_0=200 Hz の場合のホルマント周波数抽出結果。
(a)スペクトログラム(narrow)。
(b)抽出結果のトレース

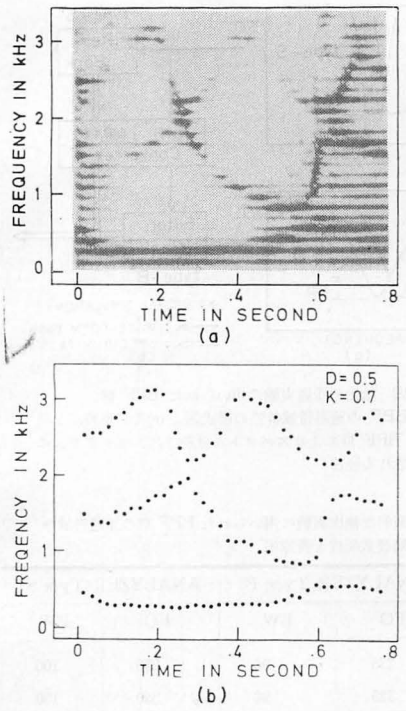


第10図 ホルマント周波数の抽出誤差の累積分布の例-Ⅰ。
合成音グループB でF_0=120, 160, 200 Hz の場合

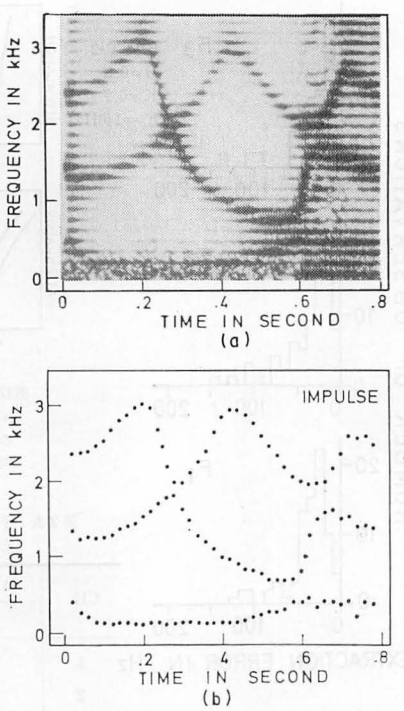


第11図 合成音グループAの K=0.5 の場合のホルマント周波数抽出結果。(a)スペクトログラム(narrow)。(b)抽出結果のトレース

くに注意する必要があるのは第11図の K=0.5 の場合である。同図(a)のソナグラムからもわかるように，第6，第12および第18高調波が，音源スペクトルの零（それも，二重零で，その実部が0のもの）によって，非常に弱められている。また同図のホルマント周波数のトレースで，



第12図 合成音グループAの K=0.7 の場合のホルマント周波数抽出結果。(a)スペクトログラム（narrow)。(b)抽出結果のトレース



第13図 合成音グループAで音源としてインパルスに 6dB/oct の低域強調をほどこしたもののホルマント周波数抽出結果。(a)スペクトログラム（narrow)。(b)抽出結果のトレース

ント周波数をよくトレースしている。グループBの場合と同様に抽出誤差の累積分布をとった。その例を第14図に示す。同図の K=0.5 場合には先に述べた事情により，第2ホルマント周波数の抽出誤差が 200 Hz あたりまでかなり広く分布している。しかし K=1.0 の場合（やはり，音源の零は K=0.5 のときと同様二重零で，その実部は0である）には，零の影響は少ないようである。

以上の考察から，第11図のように極端な悪条件が重ならないかぎり安定した抽出が期待でき，また抽出精度としては，第10図と第14図から音源基本周波数の半分（$F_0/2$）という目安がたてられよう。

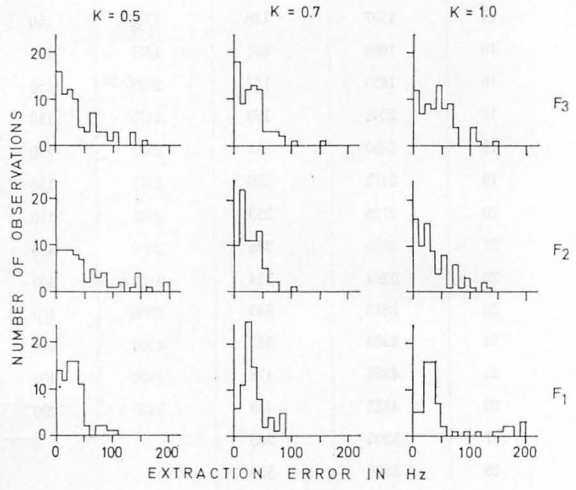
4.3. 高速化の試み

前節までの実験に用いた抽出プログラムは，抽出原理に従ってFORTRAN言語で書いたものである。したがってプログラムの構成にはかなりのむだがある。抽出過程で最も計算時間を費やすのは，単共振特性（(A-1)式）の計算である。いままでは各ステップの初めにいちいちこれを計算していた。

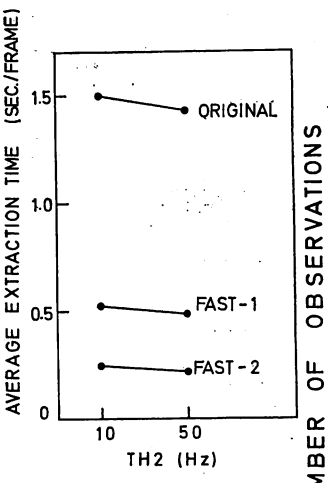
そこでまず，各ステップの初めでホルマント周波数の推定値が一つ前の推定値に近いときには，前のステップでの（A-1）式の計算結果をそのまま用いればよい。このように改良したプログラムを FAST-1 とする。さらに計算機の記憶容量に余裕のあるときには，（A-1式)の計算を予め行ない表にして必要に応じて読み出せばよい。この考え方で，共振周波数が 200 Hz から 3500 Hz まで 50 Hz おきの単共振特性の表を準備しておいたものを FAST-2 とする。

これらの改良されたプログラムで，いままでと同様の抽出実験を行なった。打切りの閾値，$\{TH2_i\}$ をパラメータとして，第7図の太線部の平均抽出時間を示すと第15図となる。FAST-2 の場合には，時間率 25 倍で改良前からはほぼ一桁速くなっている。また FAST-2 による抽出誤差の累積分布の1例を第16図に示す。これを第10図の対応する分布と比べると，抽出誤差が若干大きくなっていることがわかる。

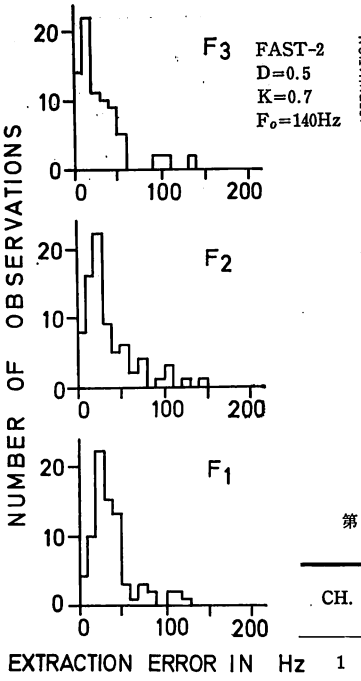
第1，第2ホルマントの近接部分での第2ホルマントの抽出がうまくいっていない。これは音源スペクトルの零による急峻な谷と第2ホルマントの一致，さらに両ホルマントの近接という悪条件が二つ重なったことによる。この例以外では，多少のずれはあっても合成時のホルマ



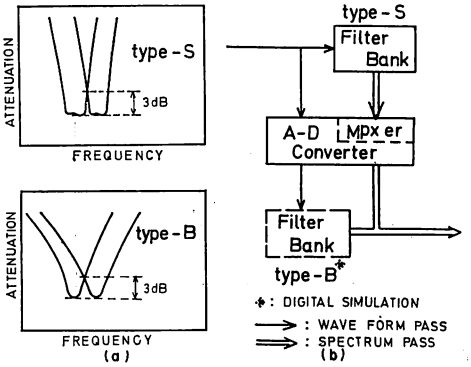
第14図 ホルマント周波数の抽出誤差の累積分布の例-Ⅱ。合成音グループAで音源三角波の非対称係数 K が 0.5, 0.7, 1.0 の場合



↑第15図 抽出プログラムの改良による抽出時間（速度）の比較



第16図 高速化した抽出プログラム（FAST-2）による抽出誤差の累積分布の1例 →



第17図 抽出の予備実験に用いられた BPF 群。(a)BPF の通過帯域特性の模式図。(b)音声資料が2種の BPF 群によりスペクトル分析され，プログラムに供される経路

第3表 予備的な抽出実験に用いられた BPF 群の通過帯域中心周波数配置と帯域幅

CH.	ANALYZER(Type B)		ANALYZER(Type S)	
	FO	BW	FO	BW
1	235	90	250	100
2	325	90	350	100
3	415	90	450	100
4	505	90	550	100
5	595	90	650	100
6	685	90	750	100
7	775	90	850	100
8	865	90	950	100
9	955	90	1050	100
10	1050	100	1150	100
11	1155	110	1275	150
12	1270	121	1425	150
13	1397	133	1575	150
14	1537	146	1725	150
15	1690	161	1875	150
16	1859	177	2025	150
17	2045	195	2175	150
18	2250	214	2325	150
19	2475	236	2475	150
20	2725	259	2625	150
21	2995	285	2900	400
22	3294	314	3300	400
23	3623	345	3700	400
24	3986	380	4100	400
25	4385	418	4500	400
26	4823	459	5300	1200
27	5305	505		
28	5836	556		
29	6419	611	in Hz	
30	7061	673		

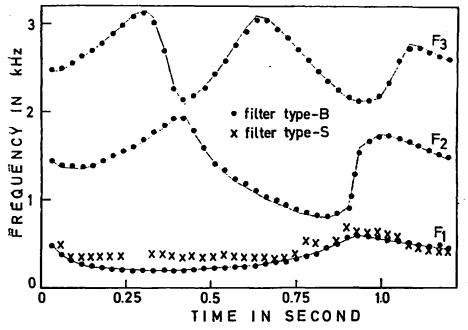
4.4. スペクトルの分析に BPF 群を用いる場合の考察

抽出実験でのスペクトル分析には，高速フーリェ変換（FFT）が用いられた。しかしより実用的な見地からは，従来の BPF 群による分析に，*3FE* 法がどう適応できるかを明らかにしておく必要がある。これについては，以前に2種の BPF 群を用いた本抽出法の予備的な実験[26]が行なわれており，ここではその結果を要約しておく。

第17図に示すような2種の BPF 群を用いたスペクトル分析によって，本実験と同様の合成音（ただし，音源としてはインパルスに 6dB/oct の低域強調をほどこしたもののみ）による抽出実験を行なった。なおこれら2種の BPF 群の，通過帯域の中心周波数の配置と帯域幅を第3表にあげておく。それらによる抽出結果の1例を第18図に示す。同図中×印は第17図の BPF 群 type-S を用いた抽出結果で，通過帯域の遮断特性が急峻であるため抽出誤差がめだつ。いっぽう・印は単共振特性を2段重ねた BPF 群（第17図の type-B）による場合で，合成時のホルマント周波数をよくトレースしている。これらの結果は，中田，松野による音声スペクトル分析用 BPF 特性の検討結果[20]とも一致している。したがって type-B のような構成にした BPF 群が望ましく，またそれは本方式にもそのまま適応できる。

5. むすび

音声の基礎的な研究に利用するために，簡便，安定，



第18図　2種の BPF 群をスペクトル分析に用いた予備実験での抽出結果の1例。
(filter type-S による結果の F_2, F_3 は，filter type-B のそれとほとんど重なるので示してない)

高速かつ高精度といった実用性に主眼をおいて，3*FE* 法と名づけたホルマント周波数抽出法を提案し，これを FORTRAN 言語でプログラミングして，計算機による抽出実験を行なった。実験にあたっては，抽出資料として自然音のほかに，一般にホルマント周波数の自動抽出がむずかしくなると考えられるいくつかの問題点を含む合成音を用意し，この諸元の明確な合成音の抽出結果を検討することにより，3*FE* 法の抽出能力のかなり厳密な評価を行なった。その結果，3*FE* 法はかなり精度がよく，また高速かつ安定であることがわかり，今後の音声研究に実用的な方法として役立つことを期待している。

抽出実験の結果と検討に基づいて，その成果と本実験の特徴を要約すれば次のようになる。

(1) F_0 と音声波形に着目した合成音，および多数の発声者による自然音による抽出結果から，個人差によらずに安定した抽出ができる。

(2)音源スペクトルでの零の存在，ホルマント相互の近接，ホルマントの急激なトランジション，および各種の F_0 値などのさまざまな悪条件のもとで，かなりよい抽出精度をもち，その目安は $F_0/2$ である。

(3)抽出に要する時間は平均 0.25秒/フレームで，時間率にして約25倍である。実時間抽出にはまだほど遠いが，従来のものよりはるかに速い。

(4) FFT の手法によるフーリェ変換を用いて，スペクトル観測機構による問題をさけた。さらにスペクトル分析に BPF 群を用いる場合の，本方式の適応について考察を行なった。

(5)抽出過程を構成しているアルゴリズムが簡明である。これは母音型スペクトルの特徴的な構造をうまく利用できたことの反映であろう。

次に 3*FE* 法の問題点，あるいは今後の課題としては，次のことがあげられる。

(1)び音などのように，伝達系に零点が含まれている場合を考慮していない。これは今後の本格的な検討を必要とする。

(2) 女声の場合のように，F_0 がとくに高い場合には，本方式はこのままでは適用できない。

(3)前述の特徴点(4)は，あくまでもそのアルゴリズムに関した話しであり，プログラムの簡素さ（たとえば他の方法に比べてプログラムの専有記憶容量が少ない）ということは意味しない。

とくに(3)は小型の計算機を利用する場合に問題となろう。したがって目的に応じたプログラムの構成，たとえば精度が要求される場合と専有メモリの少ないことが要求される場合などに分けて，最適なものを準備しておきたい。

終りに，常にご指導いただく川上部長に謝意を表する。また有益な討論と援助を下さった角川研究官，FFT のプログラムを提供していただいた高杉技官，および計算機の使用にあたって協力下さった中村技官を始め情報処理部の各位に感謝する。

参　考　文　献

(1) Fant, G.; "Acoustic Theory of Speech Production," s-Gravenhage : Mouton & Co., 1960.

(2) Delattre, P. C., Liberman, A. M. and Cooper, F. S.; "Acoustic Loci and Transitional Cues for Consonants," JASA., 27, No. 4, 1955.

(3) Fry, D. B., Abramson, A. S., Eimas, P. D. and Liberman, A. M.; "The Identification and Discrimination of Synthetic Vowels," Language and Speech, 5, p. 171, 1962.

(4) 鈴木誠史，中田和男；"日本語単音節の音韻分類と識別"，電通学誌，46, No. 11, p. 168, 1963.

(5) Hughes, G. W.; "The Recognition of Speech by Machine," M. I. T. Technical Report 395, 1961.

(6) Bell, C. G., Fujisaki, H., Heinz, J. M., Stevens, K. N. and House, A. S.; "Reduction of Speech Specra by Analysis-by-Synthesis Techniques," JASA., 33, No. 12, p. 1725, 1961.

(7) Stevens, K. N.; "Toward a Model for Speech Recognition," JASA., 32, No. 1, p. 47, 1960.

(8) 鈴木誠史，角川靖夫，中田和男；"モーメント計算によるホルマント周波数の抽出"，音響学誌，19, No. 3, p. 106, 1963.

(9) 中津井護，鈴木誠史；"連続音声識別の検討"，電波研季報，11, No. 52, p. 30, 1965.

(10) 中津井護，鈴木誠史；"消去法によるホルマント周

波数抽出", 電波研季報, 12, No.59, p.111, 1966.

(11) Cooley, J. W. and Tukey, W.; "An Algorism for the Machine Calculation of Complex Fourier Series, "Mathematics of Computation, 19, p.297, 1959.

(12) 角川靖夫, 中津井護, 鈴木誠史, 高杉敏男; "高速フーリェ変換と最近のスペクトル分析装置"(解説), 電波研季報, 15, No.76, p.43, 1969.

(13) Flanagan, J. L.; "Perceptual Criteria in Speech Processing," Speech Communication Seminar, Stockholm, 1962.

(14) 角川靖夫, 中田和男; "「合成による分析法」によるホルマント周波数抽出", 音学誌, 20, No.1, p.1, 1964.

(15) 藤崎博也; "電子計算機による母音のホルマント抽出", 情報と制御の研究, 第2,3合併号, 1962.

(16) Miller, R. L.; "Nature of Vocal Cord Wave," JASA, 31, No.6, p.667, 1959.

(17) Flanagan, J. L.; "Some Properties of the Glottal Sound Source," J. Speech and Hearing Research, 1, p.99, 1958.

(18) Flanagan, J. L.; "Speech Analysis, Synthesis and Perception," p.195, Springer-Verlag, 1965.

(19) Takasugi, T. and Suzuki, J.; "Speculation of Glottal Waveform from Speech Wave," J. Rad. Res. Labs., 15, No.82, p.279, 1968.

(20) 中田和男, 松野辰治; "音声研究用計算機入出力装置の検討, 第2部音声スペクトル分析用 BPF 特性の検討", 電波研季報, 9, No.44, p.248, 1963.

(21) Gold, B. and Rabiner, L. R.; "Analysis of Digital and Analog Formant Synthesizers," IEEE Trans. on Audio and Electroacoustics, AU-16, No.1, p.81, 1968.

(22) Fant, G.; "Formant Bandwidth Data", STL-QPSR-1/1962, p.1, Royal Institute of Technology (Sweden).

(23) Mathews, M. V., Miller, J. E. and David, E. E. Jr.; "Pitch Synchronous Analysis of Voiced Sounds," JASA., 33, p.179, 1961.

(24) Nakatsui, M. and Suzuki, J.; "Fast Formant Frequency Tracking Technique,', 6-th I. C. A., B-2-4, 1968.

付　　録

本文の(3a), (3b) 式の各項の近似式はそれぞれ,

$$H_i(f)=\frac{F_i^2+B_i^2/4}{\{(f-F_i)^2+B_i^2/4\}^{1/2}\{(f+F_i)^2+B_i^2/4\}^{1/2}} \tag{A-1}$$

$$20\log_{10}\{HP(f)\}=0.717(f/500)^2+0.00318(f/500)^4 \tag{A-2}$$

$$Q(f)\cdot R(f)=f/(1+f^2/100^2) \tag{A-3}$$

である。ここで, F_i, B_i はそれぞれ第 i ホルマントの周波数と帯域幅である。(A-2) 式では, 高次ホルマントとして, 第4ホルマント以上をとっている。