

論文 / 著書情報
Article / Book Information

論題	音声の個人性パラメータの時期的変動の解析
著者	古井 貞熙
掲載誌/書名	研究実用化報告, Vol. 25, No. 2, pp. 425-436
発行日	1976,

534. 322 : 801. 4 "4" : 534. 62 : 519. 26

音声の個人性パラメータの時期的変動の解析*

古井 貞 熙**

あらまし 音波波スペクトルに含まれる個人性情報を、音声生成のメカニズムに基づいて、音源特性に起因するものと声道特性に起因するものとに近似的に分離し、それぞれの特性を表現するパラメータの時期的変動に関する分析を行った。まず「時期的変化率」を定義してその値を求めたところ、声道特性よりも音源特性の方が、時期的にゆっくり大きく変化する傾向が強いことがわかった。つぎに話者要因に関する分散分析を行った結果、音源特性の話者効果は短期間では大きい、長期間になると変動量が増大するために相対的にかなり小さくなるということがわかった。これらの結果に基づいて話者認識実験を行い、放射および音源特性を近似的に除去する逆フィルタを用いることが、時期的に安定した個人性情報を抽出する上できわめて有効であることを確認した。10日間わたる学習サンプルを用いて、3~21か月後の未知サンプルの話者を99%の認識率で認識することができた。

1 ま え が き

音波に含まれる情報は、意味内容に関するものと、発声者の個人差に関するものとに分けることができる。後者の個人性情報、すなわち誰の声であるかという情報を音波から抽出し、客観的なパラメータで表現することは、人間-機械系としての音声研究の一環として、音声認識の高度化や合成音の高品質化とも密接な関連を有し、将来は話者の機械認識という応用も考えることができる⁽¹⁾。

一般に、音波から個人性情報を客観的なパラメータで抽出する上でもっとも困難な点の1つは、ある程度長い期間をおくと、同じ話者が同じ言葉を発声しても、それらのパラメータが変化してしまうところにある。筆者らはこれまでに、個人性情報を表現するパラメータとして、短文章の長時間平均スペクトルのケブストラムと、単語の有声音区間における基本周波数と部分自己相関係数(PARCOR係数)に関する統計量をとり上げ、主としてこれらのパラメータの長期間にわたる変動特性と、個

人差を表現する上での有効性に着目した種々の分析を行ってきた^{(2)~(5)}。その結果、これらのパラメータはいずれも、10日間程度の短期間では安定しているが、その後時期がたつにつれて、同じ発声者によるものでもゆっくり大きく変化する性質をもつことが明らかになった。このため、単語の統計的パラメータを用いた話者認識実験において、10か月程度の長期間の学習サンプルを用いれば、3か月後の未知サンプルの話者を99%の認識率で認識することができるが、10日間程度の短期間の学習サンプルを用いた場合には、認識率が大きく低下してしまうことがわかった。

個人性情報の本質に近づくという観点から考えると、短期間では安定していても時期がたつにしたがって大きく変化してしまうパラメータは、真に個人性を表現しているとはいえない。また、話者認識への適用ということを考えると、各話者について、長期間の学習サンプルを用意しなければならないということは、好ましいことではない。本論文では、音波波スペクトルに含まれる情報を、音声生成のメカニズムに基づいて、音源特性に起因するものと声道特性に起因するものとに近似的に分離し、パラメータの時期的変動の要因に関する分析を行う。これまでに、音声の個人性パラメータの時期的変動の要因について分析を行ったり、その影響を除去する方法について検討した例はほとんどない。

* Analysis of Temporal Variation of Feature Parameter Characterizing Individual Talker. By Sadaoki FURUI.

この研究は斎藤特別研究室(もと基礎研究部第4研究室)で行われたものである。

** 斎藤特別研究室研究主任

2 音響分析

2.1 音源と声道特性の分離

音声波形 $s(t)$ は、インパルス発生器または白雑音発生器の出力を適当に混合した駆動音源信号 $e(t)$ を、それぞれ音源の特性、声道の特性、放射の特性を表わすインパルスレスポンス $c(\tau)$, $t(\tau)$, $r(\tau)$ で特徴づけられる線形回路網によってろ波した結果と考えることができる。この関係を周波数領域で表示すれば、

$$S(\omega) = R(\omega)T(\omega)C(\omega)E(\omega) \quad (1)$$

と書くことができる。ただし $S(\omega)$, $E(\omega)$, $C(\omega)$, $T(\omega)$, $R(\omega)$ は、それぞれ $s(t)$, $e(t)$, $c(\tau)$, $t(\tau)$, $r(\tau)$ のフーリエ変換であり、実際の音声においては、 $E(\omega)$, $C(\omega)$, $T(\omega)$, $R(\omega)$ のいずれも時間とともに変化している。

実際に観測される音声波形 $s(t)$ およびそのフーリエ変換 $S(\omega)$ から、音源、声道、放射のそれぞれの特性を厳密に分離して取出すことはきわめて困難であるが、何らかの仮定を設ければ、近似的に分離することができる。このうち放射特性は、発声者や音韻に無関係に、約 6 dB/oct で上昇する特性で近似できることが知られているから、声道特性あるいは音源特性のいずれかが近似的に設定されれば、他方が自動的に決定される。

ここでは、つぎのような方法で音源と声道の特性の分離を試み、それぞれの特性を表現するパラメータの時期的変動特性を調べる。

〔方法Ⅰ〕この方法では、声道は滑らかで損失のない半無限長均一音響管からの微小変形として表わされるものと考え、極は全周波数帯にほぼ均一に分布するため、その周波数伝送特性の概形は、傾斜をもたず平坦であるものと仮定する。すなわち、音声波のスペクトルパタンの概形は、放射特性を含めた音源特性を表現しているものと仮定する。音声波形から、放射特性を含めた音源特性を上記の仮定のもとに適応的に (10 ms ごとに) 推定し、その特性の逆フィルタにより音声波形から声道の特性を抽出する⁽⁷⁾。

〔方法Ⅱ〕長時間にわたる音声の平均スペクトルは、種々の音韻に対応する声道の特性が平均化されるため

に、放射特性を含めた音源特性を近似的に表現してゐると考えられる。この方法では、長時間平均スペクトルパタンから、その概形として放射特性を含めた音源特性を推定し、この特性の逆フィルタに音声波を通すことにより、声道の特性を抽出する。長時間音声のサンプルとしては、長時間平均スペクトルの分析結果⁽²⁾⁽³⁾をもとに、約 10 秒長の短文章を発声したものを用いる。

〔方法Ⅲ〕方法ⅠとⅡの中間的な方法として、単語の時間平均スペクトルパタンから、その概形として放射特性を含めた音源特性を推定し、この特性の逆フィルタにより、音声波形から声道の特性を抽出する。

以上の3つの方法のブロックダイアグラムを図1に示す。いずれの場合も、音源特性を簡単のために1次系および2次臨界制動系 (2次 Critical Damping 系、以後2次 CD 系と略記) で近似した。逆フィルタの伝送関数は、つぎのように表わすことができる。

$$\left\{ \begin{array}{l} \text{1次系: } 1 + \beta Z \\ \text{2次 CD 系: } 1 + \gamma Z + \gamma^2 Z^2 / 4 \end{array} \right\} \quad (2)$$

$$Z = e^{-j\lambda}, \lambda: \text{radian/unit time}$$

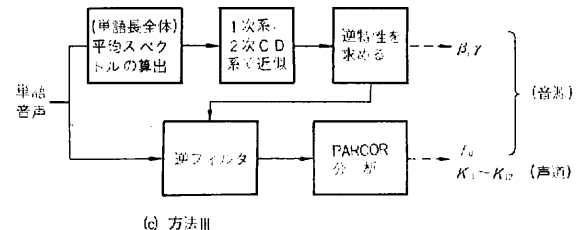
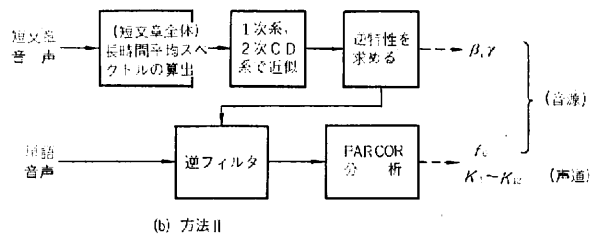
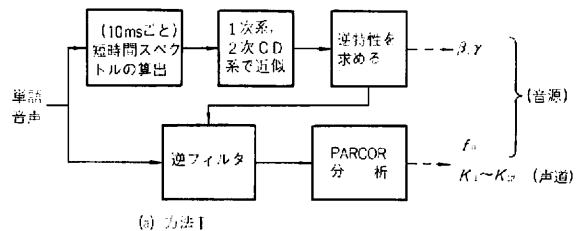


図1 音源と声道特性の分離方法

したがって、放射特性を含めた音源スペクトル特性はパラメータ β および γ によって表現され、声道の特性は逆フィルタ後の音声波形によって表現されることになる。逆フィルタ後の音声波形に関しては、図1に示すようにPARCOR分析を行い、声道の特性を表わすパラメータとしてPARCOR係数(K_i)を抽出する⁽⁸⁾。PARCOR係数は、声道の反射係数として解釈することができるから、PARCOR係数から声道断面積関数を推定することができる。逆フィルタ後の音声波形からは、さらに基本周波数(f_0)を抽出し、 f_0 についても他のパラメータと同様の分析を行う。

2.2 音声サンプル

静かな部屋で発声された5母音または単語の音声波を、3.4 kHz 低域通過フィルタに通したのち、8 kHzでサンプリングし、11 bit で量子化を行う。10 ms 間隔で32 ms 長の区間(フレーム)を抽出し、ハミングウィンドウを乗じてから、各フレームごとに上記の音源と声道特性の分離を行い、 β 、 γ 、 f_0 、 K_i のパラメータを抽出する。各パラメータの全有声音区間における平均値を求め、これらの値について、後に述べる種々の分析を行う。

5母音のサンプルとしては、/i/, /e/, /a/, /o/, /u/を単独に発声したものを用い、単語のサンプルとしては、/baŋgo:/, /namae/の2種類のサンプルを用いた。いずれも男性9名の話者が10か月～3年間の長期間にわたって繰り返し発声したものである。

3 時期的変動特性の評価法

話者 i ($1 \leq i \leq 9$) が時期 j ($1 \leq j \leq n$; 間隔は2～3日ないし3か月) に繰り返し3回発声したサンプルのうち、 k 番目 ($1 \leq k \leq 3$) のサンプルを X_{ijk} で表わす。これは有声音区間における β 、 γ 、 K_i 、 f_0 などの平均値(スカラー)、あるいはこれらの組合せからなる N 次元ベクトルである。

$$X_{ijk} = (X_{ijk1}, X_{ijk2}, \dots, X_{ijkN}) \quad (3)$$

同じ時期の3サンプルの平均値を X_{ij} で表わす。

$$X_{ij} = (X_{ij1}, X_{ij2}, \dots, X_{ijN})$$

$$= \sum_{k=1}^3 X_{ijk} / 3 \quad (4)$$

同じ時期に発声した3サンプル相互間のユークリッド距離(二乗)を、 n 時期、9名の話者および3サンプルから2サンプルを選ぶ3通りの組合せについて平均したものを、同時期変化量と定義し、つぎのように表わす。 $D_{0i}(l)$ は各成分の同時期変化量、 D_0 はベクトルとしての同時期変化量である。

$$\begin{aligned} D_{0i}(l) &= E_{i,j,k} (X_{ijk1} - X_{ijk'l})^2 \\ &= \frac{1}{27n} \sum_{i=1}^9 \sum_{j=1}^n \sum_{1 \leq k < k' \leq 3} (X_{ijk1} - X_{ijk'l})^2 \end{aligned} \quad (5)$$

$$D_0 = \sum_{i=1}^9 D_{0i}(l) \quad (6)$$

同様に、 τ だけの時間間隔を離れたサンプル間の距離の平均値を、時期的変化量と定義し、つぎのように表わす。同時期的変化量と分離するため、各時期のサンプルの平均値の間の距離として求める。

$$D_i(l, \tau) = E_{i,j} (X_{ij-l} - X_{ij+l})^2 \quad (7)$$

$$t(j) - t(j') = \tau$$

$t(j)$: 時期 j の絶対的時期

$$D(\tau) = \sum_{i=1}^9 D_i(l, \tau) \quad (8)$$

時期的変化量を同時期変化量で割った値を時期的変化率と定義し、つぎのように表わす。

$$V_i(l, \tau) = D_i(l, \tau) / D_{0i}(l) \quad (9)$$

$$V(\tau) = D(\tau) / D_0 \quad (10)$$

時期的変化率は、用いるパラメータの種類にほとんどよらずに、 $\tau \geq 6$ か月の範囲では一定となることが実験的に確かめられているので、この範囲での平均値を求め、これを長期間変化率とする。 X_{ijk} の各成分の長期間変化率を $V_{iL}(l)$ で、ベクトルとしての長期間変化率を V_L で表わす⁽⁸⁾。

$$V_{iL}(l) = E_{\tau \geq 6 \text{ 月}} V_i(l, \tau) \quad (11)$$

$$V_L = E_{\tau \geq 6 \text{ 月}} V(\tau) \quad (12)$$

時期的変化率(長期間変化率)の小さいパラメータは、時期がたっても変化量の増大しない安定したパラメータであると言える。話者の特徴を表現し、話

者認識に用いるためのパラメータとしては、個人性情報を多く含み、しかも長期間変化率の小さいパラメータであることが望ましい。

4 時期的変化率と長期間変化率の測定結果

音声波を音源と声道の特性を分離する逆フィルタに通した場合と通さない場合について、1～12次のPARCOR係数からなるベクトル、各PARCOR係数、逆フィルタの係数 β 、 γ 、および基本周波数の時期的変化率と長期間変化率を調べた。

4.1 時期的変化率

図2はサンプルとして単語を用い、逆フィルタに通さない場合の、1～12次のPARCOR係数(有声音区間の平均値)からなるベクトルの時期的変化率 $V(\tau)$ を、2単語と9名の話者について平均して示したものである。 $\tau < 6$ か月の範囲では指数関数的に増加し、 $\tau \geq 6$ か月の範囲ではほぼ一定となる性質が見られる。図中の破線は指数関数近似を行った結果である。

図3は、1次系逆フィルタの係数 β (方法I)の時期的変化率を示したものである。上と同様に、2単語と9名の話者について平均したものを示した。図2のベクトル全体の変化率に比べて、音源スペクトルの概形を示すパラメータ β の変化率がかなり大きいことがわかる。

4.2 長期間変化率

図4は、1～12次のPARCOR係数からなるベクトルの長期間変化率を、おのおのの母音について示したものである。逆フィルタを通さない場合と、種々の逆フィルタ

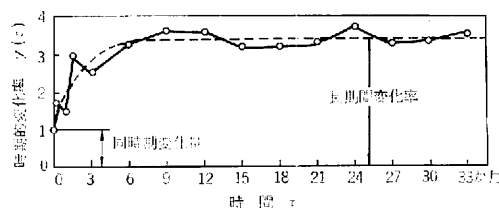


図2 単語の有声音区間における基本周波数とPARCOR係数の平均値からなるベクトルの時期的変化率(2単語平均結果)
--- 近似指数曲線

タを通した場合の結果を比較して示した。これからわかるように、音源特性を除去して近似的に声道特性のみにより、すべての母音について長期間変化率が低下し、長期間変化率の値は音韻によらずほぼ一定となる。この効果は、音源特性の近似法がかわってもほとんど変わらない。

音源特性を除去しないときには、音韻による長期間変化率の差が著しく、その内容を詳細に調べると、ホルマンの拡散の度合いの大きい音韻、あるいは調音点が口唇

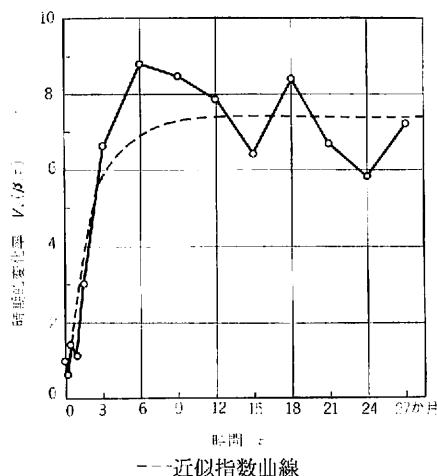
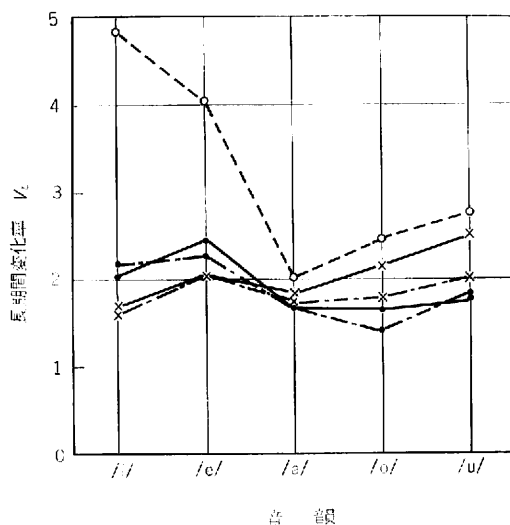


図3 1次系逆フィルタの係数 β (方法I)の時期的変化率(2単語平均結果)



音 韻
— × — 方法I (1次系近似)
-- × -- " (2次CD系近似)
— ● — 方法II (1次系近似)
-- ● -- " (2次CD系近似)
--- ○ --- 逆フィルタなし

図4 5母音の長期間変化率

に近い音韻において長期間変化率が大きくなっている。これは音源スペクトルの寄与が声道の形によって異なることを示していると考えられるが、理由は明らかではない。

図5はサンプルとして2単語を用いたときの、1~12次の各PARCOR係数の長期間変化率を示したものである。方法Ⅱによる逆フィルタを適用した場合と、逆フィルタを通さない場合の結果を示した。逆フィルタを通すこと

表1 パラメータ別長期間変化率

サンプル	逆フィルタ	パラメータ	長期間変化率	
			方法Ⅰ	方法Ⅱ
5母音	1次系	β	5.7	27.8
		f_0	4.6	4.6
		$K_1 \sim K_{12}$	2.0	1.9
	2次CD系	γ	4.2	24.8
2単語	1次系	β	6.9	27.8
		f_0	4.9	4.9
		$K_1 \sim K_{12}$	2.6	2.1
	2次CD系	γ	—	24.8
		f_0	4.9	4.9
		$K_1 \sim K_{12}$	—	2.0

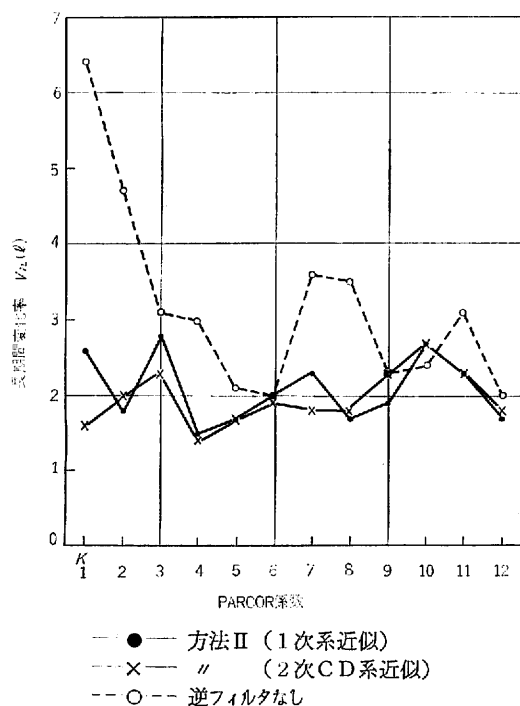


図5 各PARCOR係数の長期間変化率
(2単語平均結果)

によって全体に長期間変化率が低下しており、とくに1~2次の係数の低下が著しい。

表1はパラメータ別の長期間変化率を整理して示したものである。いずれの場合も、声道特性に因与する $K_1 \sim K_{12}$ の長期間変化率に比べて、音源特性に因与する β , γ , f_0 などの長期間変化率が大きくなっている。このことは、声道の特性よりも音源の特性の方が、時期的にゆっくり変化する傾向、すなわち時間の経過とともにばらつきが増大する傾向が強いことを示している。サンプルとして単独5母音を用いた場合の結果と、単語を用いた場合の結果はよく似ている。

以上の図3~図5および表1の結果は、音声波(スペクトル)の時期的変動の大きな部分が音源スペクトルの変動に起因していることを示している。

5 分散分析

5.1 パラメータ別分散分析

2単語の音声サンプルに関して、方法Ⅱによる逆フィルタを適用したときの β , γ , f_0 , K_i の各パラメータ別に話者要因に関する分散分析を行った。ほぼ10日間にわたる4時期のサンプルと、ほぼ10か月間にわたる4時期のサンプルについて分析を行い、サンプルの収集期間の長さの効果を調べた。

危険率1%のときの有意水準値 F_0 を1としたときの話者効果の大きさ(F/F_0)を図6に示す。音源に関するパラメータ β , γ , f_0 の結果を(a)に、2次CD系の逆フィルタを通したときの声道に関するパラメータ $K_1 \sim K_{12}$ の結果を(b)に示した。いずれも2単語に関する結果を平均したものである。これらの結果からつぎのことがわかる。すなわち、音源スペクトルの特性を示す β , γ は、10日間程度の短期間ではきわめて大きい話者効果を有するが、10か月程度の長期間では、変動量が増大するために話者効果は相対的に小さくなり、1/8程度に低下してしまう。一方、声道の特性を示すPARCOR係数 $K_1 \sim K_{12}$ ではこのような性質はあまりなく、長期間については、 β および γ とPARCOR係数とはほぼ同程度の話者効果を有する。

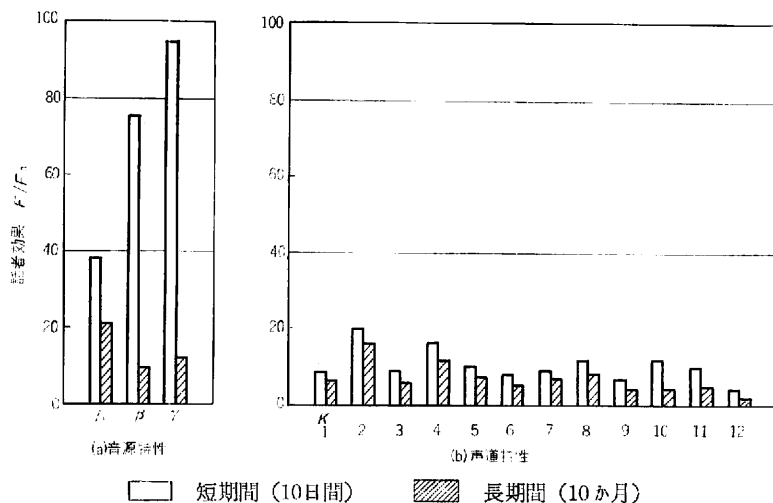
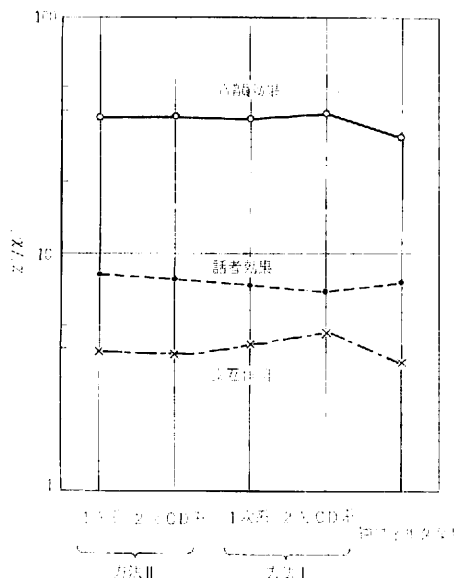


図 6 サンプルの期間と話者効果の関係

図 7 PARCOR 係数 $K_1 \sim K_{10}$ からなるベクトルの多変量分散分析結果

5.2 多変量分散分析

逆フィルタを通した場合と通さない場合の PARCOR 係数 $K_1 \sim K_{10}$ からなるベクトルについて、音韻要因、話者要因、および音韻—話者交互作用に関する多変量分散分析を行った。分析には、ほぼ3年間にわたって男性9名が発声した単独5母音のサンプルを用いた。危険率1%のときの有意水準値 $\chi^2_{0.01}$ を1としたときの各効果の大きさ ($\chi^2/\chi^2_{0.01}$) を図7に示す。この結果から、音韻効果、話者効果ともに、音源特性を除去してもほ

んど大きさが変わらないことがわかる。

6 話者認識実験

これまでの分析結果から、音声波に含まれる放射および音源特性を近似的に除去する逆フィルタを用いて、声道の特性のみを抽出すれば、パラメータが時期とともに変動する性質が少なくなり、しかも含まれる個人性情報はほとんど減少しないことがわかった。ここでは、このことを話者認識実験によって確かめる。すなわち、一般に話者認識においては、第1節でも述べたとおり、各話者の学習サンプルが1時期あるいは10日間程度の短期間のものである場合には、未知サンプルまでの時間間隔が3か月程度以上になると、急激に認識率が低下するが、本論文で用いている方法によって放射および音源特性を近似的に除去すれば、長期間たった後の未知サンプルに対する認識率の低下をおさえられることが期待される。

そこで、単語の有声音区間における基本周波数と1~12次の PARCOR 係数の統計量(平均値、標準偏差、および相関行列)を特徴パラメータとして用いる認識方法⁽⁴⁾⁽⁵⁾によって話者認識実験を行い、逆フィルタの効果を定量的に調べた。標準偏差および相関行列成分からは、学習サンプルを用いた分散分析によって20成分を選択し、平均値とともに特徴ベクトル成分として用いた。単語としては /baŋgo:/, /namae:/ の2種類を用い、1単語による方法と、2単語を組合せる方

法⁽⁴⁾による認識実験を行った。

6.1 短期間の学習サンプルによる認識 -1 ——音源特性を除去する効果——

図8に示すように、2～3日おきの3～4時期の学習サンプルを用いて、最終の学習サンプルから2～3日後、3週間後、6週間後、3か月後、および10か月後のサンプルの認識を試みた。方法Ⅰ～Ⅲの逆フィルタを適用したときの話者照合許(称者8名)の結果を、逆フィルタを適用しない場合の結果とあわせて図9～図11に示す。

これらの結果から、時期的にゆっくり大きく変化する傾向の強い音源スペクトル特性を除去することにより、認識誤りを1/2程度にへらすことができること、すなわちパラメータの時期的変動をかなりおさえることができることがわかる。方法Ⅰ、Ⅱ、およびⅢの結果を比較すると、これら

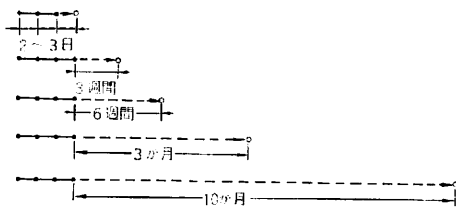
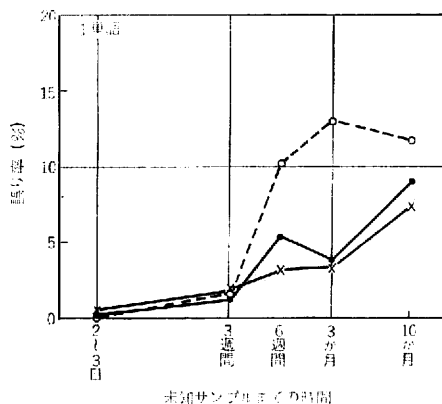


図8 短期間の学習サンプルを用いた認識における学習サンプルと未知サンプルの関係

- 学習サンプル
- 未知サンプル



の差は小さいが、方法ⅠよりもⅡあるいはⅢの方が比較的良好な結果を示している。

6.2 長期間の学習サンプルによる認識

図12に示すように、3か月おきの4時期のサンプルを学習サンプルとして、その最終のサンプルから3か月後のサンプルを認識する場合について、上と同様の実験を行った。その結果は図13に示すとおりで、この場合は、逆フィルタを通した場合と通さない場合の認識率の差はほとんどない。これは、このような10か月程度の長期間では、パラメータの時期的変動がほぼ不規則変動と見なせるためと、長期間では音源スペクトルの有する話者効果(個人性情報)がかなり小さいためと考えられる。この結果は、図7の分散分析結果と対応している。

6.3 学習サンプルが短期間のものである場合と長期間のものである場合の比較

図14に示すように、共通の6時期の未知サンプル(A～F)を用いて、学習サンプルが10日間程度の短期間のものである場合〔短期間〕と、10か月程度の長期間のものである場合〔長期間〕の比較を行った。逆フィルタの方法としては、方法Ⅱを用いた。〔短期間〕の場合は、学習サンプルを固定して未知サンプルまでの時間を3か月から21か月まで変え、〔長期間〕の場合は、未

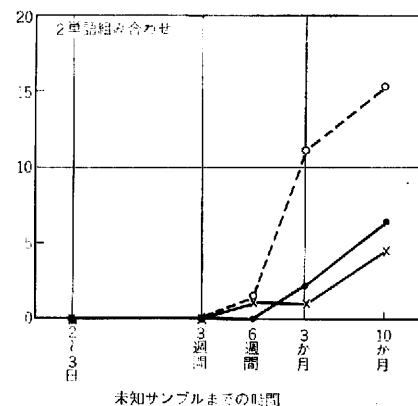


図9 短期間の学習サンプルによる話者照合(方法Ⅰによる逆フィルタの効果)

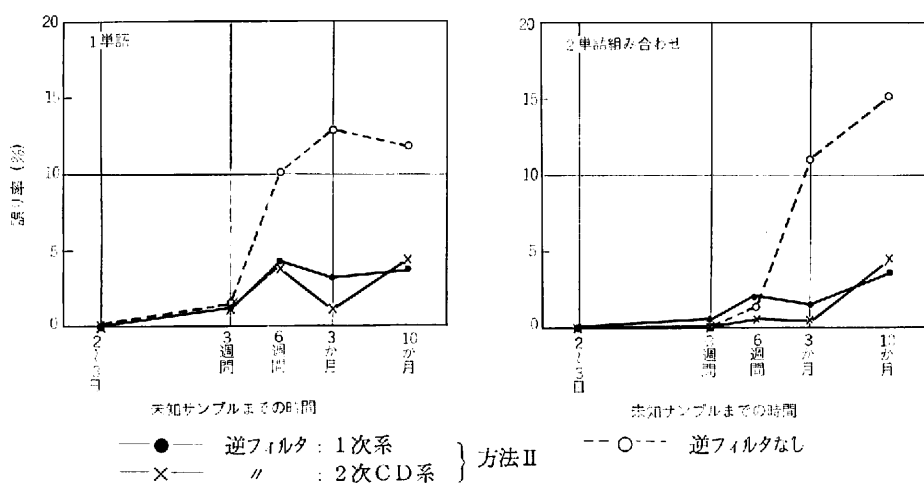


図 10 短期間の学習サンプルによる話者照合 (方法IIによる逆フィルタの効果)

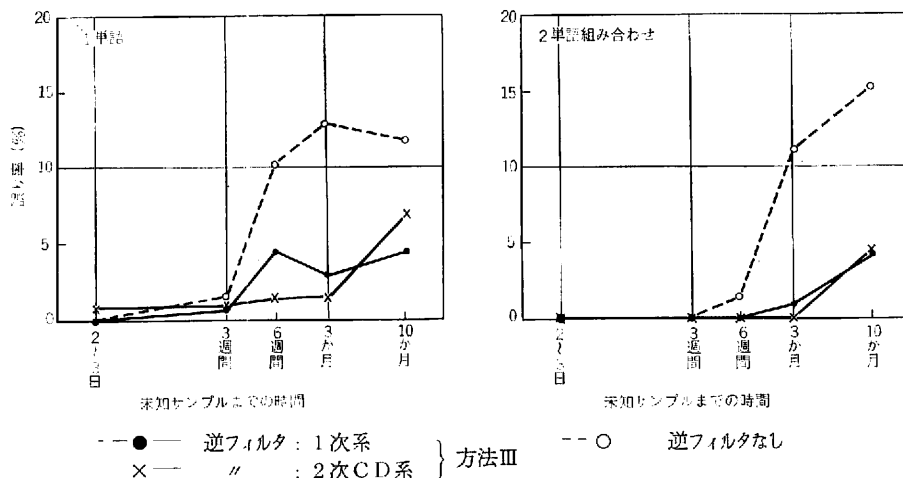
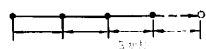


図 11 短期間の学習サンプルによる話者照合 (方法IIIによる逆フィルタの効果)



● 学習サンプル ○ 未知サンプル
図 12 長期間の学習サンプルを用いた認識における学習サンプルと未知サンプルの関係

知サンプルまでの時間を3か月に固定して、学習サンプルを変えた。なお学習サンプルが10か月程度の長期間のものなら、未知サンプルまでの時間が3か月から12か月まで変化しても、認識率はほとんど変わらないことが、逆フィルタを通さない場合について、すでに確認されている⁽⁴⁾。

認識実験の結果、認識方法によらず、〔短期間〕で

未知サンプルまでの時間を3か月から21か月まで変えても、単調に認識誤りが増加するような傾向はないことがわかったので、〔短期間〕〔長期間〕いずれの場合も、これらの6時期の未知サンプルに関する認識結果を平均して表2に示した。この結果から、学習サンプルが短期間のものであっても、音源スペクトル特性を除去して、2単語を組み合わせる方法を用いることにより、99%の認識率(表2の下線部分)が得られることがわかる。この認識率は、長期間の学習サンプルを用いる場合の結果と同程度であり、ここで用いた方法が、話者認識におけるパラメータの変動の問題に対処し学習サンプルを収集する期間を短縮するという点で、きわめて有効であることが確認された。

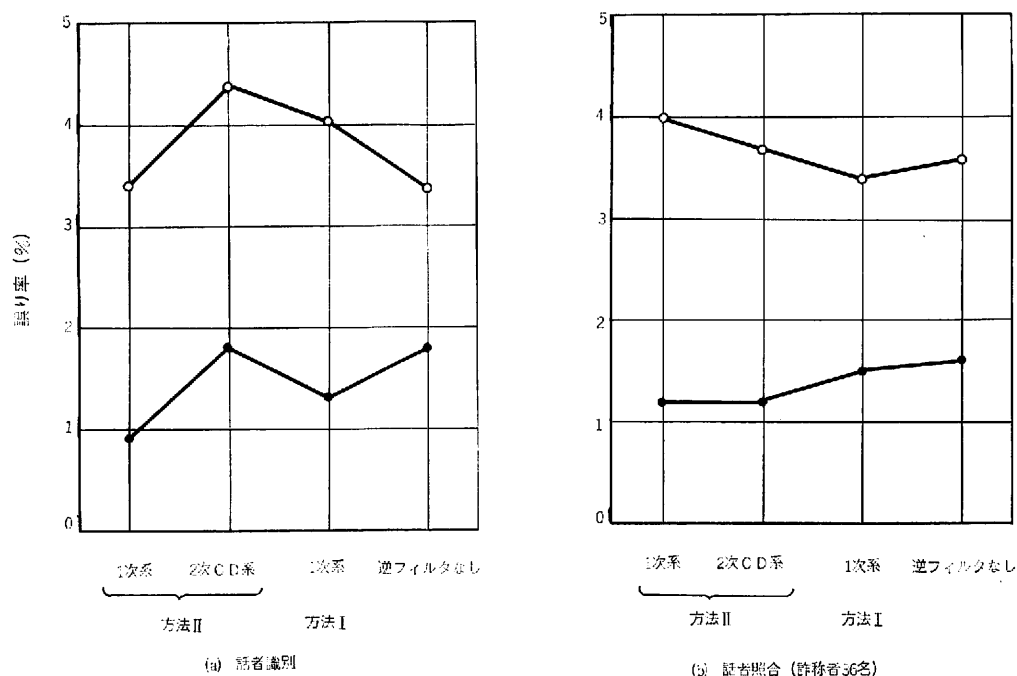


図 13 長期間の学習サンプルによる話者認識

表 2 学習サンプルが短期間の場合と長期間の場合の認識率

(逆フィルタ：方法Ⅱ)

学 習 サ ャ プ ル	未知サンプル までの時間	逆 フ ィ ル タ	話 者 識 別 (9 名)		話者照合(詐称者 8 名)	
			1 単 語	2 単 語 組 合 せ	1 単 語	2 単 語 組 合 せ
短 期 間 (10 日 間)	3 ~ 21 か 月	1 次 系	90.8%	99.4%	97.4%	99.0%
		2 次 CD 系	92.3	99.4	97.2	98.9
		な し	88.3	94.5	92.7	92.8
長 期 間 (10 か 月)	3 か 月	1 次 系	97.2	99.4	97.6	99.5
		2 次 CD 系	96.0	98.8	97.1	98.8
		な し	98.2	99.4	97.9	99.5

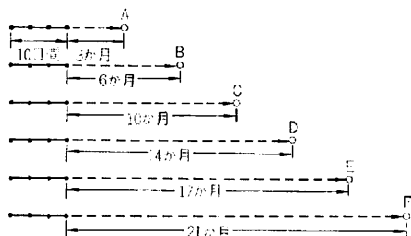
6.4 短期間の学習サンプルによる認識-2 ——学習サンプルの数の効果——

上の実験では、短期間の学習サンプルとして2~3日おきのほぼ10日間にわたる4時期のサンプルを用いたが、長期間たったのちの未知サンプルを良い認識率で認識するために、何時期程度の学習サンプルが必要かを調べるため、つぎのような補足的認識実験を行った。すなわち、学

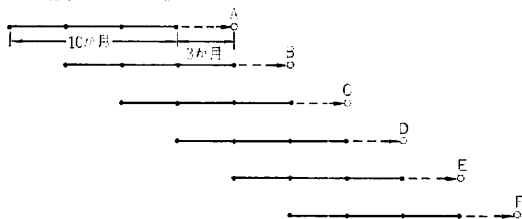
習サンプルから未知サンプルまでの時間間隔は、図14(a)と同様に3か月から21か月までとし、学習サンプルを求める時期(の数)を1時期から2~3日おきの4時期までの範囲で変えて、認識率の変化を調べた。逆フィルタの方法としては、方法Ⅱ(2次CD系近似)を用いた。

結果は図15のとおりで、学習サンプルが1時期のみのときには誤り率が極めて大きく、2時期以上の場合との

(a) 学習サンプルが短期間のものである場合

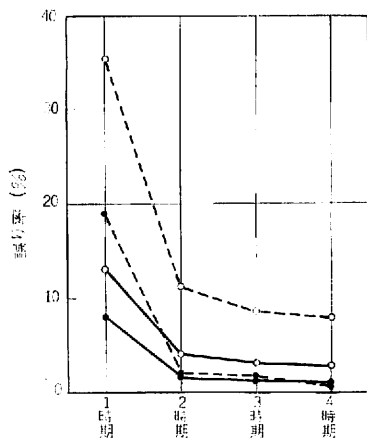


(b) 学習サンプルが長期間のものである場合



● 学習サンプル ○ 未知サンプル

図 14 学習サンプルが短期間のものである場合と長期間のものである場合の学習サンプルと未知サンプルの関係



学習サンプルの数 (2 ~ 3日おき)

- 話者識別 (1 単語)
- " (2 単語組み合わせ)
- 話者照合—詐称者 8 名— (1 単語)
- " (2 単語組み合わせ)

図 15 学習サンプル (短期間) の数と認識誤り率の関係 (方法 II, 2 次 CD 系)

間には大きな差があるが、3 ~ 4 時期で飽和する傾向が見られる。この結果から、短期間の場合でも、4 時期程度のサンプルを用いることが、ほぼ必要かつ十分であることがわかる。

6.5 逆フィルタの係数 β, γ を特徴ベクトルに加えた場合

音源スペクトルの概形を近似するパラメータ β および γ の有する個人性情報を評価するため、これらのパラメータを特徴ベクトルに加えて認識実験を行った。その結果、これらを用いる効果はほとんどなく、学習サンプルが短期間のものである場合には、逆に認識率がやや低下する傾向が見られ、学習サンプルが長期間のものである場合には、ほとんど認識率に差は見られなかった。これは主として、これまでの分散分析や認識実験ですでに明らかになっているように、長期間では β や γ の変動がかなり大きく、これらのパラメータの有する個人性情報がかなり少ないためと考えられる。

6.6 考 察

これまでの話者認識実験により、音声波に含まれる放射および音源スペクトル特性を近似的に除去する逆フィルタを用いることが、音声の特徴パラメータの時期的な変動量をへらし、安定した個人性情報を抽出するうえで、きわめて有効であることが確認された。

逆フィルタの方法 I ~ III の中では、方法 II または III が比較的良好な結果を示し、この両者の差はほとんどない。方法 I では、声道の周波数伝送特性の概形がほぼ平坦であることを仮定しているが、この仮定は、本研究のように 0 ~ 3.4 kHz 程度の比較的狭い帯域内で考えているときには必ずしも十分に満たされず、いわゆるホルマントなどの影響をかなり受けるために、この方法で求まる音源特性は、音韻や分析フレームごとにより変化してしまう。このことが認識実験の際に悪影響をおよぼしていると考えられる。

方法 II にはこのような欠点はないが、分析するサンプル以外に長時間のサンプルを用意しなければならないという問題点があり、話者認識という応用を考えた場合には、障害となる可能性がある。この点で方法 III は、認識に用いる単語の平均スペクトルで逆フィルタをすればいいうえ、方法 II とほとんど変わらぬ結果が得られるので、もっとも現実的で好ましい方法であると考えられる。

音源と声道の特性を近似的に分離するという意味で

は、方法Ⅱがもっとも妥当であると思われるが、方法ⅡとⅢにほとんど差がないのは、ここでは音源の特性としてスペクトルの全体的な特徴のみを用いているためであろう。

7 む す び

音波スペクトルに含まれる情報を、音声生成のメカニズムに基づいて、音源特性に起因するものと声道特性に起因するものとに近似的に分離し、それぞれの特性を表現するパラメータの時期的変動に関する分析を行った。まず、個人性パラメータの時期的変動の大きさを定量的に評価する尺度として、時期的変化率および長期間変化率を定義し、各パラメータについて、その値を求めた結果、声道特性よりも音源特性の方が、時期的にゆっくり大きく変化する傾向が強いことがわかった。つぎに話者要因に関する分散分析を行った結果、音源特性の話者効果は短期間では大きい、長期間になると変動量が增大するために相対的にかなり小さくなることが明らかになった。

これらの結果に基づいて話者認識実験を行い、放射および音源特性を近似的に除去する逆フィルタを用いることが、短期間の学習サンプルに基づいて、長期間たったのちにも安定な個人性情報を抽出する上で、きわめて有効であることを確認した。ほぼ10日間にわたる学習サンプルに基づいて、3～21か月後の未知サンプルを認識したところ、2単語の情報を組合せれば、話者識別と話者照合のどちらの場合も、99%の認識率が得られた。

上の結果は、スペクトルバタンの時期的変動は主としてその全体的な形の変動に起因し、やや細かいバタンに安定した特徴が保存されることを示している。これは長時間平均スペクトルの検討の際に得られた結果⁽²⁾と一致している。短期間では安定した個人差を有していても、長期間で大きく変化するようなパラメータは、真に個人性を表現しているとはいえない。すなわち、音源スペクトル特性の個人差は、あまり本質的ではないと考えられる。聴覚的な実験でも、音源の個人差を除去しても高い話者認識率の得られることが報告されている⁽⁹⁾。

発声時の声道の形状については、千葉、梶山⁽¹⁰⁾、Fant⁽¹¹⁾らがレントゲン写真を用いて行った観測結果などがすでに報告されている。ここで検討した方法が、音源お

よび放射特性と声道特性を分離するという意味で妥当であるかどうかを簡単に調べるため、方法Ⅱによって音源および放射特性を除去したのちに得られるPARCOR係数から、声道断面積関数を求めたところ、上記の観測結果とよく似ていることが確かめられた。その詳細については別の機会に報告する。今後さらに、より適切な声道情報と音源情報の分離方法の検討と、それに基づく個人性情報の分析を進めて行く予定である。

謝 辞

日ごろ、ご指導ごべんたついただく武蔵野電気通信研究所基礎研究部野田部長、斎藤特別研究室斎藤室長、ならびに熱心に討論していただく斎藤特別研究室の諸氏に感謝の意を表する。

文 献

- (1) 斎藤：話者認識，数理科学，No. 116，p. 49，1973.
- (2) 古井・板倉・斎藤：長時間平均スペクトルによる話者認識，信学誌，55-A，No. 10，p. 549，1972.
- (3) 古井・板倉・斎藤：音声の長時間平均スペクトルに含まれる個人性情報，通研実報，23，No. 6，p. 1199，1974.
- (4) 古井・板倉：単語の統計的パラメータによる話者認識，信学誌，56-A，No. 11，p. 717，1973.
- (5) 古井・板倉：音声スペクトルの統計的パラメータに含まれる個人性情報，通研実報，24，No. 5，p. 959，1975.
- (6) 古井：音声の個人性パラメータの時期的変動と話者認識，信学誌，57-A，No. 12，p. 880，1974.
- (7) T. Nakajima, et al.: Estimation of Vocal Tract Area Functions by Adaptive Inverse Filtering Methods, Bul. Electrotech. Lab., 37, No. 4, 1973.
- (8) 板倉：音声の特徴パラメータによる通信，昭47連大，No. 219，1972.
- (9) R. O. Coleman: Speaker Identification in the Absence of Inter-Subject Differences in Glottal Source Characteristics, J. Acoust. Soc. Amer., 53, No. 6, p. 1741, 1973.
- (10) T. Chiba and M. Kajiyama: The Vowels Its Nature and Structure, 1942, Tokyo-Kaiseikan Pub. Co., Ltd.

- (11) G. Fant : Acoustic Theory of Speech Production, 1960, Mouton & Co.

(1975, 8, 28 受付)