

論文 / 著書情報
Article / Book Information

論題(和文)	音声の個人性パラメータの時期的変動と話者認識
Title(English)	
著者(和文)	古井貞熙, 板倉文忠
Authors(English)	SADAOKI FURUI
出典(和文)	電子情報通信学会論文誌, Vol. J57-A, No. 12, pp. 880-887
Citation(English)	, Vol. J57-A, No. 12, pp. 880-887
発行日 / Pub. date	1974,
URL	http://search.ieice.org/
権利情報 / Copyright	本著作物の著作権は電子情報通信学会に帰属します。 Copyright (c) 1974 Institute of Electronics, Information and Communication Engineers.

音声の個人性パラメータの 時期的変動と話者認識

正 員 古井 貞熙†

An Analysis of Long-term Variation of Feature Parameters of Speech and Its Application to Talker Recognition

Sadaaki FURUI†, Regular Member

あらまし 音声の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、時期的変化率を提案し、音声波から音源特性と声道特性を近似的に分離して、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(スペクトル)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。次に話者要因に関する分散分析を行った結果、音源特性が有する個人性は、短期間では大きい、長期間になるとかなり小さくなるということが明らかになった。これらの結果に基づいて、音源特性を除去したのちのパラメータだけを用いて話者認識実験を行い、既知サンプルが短期間のものである場合に、この方法が有効であることを確認した。ほぼ10日間にわたる既知サンプルを用いて、3~21か月後の未知サンプルを認識したところ、2単語を組合わせれば、話者識別、話者照合のどちらの場合も99%の認識率が得られることがわかった。

1. ま え が き

一般に話者認識の問題を取り扱う上で最も困難な点の一つは、ある程度長い期間をおくと特徴パラメータが変化してしまうところにある。筆者らは、既に音声の長時間平均スペクトルの長期間にわたる変動特性を明らかにするとともに⁽¹⁾、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)⁽⁹⁾の統計量を用いた話者認識方法を提案し^{(2),(6)}、3か月おきの4時期程度の既知サンプル(学習サンプル)を用いれば、その最終の既知サンプルから3か月後のサンプルに関する話者識別および話者照合において、99%の認識率が得られることを示した。この結果は、必ずしも1年間程度の既知サンプルの必要性を示すものではないが、既知サンプルが10日間程度の短期間のも

のである場合には、3か月程度の時間が経過すると、未知サンプルに対する認識率は急激に低下する。このことは特徴パラメータのゆっくりした時期的変動特性に起因する^{(3),(4)}。

各登録話者について、長期間の既知サンプルを用意しなければならないということは、実際の話者認識システムを構成する際には好ましいことではなく、又、時期がたつと大きく変わるパラメータは、真に個人性を表現しているとはいえない。本論文では特徴パラメータの変動特性を音声生成のメカニズムに基づいて分析し、その結果をもとに、既知サンプルを求める期間を短縮するための方法を検討する。

2. 時期的変動特性の評価法

静かな部屋で発声された音声を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、PARCOR係数と基本周波数を求める。

† 電電公社武蔵野電気通信研究所基礎研究部、武蔵野市
Musashino Electrical Communication Laboratory,
N.T.T., Musashino-shi, 180 Japan
論文番号: 昭 49-488[A-156]

話者 i ($1 \leq i \leq 9$) が時期 j ($1 \leq j \leq n$; 間隔は 3 か月) に繰返し 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} で表す。これは、単語音声の有声音区間における PARCOR 係数と基本周波数の平均値からなるベクトルである。

$$X_{ijk} = (X_{ijk1}, X_{ijk2}, \dots, X_{ijkN}) \quad (1)$$

同じ時期の 3 サンプルの平均値を $X_{ij\cdot}$ で表す。

$$X_{ij\cdot} = (X_{ij\cdot 1}, X_{ij\cdot 2}, \dots, X_{ij\cdot N}) = \frac{1}{3} \sum_{k=1}^3 X_{ijk} \quad (2)$$

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 (の 2 乗) を、 n 時期および 9 名の話者について平均したものを同時期変化量と定義し、次のように表す。 $D_{0L}(l)$ 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。

$$D_{0L}(l) = \frac{1}{27n} \sum_{i=1}^9 \sum_{j=1}^n \sum_{1 \leq k < l \leq 3} (X_{ijkL} - X_{ijlL})^2 \quad (3)$$

$$D_0 = \sum_{l=1}^N D_{0L}(l) \quad (4)$$

同様に、 $(\tau \times 3)$ か月の間隔をおいたサンプル間の距離を次式のように求め、時期的変化量と定義する。

$$D_L(l, \tau) = \frac{1}{9(n-\tau)} \sum_{i=1}^9 \sum_{j=1}^{n-\tau} (X_{ij+\tau, L} - X_{ij, L})^2 \quad (5)$$

$$D(\tau) = \sum_{l=1}^N D_L(l, \tau) \quad (6)$$

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、次のように表す。

$$V_L(l, \tau) = D_L(l, \tau) / D_{0L}(l) \quad (7)$$

$$V(\tau) = D(\tau) / D_0 \quad (8)$$

時期的変化率を用いるパラメータの種類にほとんどよらずに、 $\tau \geq 2$ の範囲でほぼ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 で表す⁽⁵⁾。

$$V_{iL}(l) = \frac{1}{n-\tau_0} \sum_{\tau=\tau_0}^{n-1} V_L(l, \tau) \quad (9)$$

$$V_L = \frac{1}{n-\tau_0} \sum_{\tau=\tau_0}^{n-1} V(\tau) \quad (\tau_0 = 2) \quad (10)$$

話者認識に用いるパラメータとしては、時期的変化率 (長期間変化率) が小さいものが好ましい。

3. 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られる。この関係を周波数領域で表示すれば、

$$S(\omega) = R(\omega)T(\omega)C(\omega)E(\omega) \quad (11)$$

と書くことができる。但し $S(\omega)$, $E(\omega)$, $C(\omega)$, $T(\omega)$, $R(\omega)$ 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 $s(t)$ 及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り出すことは極めて困難であるが、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、次のような方法で音源と声道の特性の分離を試み、これに基づいてパラメータの時期的変動特性の解明を行う。

〔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ーンの概形は、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下に適応的に (分析フレームごとに) 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いると考えられる。この方法では、長時間平均スペクトルパターンから、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。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⁽¹⁾をもとに、約 10 秒長の短文章を発声したものをを用いる。

〔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ーンから、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。

以上の三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、次のように表すことができる。

$$\left\{ \begin{array}{ll} \text{1次系:} & 1 + \beta Z \\ \text{2次CD系:} & 1 + \gamma Z + \frac{\gamma^2}{4} Z^2 \end{array} \right\} \quad (12)$$

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

方法Ⅰには、声道の周波数伝送特性の概形が平坦であるという仮定が、本研究のように0~3.4 kHz程度の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、ホルマントら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が、音韻や分析フレームごとにより変化してしまうという欠点がある。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る。

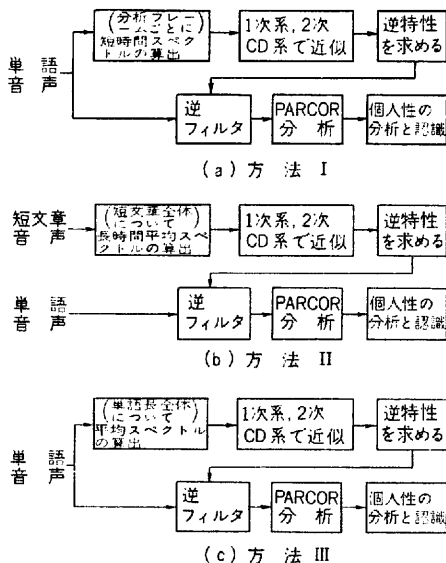


図1 分析方法
Fig.1-Block diagram of the analyzing method.

4. 時期的変化率と長期間変化率の測定結果

音声を音源特性を除去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係数および逆フィルタの係数 β , γ の時期的変化率と長期間変化率を調べた。分析は、男性9名が長期間にわたって発声した、単独5母音(/i/, /e/, /a/, /o/, /u/)と、2種類の単語(/bango/: /namae/)を用いて行った。

4.1 時期的変化率

図2は、サンプルとして単語を用いたときの、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。 $\tau < 2$ の範囲では指数関数的に増加し、 $\tau \geq 2$ の範囲ではほぼ一定となる性質が現れている。

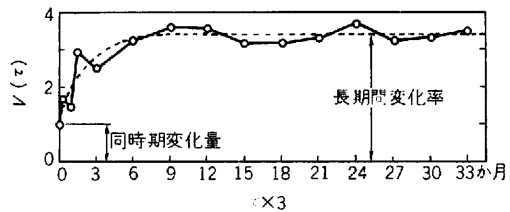


図2 単語の有声音区間における基本周波数とPARCOR係数の平均値からなるベクトルの時期的変化率(2単語平均結果)

Fig.2-Time-variation-rate of the vector which consists of fundamental frequency and PARCOR coefficients.

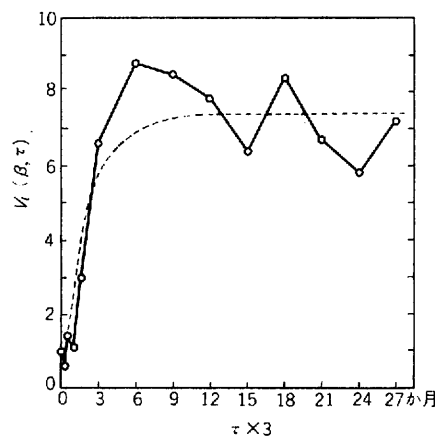


図3 1次系逆フィルタの係数 β (方法Ⅰ)の時期的変化率(2単語平均結果)

Fig.3-Time-variation-rate of coefficient β , parameter of 1st order filter.

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、各々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(逆フィルタ)して近似的に声道特性のみにすることにより、すべての母音について長期間変化率が大きく低下する。音源特性を近似する方法によるこの効果の差は小さい。

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、ホルマントの拡散の度合いの大きい音韻ほど、あるいは調音点が口唇に近い音韻ほど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ためと考えることができるが、明らかでない。

以上の図3～4の結果は、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。

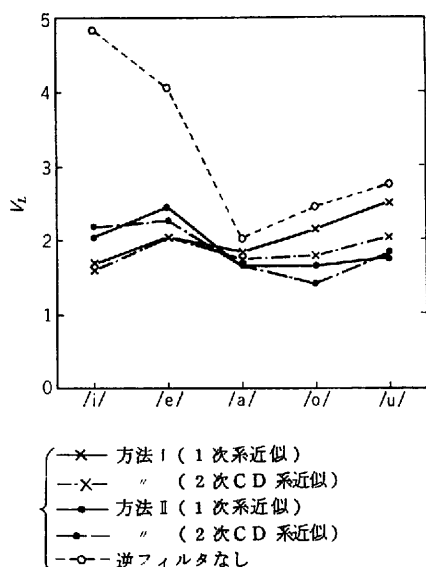


図4 5母音の長期間変化率

Fig.5-Long-term variation rate of five vowels.

表1は、5母音を分析したときの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、声道特性に關与する $K_1 \sim K_{12}$ の長期間変化率に比べて、音源特性に關与する β , r , f_0 (基本周波数)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共にばらつきが増大する傾向が強いことを示している。

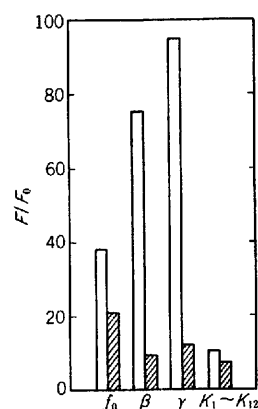
表1 パラメータ別長期間変化率
(5母音平均結果)

逆フィルタ	パラメータ	分析 方法	
		方法Ⅰ	方法Ⅱ
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	r	4.2	24.8
	f_0	4.6	4.6
	$K_1 \sim K_{12}$	1.9	1.9

5. 分散分析

5.1 パラメータ別分散分析

方法Ⅱによる逆フィルタを適用したときの各パラメータ別に、分析に用いるサンプルがほぼ10日間にわたる4時期のものである場合について、話者要因に関する分散分析を行った。分析には2単語を用い、危険率1%のときの有意水準 F_0 を1としたときの話者効果



□ 短期間(10日間) ▨ 長期間(10か月)

図5 サンプルの期間と話者効果の関係

Fig.5-Relations between talker effect and period for which samples are taken.

の大きさ (F/F_0) を、2 単語に関して平均して図 5 に示した。音源に関するパラメータとして、 β , γ , f_0 の結果を、声道に関するパラメータとして、2 次 CD 系の逆フィルタを通したのちに得られる 1~12 次の PARCOR 係数 $K_1 \sim K_{12}$ の結果を平均したものを示した。

これらの結果から次のことがわかる。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極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その効果は $1/8$ 程度に低下してしまう。一方 $K_1 \sim K_{12}$ ではこのような性質は余りなく、長期間については、 β 及び γ と PARCOR 係数とはほぼ同程度の話者効果を有する。

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、ほぼ 3 年間にわたって男性 9 名が発声した単独 5 母音を用いて、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。危険率 1% のときの有意水準 χ^2_0 を 1 としたときの各効果の大きさ (χ^2/χ^2_0) を図 6 に示す。この結果から、音韻効果、話者効果共に、音源特性を除去してもほとんど変わらないことがわかる。

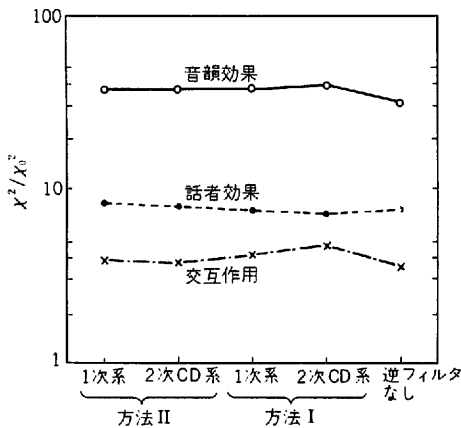


図 6 PARCOR 係数 $K_1 \sim K_{10}$ からなるベクトルの多変量分散分析結果

Fig. 6-Results of multi-variate analysis of variance for vector which consists of $K_1 \sim K_{10}$.

6. 認識実験

前述の結果から、既知サンプルが短期間のものではあ

る場合には、音源特性を除去することによって、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抑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、相関行列) を特徴パラメータとして用いる認識方法^{(2), (6)}によって、音源特性を除去する効果を調べた。標準偏差および相関行列成分からは、既知サンプルを用いた分散分析によって 20 成分を選択し、平均値と共に特徴ベクトル成分として用いた。音源および放射特性を除去したのちに得られる PARCOR 係数は、声道の形状と対応づけられる⁽⁹⁾から、このときに用いられるパラメータは、基本周波数と声道形状の、単語有声音区間における平均値および運動上の特徴を表現するものとなる。単語として /baŋgo:/, /namae/ の 2 種類を用い、1 単語による方法と、2 単語を組み合わせる方法⁽⁶⁾による認識実験を行った。

6.1 短期間の既知サンプルによる認識-1

—— 音源特性を除去する効果 ——

図 7 に示すように、2~3 日おきの 3~4 時期の既知サンプルを用いて、最終の既知サンプルから、2~3 日後、3 週間後、6 週間後、3 か月後、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合せて図 8 (a)~(c) 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特性を除去することにより、認識誤りを $1/2$ 程度に減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抑えることができることがわかる。方法 I, II, 及び III の結果を比較すると、これらの差は小さいが、方法 I よりも II あるいは III の方が比較的

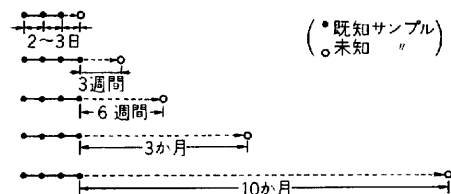
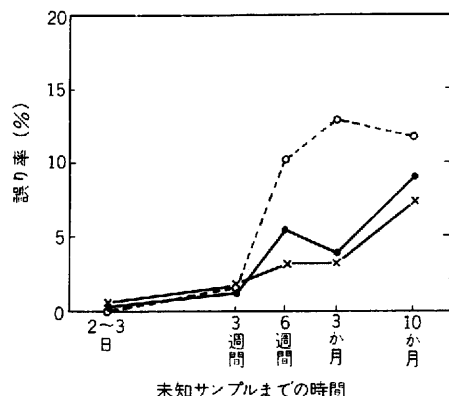
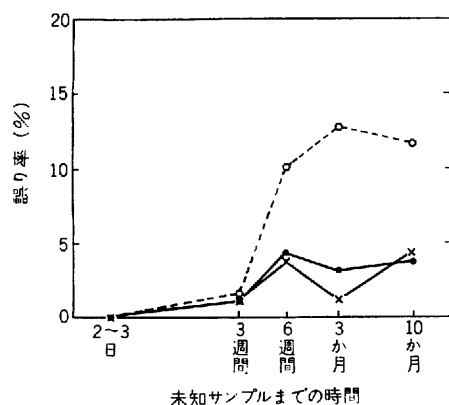


図 7 短期間の既知サンプルによる認識における既知サンプルと未知サンプルの関係

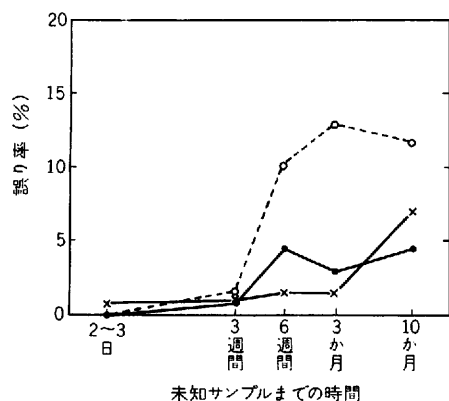
Fig. 7-Relations between reference samples and unknown samples in recognition experiment using short-term reference samples.



(a) 方法Ⅰによる逆フィルタの効果



(b) 方法Ⅱによる逆フィルタの効果



(c) 方法Ⅲによる逆フィルタの効果

—○— 逆フィルタ：1次系 —×— 逆フィルタ：2次CD系
- - - 逆フィルタなし

図8 短期間の既知サンプルによる話者照合
Fig.8-Results of talker verification using short-term reference samples.

よい結果を示している。

6.2 長期間の既知サンプルによる認識

図9に示すように、3か月おきの4時期のサンプルを既知サンプルとして、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10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みなせるためと、長期間では音源スペクトルの有する話者効果はかなり小さいためと考えられ、この結果は図6の分散分析結果と対応している。

6.3 既知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11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、既知サンプルが10日間程度の

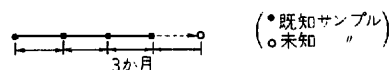
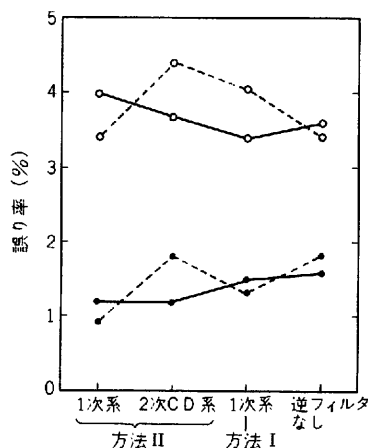


図9 長期間の既知サンプルによる認識における既知サンプルと未知サンプルの関係

Fig.9-Relations between reference samples and unknown samples in recognition experiment using long-term reference samples.



—○— 話者識別 (1 単語)
- - - " (2 単語組合せ)
—○— 話者照合 - 詐称者 36 名 - (1 単語)
- - - " (2 単語組合せ)

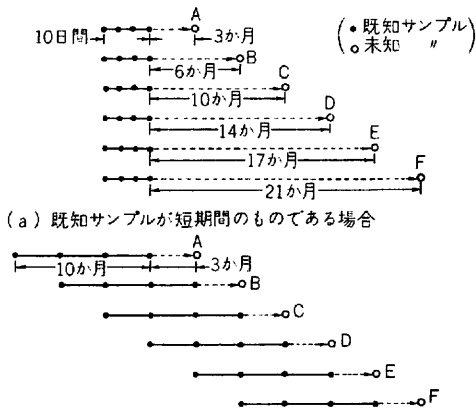
図10 長期間の既知サンプルによる話者認識
Fig.10-Results of talker identification and verification using long-term reference samples.

表2 既知サンプルが短期間の場合と長期間の場合との比較
(逆フィルタ: 方法Ⅱ)

既知 サンプル	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	1 単語	2 単語
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

短期間のものである場合(a)と、10か月程度の長期間の
ものである場合(b)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。(a)の場合は、既知サン
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、(b)の場合は、未知サンプルまでの
時間を3か月に固定して既知サンプルを変えた。なお
既知サンプルが10か月間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、既に確認されている⁽⁶⁾。

認識実験の結果、認識方法によらず、(a)で未知サン
プルまでの時間が3か月から21か月まで増加しても、
単調に認識誤りが増加する傾向はないことがわかった



(b) 既知サンプルが長期間のものである場合

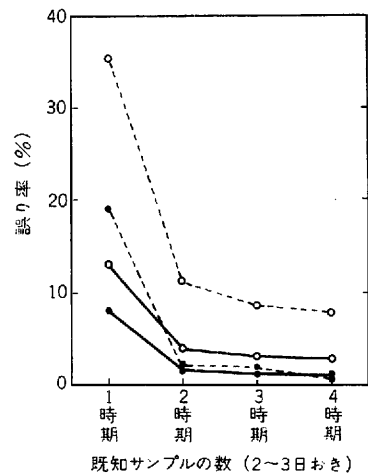
図 11 既知サンプルが短期間のものである場合と
長期間のものである場合の既知サンプルと
未知サンプルの関係Fig. 11-Relations between reference samples
and unknown samples in recognition
experiment about the effects of period
for which reference samples are taken.

ので、(a)、(b)いずれの場合も、これら
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、既知サンプルが短期間のもの
である場合でも、音源特性を除去し
て 2 単語を組合わせることにより、既
知サンプルが長期間のものである場
合同程度の 99 % の認識率 (表 2
の下線部分参照) が得られることが
わかり、ここで検討した方法の有効
性が確認された。

6.4 短期間の既知サンプルによる認識-2

— 既知サンプルの数の効果 —

上の実験では、短期間の既知サンプルとして 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、長期間たった後の未知サンプルを良い認識率
で認識するために、何時期程度の既知サンプルが必要
かを調べるため、次のような補足的認識実験を行った。
すなわち、既知サンプルから未知サンプルまでの時間
間隔は、図 11 (a) と同様に 3 か月から 21 か月までと
し、既知サンプルを求める時期を 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

既知サンプルの数 (2~3日おき)
---○--- 話者識別 (1 単語)
---□--- " (2 単語)
—○— 話者照合 (1 単語)
—□— " (2 単語)
未知サンプル: 3 ~ 21 か月後図 12 既知サンプル (短期間) の数と認識
誤り率の関係 (方法Ⅱ, 2 次 CD 系近似)Fig. 12-Relations between the number of short-
term reference samples and error rate.

近似)を用いた。

結果は図12のとおりで、既知サンプルが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 及び γ 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、既知サンプルが短期間のものである場合に、逆に認識率がやや低下する傾向が見られたほかは、ほとんど認識率に差は見られなかった。長期間では、図5に示したように、 β や γ の有する話者効果はかなり小さいと思われるが、他の声道に関するパラメータと独立でないためとも考えられる。

7. むすび

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”を提案し、音波から音源特性と声道特性を近似的に分離して、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、時期的にゆっくり変化する傾向が強いことがわかった。次に話者要因に関する分散分析を行った結果、音源特性が有する個人性は、短期間では大きい、長期間になるとかなり小さくなるということが明らかになった。これらの結果に基づいて、音源特性を除去したのちのパラメータだけを用いて話者認識実験を行い、既知サンプルが短期間のものである場合に、この方法が有効であることを確認した。ほぼ10日間にわたる既知サンプルを用いて、3～21か月後の未知サンプルを認識したところ、2単語を組み合わせれば、話者識別と話者照合のどちらの場合も99%の認識率が得られることがわかった。

上の結果は、長期間になるに従って、スペクトルパターンの全体的な形の変動が、比較的細かい構造の変動に比べて大きくなることを示しており、長時間平均スペクトルの検討の際に得られた結果⁽¹⁾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ず、聴覚的な実験でも、音源の個人性を除去しても高い話者認識率が得られるという結果が報告されている⁽⁸⁾。

発声時の声道の形状については、千葉、梶山⁽⁹⁾、Fant⁽¹⁰⁾らがレントゲン写真を用いて行った観測結果などが既に報告されている。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、それらの結果とよく似ていることが確かめられた。今後更に、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 日ごろ御指導ご鞭撻を頂く、野田基礎研究部長、斎藤第4研究室長、板倉調査員、ならびに熱心に討論して頂いた第4研究室的の諸氏に深く感謝する。

文 献

- (1) 古井, 板倉, 斎藤: “長時間平均スペクトルによる話者認識”, 信学論(A), 55-A, 10, p.549 (昭47-10).
- (2) 古井, 板倉: “スペクトルパターンの時間変動特性を考慮した話者認識”, 音響学会講演, 2-2-20 (昭47-10).
- (3) 古井, 板倉: “スペクトルパラメータによる話者認識におけるキーワード数と時期的変動の効果”, 昭48信学全大, 1258.
- (4) 古井, 板倉: “話者認識におけるスペクトルパラメータの時期的変動の効果”, 音響学会講演, 2-1-18 (昭48-05).
- (5) 古井, 板倉: “音声の個人性パラメータの時期的変動特性の解析”, 音響学会講演, 2-3-1 (昭48-10).
- (6) 古井, 板倉: “単語の統計的パラメータによる話者認識”, 信学論(A), 56-A, 11, p.717 (昭48-11).
- (7) T. Nakajima, et al.: “Estimation of vocal tract area functions by adaptive inverse filtering methods”, Bul. Electrotech. Lab., 37, 4 (1973).
- (8) R. O. Coleman: “Speaker identification in the absence of inter-subject differences in glottal source characteristics”, J. Acoust. Soc. Amer., 53, 6, p.1741 (1973).
- (9) 板倉: “音声の特徴パラメータによる通信”, 昭47連大, 219.
- (10) T. Chiba and M. Kajiyama: “The vowels its nature and structure”, Tokyo-Kaiseikan Pub. Co., Ltd. (1942).
- (11) G. Fant: “Acoustic theory of speech production”, Mouton & Co. (1960).

(昭和49年4月4日受付)