T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

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	
著者(和文)	古井 貞熙, 板倉 文忠
Author	SADAOKI FURUI
出典(和文)	資料番号S73-18,　, No. S73-18, pp. 1-10
Journal/Book name	,　, No. S73-18, pp. 1-10
発行日 / Issue date	1973, 11

Powered by T2R2 (Science Tokyo Research Repository)

昭和48年度科研費
音声総合研究委員会　資料
音響学会研究委員会
資料番号S73-18(1973-11)

音声の個人性パラメータの時期的変動と話者認識

古井貞熙・板倉文忠

(電電公社・武蔵野通研)

1973年11月20日

昭和48年度　音声の生成・知覚過程およびその
科研費　自動化に関する総合的研究委員会

日本音響学会　音声研究委員会

音声の個人性パラメータの時期的変動と話者認識

古井貞熙・板倉文忠（日本電信電話公社　武蔵野電気通信研究所）

昭和48年11月20日　　於　国際電電研究所

内容梗概：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，音声波から声帯波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，声道特性よりも声帯特性の方が，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なった結果，声帯波特性が有する個人性は，短期間では大きいが長期間になるとかなり小さくなることが明らかになった．これらの結果にもとづいて，声帯波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ない，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．

Analysis of the Long-term Variations of the Characteristic Parameters and Its Application to Talker Recognition

Sadaoki FURUI and Fumitada ITAKURA

(Musashino E.C.L., N.T.T.)

November 20, 1973　at K.D.D. Research Laboratory

Abstract: As a measure to evaluate the long-term variations of characteristic parameters of speech sounds, "time-variation rate" is proposed. Vocal cord wave and vocal tract characteristics are separated approximately from speech wave, and time-variation rate is measured for each parameter. From these results, it becomes clear that the long-term variational characteristics of vocal cord wave are more evident than that of vocal tract characteristics. Then, analysis of variance for talker factor is made, and its results reveal that the amount of individual character included in vocal cord wave is large for a short period and it becomes small for a long period. On the basis of these results, talker recognition experiments using parameters which ignore vocal cord characteristics are made, and it is ascertained that when the known sample is of short observations this recognition method leads to good results.

1. まえがき

筆者らはすでに，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）の統計量を用いた話者認識方法を提案し，[1][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．3か月程度以上の時間を経過した後の未知サンプルに対して良い認識率を得るために，それ以前のかなり長期間の既知サンプルが必要とされるということは，特徴ベクトルのゆっくりした時期的変動特性に起因する．その特性の概略については，すでに何回か報告したが，[3][4] ここではその特性を音声生成のメカニズムにもとづいて分析し，その結果を用いて，既知サンプルを求める期間を短縮することを試みた．

2. 時期的変動特性の評価法[5]

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 $\mathbf{X}_{ijk}$ であらわす．これは，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である

$$\mathbf{X}_{ijk}=(X_{ijk1}, X_{ijk2}, \cdots, X_{ijkN}) \tag{1}$$

同じ時期の3サンプルの平均値を $\mathbf{X}_{ij\cdot}$ であらわす．

$$\mathbf{X}_{ij\cdot}=(X_{ij\cdot 1}, X_{ij\cdot 2}, \cdots, X_{ij\cdot N})=\frac{1}{3}\sum_{k=1}^{3}\mathbf{X}_{ijk} \tag{2}$$

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する．

$$D_{0l}(l)=\frac{1}{27n}\sum_{i=1}^{9}\sum_{j=1}^{n}\sum_{\substack{1\leq k,k'\leq 3\\ k\neq k'}}(X_{ijkl}-X_{ijk'l})^2 \tag{3}$$

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

$D_{0l}(l)$ 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．

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．

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

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

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する．

$$V_l(l,\tau)=D_l(l,\tau)/D_{0l}(l) \tag{7}$$

$$V(\tau)=D(\tau)/D_0 \tag{8}$$

S-1-18

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 $\tau \geq 2$ 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．

$$V_{IL}(I)=\frac{1}{n-\tau_0}\sum_{\tau=\tau_0}^{n-1} V_I(I,\tau) \quad (\tau_0=2) \tag{9}$$

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

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. 音源と声道特性の抽出

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，

$$S(\omega)=R(\omega)T(\omega)C(\omega)E(\omega) \tag{11}$$

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．

実際に得られる音声波形 $s(t)$ 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，これにもとづいてパラメータの時期的変動特性の解明を行なう．

〔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．音声波形から放射特性を含めた音源特性を上の仮定のもとに適応的に推定

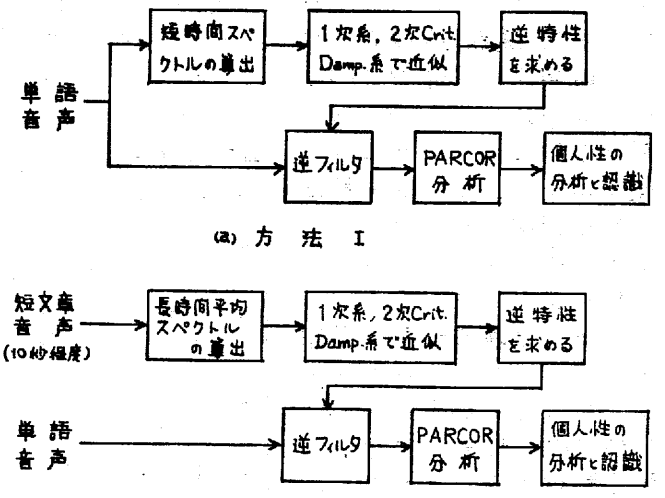


図 1. 分析方法

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] この方法には，上記の周波数特性が平坦であるという仮定が，本報告のように0～3.4KHz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，フォルマント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，音韻やフレーム毎にかなり変化してしまうという欠点がある.

〔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から，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.

2つの方法のブロックダイヤグラムを図1に示す．どちらの場合も，音源特性を1次系および2次Critical Damping系で近似した．逆フィルタの伝送関数は，次のように表わすことができる.

$$\begin{cases} 1\text{次系}: & 1+\beta Z \\ 2\text{次C.D.系}: & 1+\gamma Z+\frac{\gamma^2}{4}Z^2 \end{cases} \qquad \left(\begin{array}{l} Z=e^{-j\lambda} \\ \lambda=\text{radian/unit time} \end{array} \right)$$

4. 時期的変化率

音声波を，声帯波特性（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.

図2は，サンプルとして単語を用いたときの，1～12次のPARCOR係数からなるベクトルの時期的変化率 V(τ) である．/bango:/,/namae/ の2種類の単語について平均した結果を示した．τ<2の範囲では指数関数的に増加し，τ≧2の範囲でほぼ一定となる性質が表われている.

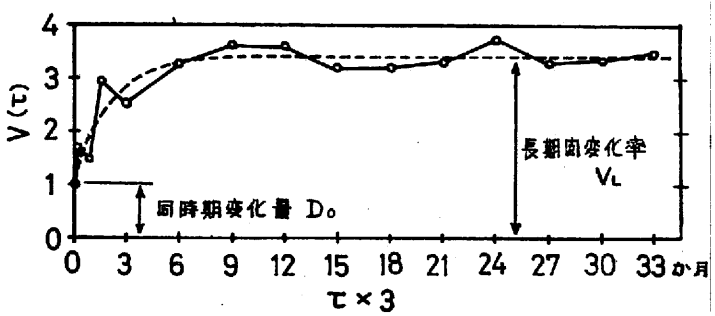
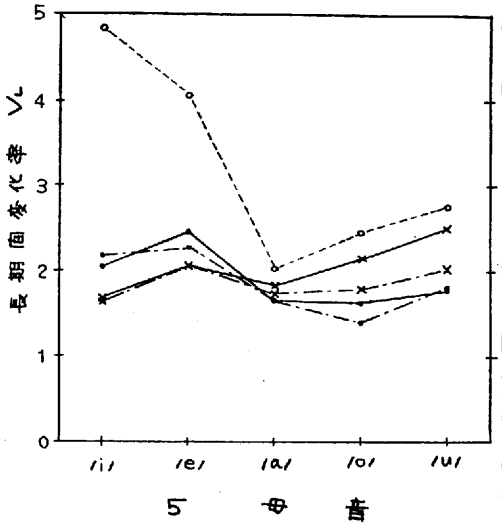


図2. 1～12次のPARCOR係数からなるベクトルの時期的変化率（2単語平均結果）

図3は，単独5母音の長期間変化率 V_L を示したものである．すべての母音について，声帯波特性を除去して近似的に声道特性のみにすることにより，長期間変化率が大きく低下することがわかる．この効果の，声帯波特性を近似す

図3. 5母音の長期間変化率



3方法による差は少ない．声帯波特性を除去しないときに，音韻による長期間変化率の差が著しい理由は，声帯波スペクトルの寄与が音韻によって異なるためと考えることもできるが，明らかではない．

図4は，1次系逆フィルタの係数βの時期的変化率である．2単語に関する結果を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，声帯波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．

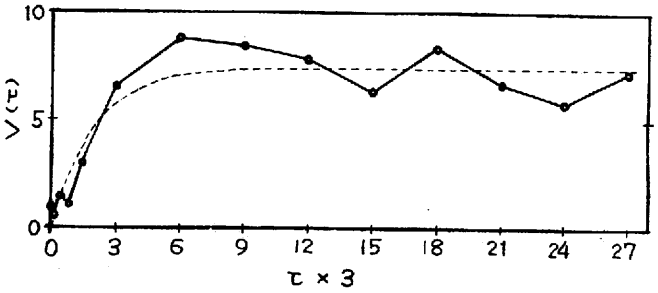


図4．1次系逆フィルタの係数β（方法I）の時期的変化率（2単語平均結果）

図5は，方法IIを用いたときの，1～12次の各PARCOR係数の長期間変化率を，2単語について平均した結果である．逆フィルタに通すことによって，全体に長期間変化率が低下しており，特に1～2次の係数の低下が著しい．図3～5の結果は，時期的変動の大きな部分が，声帯波スペクトルの変動に起因していることを示している．

表1は，5母音を用い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，声道特性に関与するk_1～k_{12}の長期間変化率に比べて，音源特性に関与するβ，γ，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，声道の特性よりも声帯の特性の方が，時期的にゆっくり変化する傾向

表1．各パラメータの長期間変化率（5母音平均結果）

方法	近似系	パラメータ	長期間変化率
方法I	1次系	β	5.7
		f_0	4.6
		k_1～k_{12}	2.0
	2次 C.D.系	γ	4.2
		f_0	4.6
		k_1～k_{12}	1.9
方法II	1次系	β	27.8
		f_0	4.6
		k_1～k_{12}	1.9
	2次 C.D.系	γ	24.8
		f_0	4.6
		k_1～k_{12}	1.9

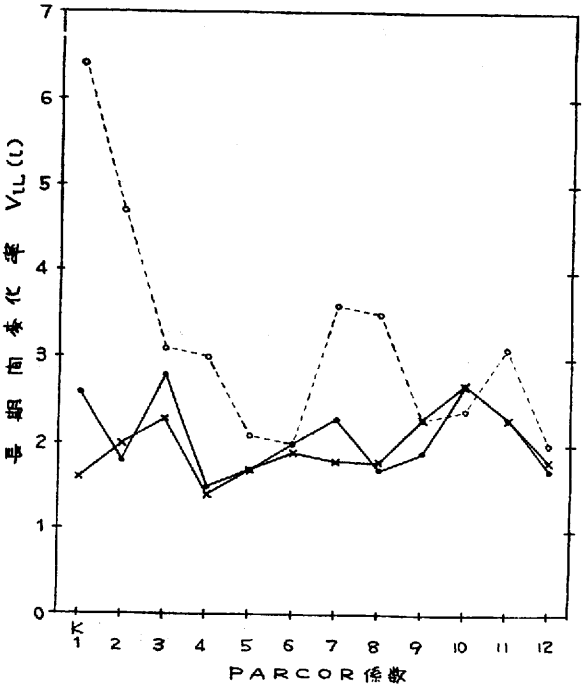


図5．1～12次の各PARCOR係数の長期間変化率（2単語平均結果）

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

5. 分散分析

(1) パラメータ別分散分析

2単語に関して，分析に用いるサンプルの期間を10日から24か月まで変えて，各パラメータ別に話者要因に関する分散分析を行なった．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の大きさ（F/F_0）を，図6～7に示す．図6は，声帯に関するパラメータβ，γ，f_0の結果，図7は，方法IIで2次C.D.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のうち，1～2次のPARCOR係数K_1，K_2の結果を示したものである．

声帯波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，3か月以上の長期間では，その効果の大きさは1/7程度に低下してしまう．一方K_1およびK_2では，このような性質はあまりなく，長期間については，β，γよりもK_2の方が単独のパラメータとして大きな話者効果を示している．

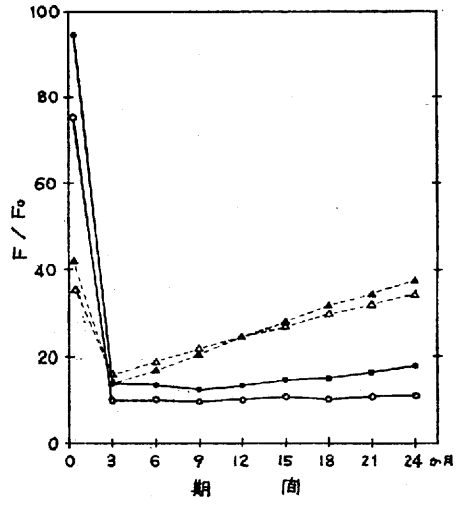


図6. サンプルの期間と話者効果の関係

--▲-- 基本周波数（/namae/）
--△-- 基本周波数（/bango:/）
—●— 逆フィルタの係数γ（方法II）
—○— 逆フィルタの係数β（方法II）

(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

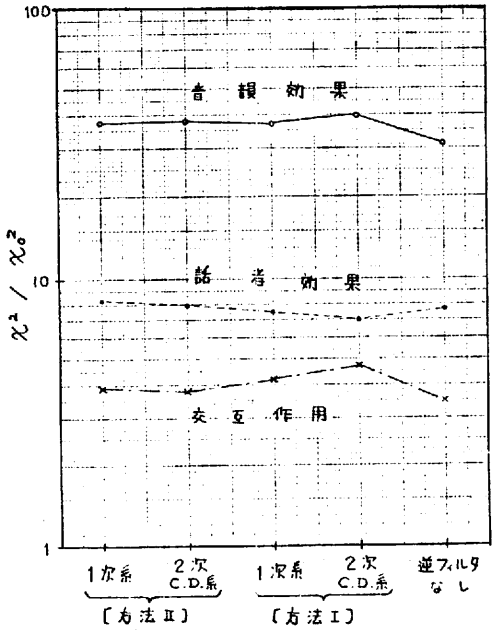


図8. 1～10次のPARCOR係数からなるベクトルの多変量分散分析結果

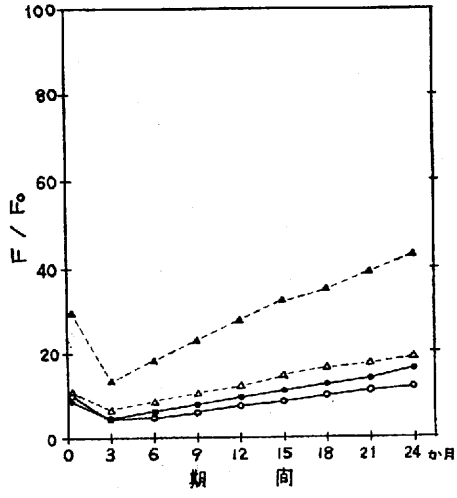


図7. サンプルの期間と話者効果の関係

--▲-- K_2（/namae/）
--△-- K_2（/bango:/）
—●— K_1（/namae/）
—○— K_1（/bango:/）

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，音韻効果，話者効果，および音韻－話者交互作用に関する多変量分散分析を行なった．

危険率1％のときの有意水準値 χ_0^2 を1としたときの各効果の大きさ（χ^2/χ_0^2）を図8に示す．この結果から，音韻効果，話者効果ともに，声帯波特性を除去してもほとんど変わらないことがわかる．

6．認識実験

上記の結果から，既知サンプルが短期間のものである場合には，声帯波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，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，分散，相関行列）を用いた認識方法[1]によって，声帯波特性を除去する効果を調べた．声帯波特性を除去したときに認識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，/bango:/，/namae/ の2種類を用いた．

(1) 短期間の既知サンプルによる認識

下図のように，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．

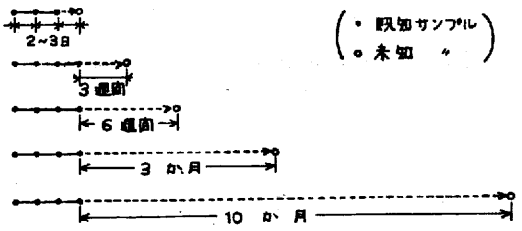


図9は，方法Ⅰによる場合，図10は，方法Ⅱによる場合の話者照合の結果である．これらの結果から，時期的にゆっくり大きく変化する傾向の強い声帯波特性を除去することにより，認識誤りを1/2程度にへらすことができること，すなわちパラメータの時期的変動の影響をかなりおさえることができることがわかる．

方法Ⅰと方法Ⅱとでは，後者の方が比較的よい結果を示している．

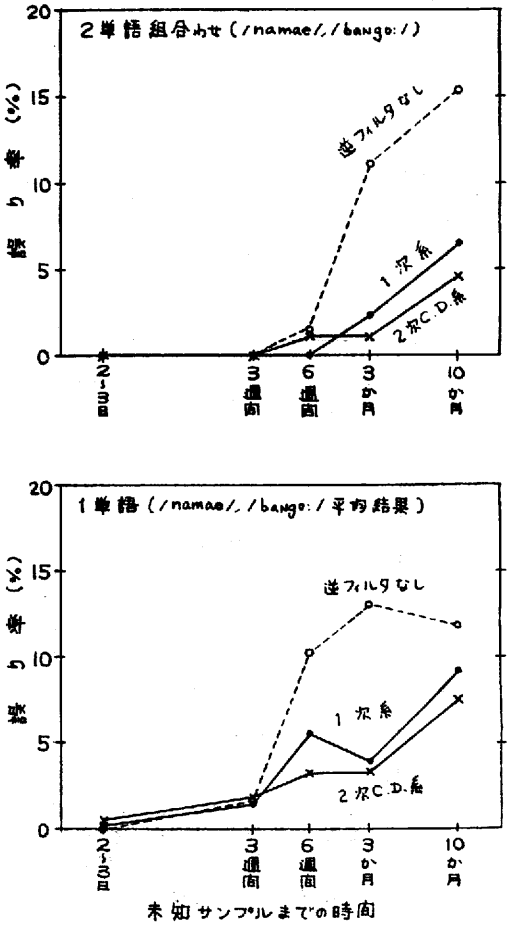
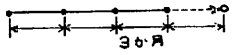


図9．短期間の既知サンプルによる話者照合（方法Ⅰ）

(2) 長期間の既知サンプルによる認識

下図のように，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なった．



その結果は図11に示す通りで，この場合は種々の方法による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見なせるためと，長期間では声帯波スペクトルの有する話者効果がかなり小さいためと考えられ，この結果は，図8の分散分析結果と対応している．

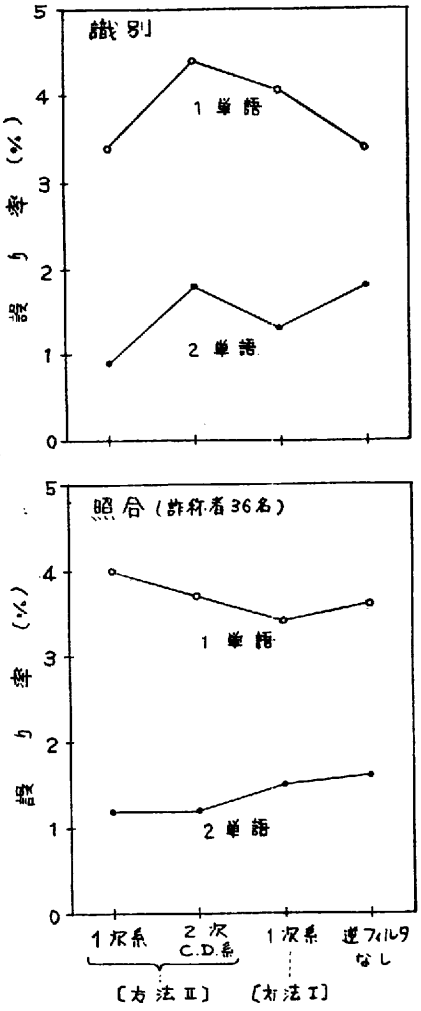


図11. 長期間の既知サンプルによる話者認識

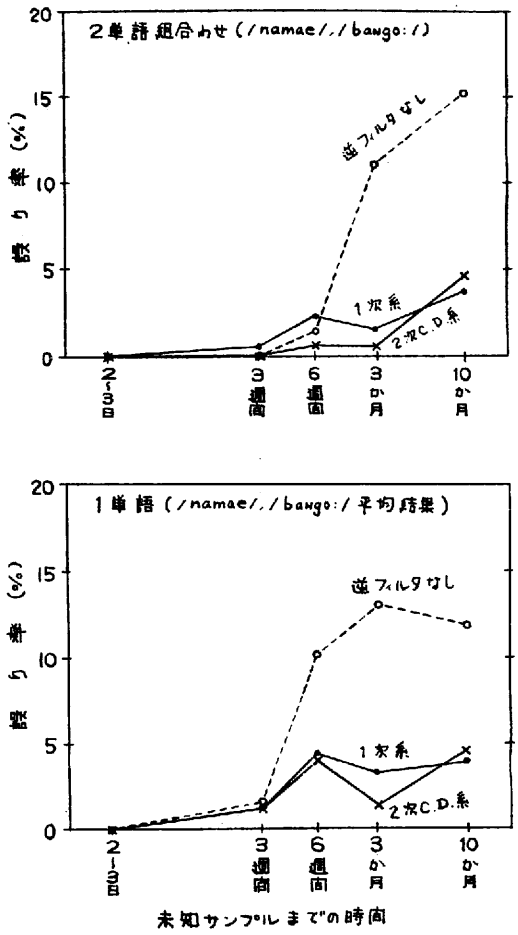


図10. 短期間の既知サンプルによる話者照合（方法Ⅱ）

(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

下に示すように，共通の6時期のサンプル（A～F）を用いて，既知サンプルが10日間程度の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)の比較を行なった．(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．

その結果，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，声帯波特性を除去して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と同程度の99%の認識率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．

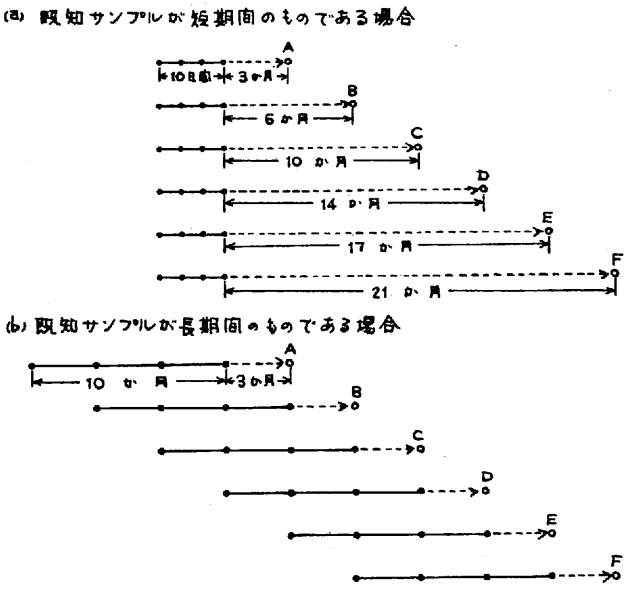


表2．既知サンプルが短期間の場合と長期間の場合の比較

既知サンプル	未知サンプルまでの時間	逆フィルタ	識別 1単語	識別 2単語	照合（許容者8名） 1単語	照合（許容者8名） 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C.D.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C.D.系	96.0	98.8	97.1	98.8
		なし	98.2	99.4	97.9	99.5

(4)逆フィルタの係数β，γを特徴ベクトルに加えた場合

声帯波特性の概形を近似するパラメータβおよびγを特徴ベクトルに加えて認識実験を行なった．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β，γの有する話者効果がかなり小さいためと思われるが，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，音声波から声帯波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，声道特性よりも声帯特性の方が，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なった結果，声帯波特性が有する個人性は，短期間では大きいが，長期間になるとかなり小さくなることが明らかになった．これらの結果にもとづいて，声帯波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ない，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か月～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にも99%の認識率が得られることがわかった．

今後さらに種々の問題点について，検討を進めて行く予定である．

［謝辞］　日頃御指導御鞭撻を頂く，当研究所野田基礎研究部長，斉藤才4研究室長，ならびに熱心に討論して頂いた才4研究室の諸氏に深く感謝する．

文　　献

(1) 古井，板倉：スペクトルパタンの時間変動特性を考慮した話者認識．音響学会講演論文集，2－2－20　（昭47－10）
(2)　　同　　：スペクトルパラメータの相関を用いた話者認識．音響学会音声研究会資料　（昭47－11）
(3)　　同　　：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．通信学会全国大会，1258　（昭48－03）
(4)　　同　　：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．音響学会講演論文集，2－1－18　（昭48－05）
(5)　　同　　：音声の個人性パラメータの時期的変動特性の解析．音響学会講演論文集，2－3－1　（昭48－10）
(6)　　同　　：単語の統計的パラメータによる話者認識．通信学会論文誌，56－A，11　（昭48－11）　（予定）
(7) 中島他　　：適応逆フィルタ法による声道断面積関数の推定．音響学会音声研究会資料　（昭48－02）

盛光印刷所