論文 / 著書情報
Article / Book Information

論題(和文)	音源・声道特徴を用いたテキスト独立形話者認識
Title(English)	
著者(和文)	松井知子, 古井貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	電子情報通信学会論文誌, Vol. J75-A, No. 4, pp. 703-709
Citation(English)	, Vol. J75-A, No. 4, pp. 703-709
発行日 / Pub. date	1992, 4
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) 1992 Institute of Electronics, Information and Communication Engineers.

論　文

音源・声道特徴を用いたテキスト独立形話者認識

正　員　松井　知子[†]　　正　員　古井　貞熙[†]

Text-Independent Speaker Recognition Using Vocal Tract and Pitch Information

Tomoko MATSUI[†] *and* Sadaoki FURUI[†], *Members*

あらまし　ベクトル量子化による話者認識において，声道，音源に関する特徴量を，テキスト独立に，かつ時期的な変動に対して頑健に組み合わせる方法について検討した．本方式には次の三つの特徴がある．第1は，声道，音源に関する特徴量を組み合わせて用いていること，第2は，話者間・話者内のばらつきを考慮した特徴量正規化法の提案，第3は，類似度を計算するための新しい距離尺度，「ひずみ・重なり尺度」の提案である．本方法によるテキスト独立形話者認識実験を行った結果，男性9名の約3年間にわたって収集されたデータに対して，1時期のデータのみで学習を行う条件において，99.0%の話者識別率，98.7%の話者照合率を得た．

キーワード：話者認識，テキスト独立，ベクトル量子化，時期差，距離尺度

1. まえがき

話者認識の研究において，音声波に含まれる個人性情報をいかにうまく抽出するかが最重要課題であることはいうまでもない．個人性情報は音韻性情報と深くかかわり合っていて，両者を分離することは難しい．また，個人性情報は同じ言葉でも発声の度に変動する(時期差による変動)．これらの問題が個人性情報の抽出を困難なものとしている．このため，現状ではテキスト独立な話者認識法，時期的な変動に対して頑健な話者認識法はまだ確立されていない．

人間は，声の音色，高さ，大きさ，発声速度，イントネーション，アクセント，語彙等のさまざまな情報から個人性を知覚しており，発声内容，長さ，時期差や体調による変動等の状況の変化にも対応できる．このような観点から，個人性情報は複数の特徴量の関係として表現する必要があると考える．

従来の話者認識に関する研究[1]~[10]では，声道に関する特徴量が用いられることが多く，それと音源に関する特徴量を組み合わせた例は少ない[3]~[5]．この要因としては音源に関する特徴量の抽出の難しさ，その適切な利用法の難しさ，大量データの処理能力の不足等が考えられる．

本研究では，音源と声道に関する特徴量を同時に扱って，発声内容に依存しない，かつ時期的な変動に対して頑健な話者認識法について検討する．

† NTTヒューマンインタフェース研究所，武蔵野市
NTT Human Interface Laboratories, Musashino-shi, 180 Japan

2. 時期的な変動を考慮した話者認識法

2.1 話者認識システムの構成

ここで検討する話者認識法は，ベクトル量子化に基づく方法である．一般にベクトル量子化による話者認識では，入力音声から抽出した特徴量(テストデータ)と，登録している各話者の特徴量の標準パターン（符号帳）との類似度（平均量子化ひずみ）を計算し，最も類似度が大きい（ひずみが小さい）話者を選び出して，その人の音声と判定する[6],[7]．

ここでは，声道に関する特徴量としてケプストラムとΔケプストラム，音源に関する特徴量としてピッチ

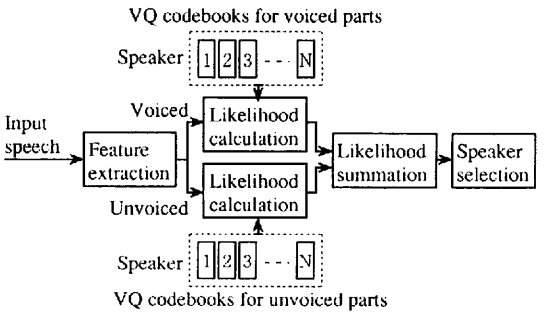

図1　話者認識システムの構成
Fig. 1　Speaker recognition system block diagram.

とΔピッチを扱う．ピッチ特徴量は，無声音区間などでは抽出できないので，そのような区間の処理が問題となる．ここでは，ピッチが抽出できる区間とできない区間とで，別々に符号帳を作成し，それぞれの区間における類似度を区間長の比で重み付けして足し合わせることにより，一つの類似度に統合した．話者認識システムの構成を図1に示す．

2.2 特徴量正規化法 TVN

特徴量は，一つのベクトル $\vec{x}=(x_1, x_2, \cdots, x_i \cdots)$に統合して表す．複数の特徴量を扱う場合，特徴量間で分布が異なるので，平均量子化ひずみを計算するときに，各特徴量を正規化する必要がある．話者認識においては，話者間のばらつきが大きく，話者内のばらつきが小さく，かつ時期的な変動の少ない特徴量が有効である．各特徴量を正規化するときには，そのような有効性を考慮する必要がある．従来の話者認識では，簡易マハラノビス法が用いられることが多い．この方法は，各特徴量を話者内標準偏差で割るという方法で，各特徴量の話者内の分布のばらつきを正規化している．

ここでは，各特徴量の話者間の分布のばらつきも考慮した正規化法として，TVN (Talker Variability Normalization)を提案する．TVNでは，正規化特徴量 $\vec{y}$ を，次に示す x_i に対する重み w_i を用いて，$y_i=w_i \cdot x_i$ で与える．

$$w_i=\sum_{n=1}^{N}\sum_{\substack{m=1\\ m\neq n}}^{N}\frac{\sigma_{ni}}{L_{mni}^2} \qquad (1)$$

ここで，

N ：話者数

m, n：話者の識別子

σ_{ni} ：話者 n の x_i の標準偏差

L_{mni}：話者 m と話者 n の x_i に関する分布の重なりを示す尺度

である．

σ_{ni} は，各話者の学習データから計算する．L_{mni} は，話者 m と話者 n の特徴量 x_i に関する分布の重なり方に応じて，次のように与える．

$$L_{mni}=\begin{cases}3(\sigma_{mi}+\sigma_{ni})-|\mu_{mi}-\mu_{ni}| & \text{if} \quad (2)\\ 6\cdot\min(\sigma_{mi}, \sigma_{ni}) & \text{if} \quad (3)\\ \varepsilon_i & \text{if} \quad (4)\end{cases}$$

ここで各条件は，

$$3|\sigma_{mi}+\sigma_{ni}|-\varepsilon_i \geq |\mu_{mi}-\mu_{ni}| > 3|\sigma_{mi}-\sigma_{ni}| \qquad (2)$$

$$3|\sigma_{mi}-\sigma_{ni}| \geq |\mu_{mi}-\mu_{ni}| \qquad (3)$$

$$|\mu_{mi}-\mu_{ni}| > 3|\sigma_{mi}+\sigma_{ni}|-\varepsilon_i \qquad (4)$$

μ_{ni}：話者 n の x_i の平均値

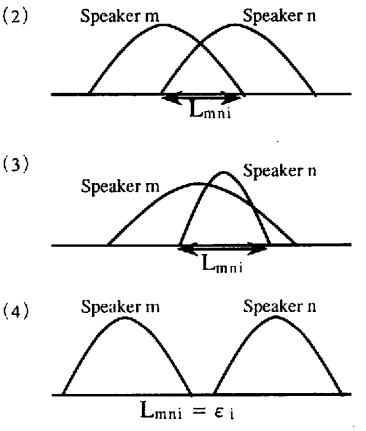

図2 L_{mni} に関する条件(2)，(3)，(4)
Fig. 2 Illustration of L_{mni} under three different conditions: (2), (3), and (4).

ε_i ：定数 $(\varepsilon_i>0)$

である．

条件(2)，(3)，(4)は図2のように，話者 m と話者 n の特徴量 x_i の分布を正規分布で近似したときに，重なりをもつ状態，包含関係にある状態，重なりをもたない状態に対応している．また，ε_i は特徴量 x_i の分布に応じて，$\varepsilon_i>0$ の十分小さな値に設定する．μ_{ni} も σ_{ni} と同様に，各話者の学習データから計算する．

正規分布で近似した特徴量 x_i の分布を更に三角形で近似した場合，話者 n の分布は底辺が $6\cdot\sigma_{ni}$ の三角形になり，話者 m，話者 n の分布の重なり部は底辺が L_{mni} の三角形になる．各話者の分布の形が極端に異なることは少ないので，重なり部の三角形と各話者の分布の三角形はほぼ相似の関係にあり，その面積は L_{mni}^2/σ_{ni}^2 にほぼ比例する．分布の重なりが小さい(話者間のばらつきが大きく，かつ話者内のばらつきが小さい)場合，特徴量 x_i の話者認識における有効性が高いと考えられる．よって，L_{mni}^2/σ_{ni}^2 の逆数は話者認識における有効性を示す．重み w_i は，話者認識における有効性に対応した L_{mni}^2/σ_{ni}^2 の逆数と簡易マハラノビス法で用いられる $1/\sigma_{ni}$ との積として，次のように定義される．

$$w_i=\sum_{n=1}^{N}\sum_{\substack{m=1\\ m\neq n}}^{N}\frac{1}{\dfrac{L_{mni}^2}{\sigma_{ni}^2}}\frac{1}{\sigma_{ni}}=\sum_{n=1}^{N}\sum_{\substack{m=1\\ m\neq n}}^{N}\frac{\sigma_{ni}}{L_{mni}^2} \qquad (5)$$

以上より，重み w_i は話者間のばらつきを強調し，かつ話者内のばらつきを吸収する効果をもつ．これまで，話者間/話者内分散比 (F 比) によって，特徴量の有効性を評価したり，特徴量を選択したりした例はあるが，

この考えを特徴量の重み付けに適用して成功した例は少ない[5].

2.3 「ひずみ・重なり尺度」DIM

時期差を含んだ，テキスト独立な話者認識では，テストデータの分布の中で，符号帳の分布から外れた部分が生ずることがある．この部分は，時期的な変動を受け易い部分，若しくは符号帳の作成には出現しなかった音韻が含まれている部分である可能性が大きい．

ここでは，類似度を計算する方法として，テストデータの分布と符号帳の分布の重なり部に含まれるテストデータのサンプル数，およびその重なり部での平均量子化ひずみの両者を考慮する「ひずみ・重なり尺度」DIM (Distortion-Intersection Measure) を提案する．

DIM では，正規化したテストデータ系列 $\{\vec{y}_j\}$ (サンプル数：J) と話者 n の符号帳 $\{\vec{c}_{nk}\}$ (符号帳サイズ：K) のひずみ量 $\mathcal{D}(\{\vec{y}_j\},\ \{\vec{c}_{nk}\})$ を次式で表す．

$$\mathcal{D}(\{\vec{y}_j\},\ \{\vec{c}_{nk}\})=\frac{\sum_{j=1}^{J} d_{nj}+R_n^2(U-u_n)}{U} \tag{6}$$

ここで，

$$d_{nj}=\begin{cases}\min\limits_k \|\vec{y}_j-\vec{c}_{nk}\|^2 \\ \quad \text{if} \quad \min\limits_k \|\vec{y}_j-\vec{c}_{nk}\|^2 \leqq r_{nk'}^2 \end{cases} \tag{7}$$

$$k'=\arg\min_k \|\vec{y}_j-\vec{c}_{nk}\|^2 \tag{8}$$

$$0 \quad \text{otherwise}$$

$$R_n=\max_k r_{nk} \tag{9}$$

$$U=\max_n u_n \tag{10}$$

u_n：(7)に該当するサンプル数

r_{nk}：k 番目のコードワードの覆う範囲

$\|\cdot\|$：ユークリッド距離

である．

$\sum_{j=1}^{J} d_{nj}$ は，テストデータの分布と符号帳の分布の重なり部の量子化ひずみの和を表す．重なり部は，符号帳に含まれる各コードワードに対して，それぞれの覆う範囲 r_{nk} を定めることによって表す．ここでは，コードワード $\vec{c}_{nk}$ をセントロイドとするクラスタに属するサンプルの各々について，セントロイドからのユークリッド距離を計算し，r_{nk} はその標準偏差の3倍に設定した．図3(7)の場合のようにサンプル $\vec{y}_j$ がどれかのコードワードの覆う範囲に入れば，(7)のようにひずみを計算する．図3(8)の場合のようにどのコードワードの覆う範囲にも入らないならば，(8)のようにひずみは計算しない．

$R_n^2(U-u_n)$ は，重ならない部分の大きさに対するペナルティを表す．重なり部の大きさは，(7)に該当するサンプル数で示される．U は全話者の符号帳における(7)に該当するサンプル数の最大値である．$R_n^2(U-u_n)$ は，重なり部の大きさが最大の符号帳と比べ，その大きさの差に比例した値である．重なり部のサンプル数が多い場合，「ひずみ・重なり尺度」DIM は小さくなる．R_n は類似度を計算するときに，重なり部の平均量子化ひずみと大きさのどちらを重視するかを表す重みと考えることもできる．

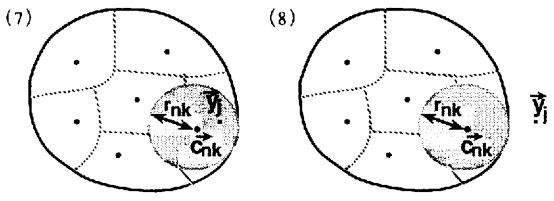

図3 テストデータの分布と符号帳の分布の重なり部の量子化ひずみを計算するための条件(7)，(8)

Fig. 3 Illustration of two conditions, (7) and (8), for calculating the quantization distortion of the intersection space between a set of test vectors and a set of VQ codebook vectors.

3. 認識実験

3.1 概　　要

男性9名が約3年間にわたる四つの時期に，繰り返し発声した5母音，5単語，3文章（表1）を対象とする．

ケプストラムは標本化周波数 12 kHz，フレーム長 32 ms，フレーム周期 8 ms，LPC 分析次数 15 で抽出し

表1 音声データ

5母音	あ，い，う，え，お
5単語	高原，爆音，海，番号，名前
3文章	爆音が銀世界の高原に広がる． (A. 約5秒)
	昔々ある所におじいさんとおばあさんがいました．おじいさんは山へ芝刈りにおばあさんは川へ洗濯に出掛けました． (B. 約12秒)
	朝鮮南部に低気圧があって，急速に発達しながら，東ないし東北東に進んでいます．関東地方ではまだ晴れている所もありますが，西日本では全般的に曇りとなりました．気温は平年より5度から6度高く，明日も引き続き高めでしょう．なお，明日は全般に南よりの風が強まり，海上は風波がかなり高くなりますから，船舶は十分ご注意ください． (C. 約30秒)

た．Δケプストラムはケプストラムの短区間（88 ms）の回帰係数である．ピッチは，上と同じLPC分析条件による変形相関法で抽出した．Δピッチはピッチの短区間（152 ms）の回帰係数である．それらの特徴量を2.2で述べた特徴量正規化法TVNを用いて正規化した後，一つの特徴ベクトルに統合して表す．

符号帳は，5母音，5単語，2文章(A, B)を用い，LBGアルゴリズム[11]で各話者，各時期ごとに作成する．テストには，符号帳の作成に使った文章とは別の1文章(C)を用いる．各話者，各時期ごとにその1文章を，三つの異なる時期の符号帳に対して，2.3で述べた「ひずみ・重なり尺度」DIMを用いてテストする．すなわち，各話者ごとに，テストデータとは異なる1時期の音声から作成した符号帳を用いてテストを行う．結果は全時期における，異なる時期の符号帳に対する平均識別誤り率，平均照合誤り率で評価する．

話者識別の場合は，9名の話者の中から最も類似度の大きい話者を選択した．話者照合の場合は，類似度の値にしきい値を設けて，しきい値よりも類似度の大きいテストデータは本人の音声として受理し，それ以外は棄却した．このしきい値は，話者ごとに2種類の誤り（詐称者受理率，本人棄却率）が等しくなる値に事後的に設定した．

参考のために，学習データ，テストデータに含まれる音韻の数をヘボン式ローマ字つづりにより比較した．結果を表2に示す．学習データに含まれる音韻は19種類，テストデータは21種類であり，テストデータ中の2種類の音韻は，学習データに全く含まれていないことがわかる．また，各音韻の出現率も学習データとテストデータとで大きく異なることがわかる．

表2 ヘボン式ローマ字つづりによる音韻の数の比較

音韻	学習	テスト	音韻	学習	テスト
a	33	57	y	1	9
i	18	37	r	3	15
u	9	27	w	3	6
e	11	17	g	7	8
o	14	28	z	2	6
k	12	24	d	1	8
s	11	17	b	6	2
t	4	21	p	0	3
n	5	17	N	11	13
h	1	10	Q	0	2
m	7	12	合計	159	339

表3 認識結果

平均識別誤り率（%）	1.0
平均照合誤り率（%）	1.3

3.2 結　果

認識結果を表3に示す．ここでは，特徴量正規化法TVNの中のε_iの値は，すべての特徴量に共通に0.01に設定した．話者照合実験において，特徴量正規化法TVNの重みw_iは，話者識別実験の場合と同様に，テスト用詐称者を含む全9名の話者の学習データを用いて求めた．また，話者照合実験において，「ひずみ・重なり尺度」DIMの中のUの値を設定するために，本人だけでなく，詐称者8名の学習データとの類似度も計算し，各話者のu_nを求めた．

従来の長時間平均スペクトルによる話者認識実験例[2]では，本実験のように発声内容に依存せず，1時期の音声によって学習を行って，時期差のある音声をテストに用いた場合，平均認識誤り率は10%程度であった．これに比べると，本実験の認識誤り率は約1/10程度に小さい．

4. 考　察

4.1 特徴量の組合せによる効果

上記の実験では，声道に関する特徴量としてケプストラム，Δケプストラムを，音源に関する特徴量としてピッチ，Δピッチを組み合わせて用いたが，組み合わせる特徴量を変えて，特徴量の組合せによる効果を調べる話者識別実験を行った．本実験では，特徴量を組み合わせる方法として簡易マハラノビス法を用いた．また，類似度を計算する方法としては「ひずみ・重なり尺度」DIMは用いず，全テストデータによる平均量子化ひずみを用いた．表4に平均識別誤り率を示す．

表4の結果から，組み合わせる特徴量の数の増加に伴い，平均識別誤り率が低下することがわかる．音源に関する特徴量は，単独で用いた場合には平均識別誤り率が非常に大きいが，声道に関する特徴量と適切に

表4 特徴量の組合せの効果

特徴量	平均識別誤り率%
ケプストラム	9.4
Δケプストラム	11.8
ピッチ	74.4
Δピッチ	85.9
ケプストラム，Δケプストラム	7.9
ケプストラム，Δケプストラム ピッチ，Δピッチ	2.9

組み合わせることにより，高い話者識別効果が得られることがわかる．ケプストラム，Δケプストラム，ピッチおよびΔピッチを組み合わせたときの誤り率は，ケプストラムのみの場合の約1/3である．

4.2 特徴量正規化法TVNによる効果

4.2.1 簡易マハラノビス法との比較

特徴量正規化法TVNの効果を調べるために，簡易マハラノビス法との比較実験を行った．表5に，話者識別実験および話者照合実験の結果を示す．いずれも類似度を計算する方法として，「ひずみ・重なり尺度」DIMは用いず，全テストデータによる平均量子化ひずみを用いた．

これらの結果から，特徴量正規化法TVNは話者識別，話者照合いずれにおいても有効であることがわかる．すべての特徴量を組み合わせ，TVNを用いたときの誤り率は，簡易マハラノビス法の場合の約2/3である．

表5 平均誤り率(%)によるTVNとマハラノビス法の比較

特徴量	実験	TVN	マハラノビス法
ケプストラム	識別	5.5	9.4
	照合	5.1	5.7
ケプストラム Δケプストラム	識別	6.0	7.9
	照合	4.8	6.8
ケプストラム Δケプストラム ピッチ Δピッチ	識別	1.9	2.9
	照合	5.6	9.9

4.2.2 特徴量の分布に関する検討

特徴量正規化法TVNでは，話者nと話者mの特徴量x_iに関する分布の重なりを示す尺度L_{mni}を求めるときに，特徴量x_iの分布を正規分布で近似した．正規分布で近似することの妥当性を評価するため，各特徴量のヒストグラムを求め，その概形を調べた．図4は，ある話者のある時期の学習データについて，有声音，無声音区間別にケプストラムの1次係数(C_1)，2次係数(C_2)，Δケプストラムの1次係数(ΔC_1)，2次係数(ΔC_2)，ピッチ(P)，Δピッチ(ΔP)のヒストグラムを表示したものである．

図4から，各特徴量のヒストグラムを正規分布で近似することはほぼ妥当であると思われる．

4.3 「ひずみ・重なり尺度」DIMによる効果

4.3.1 全テストデータによる平均量子化ひずみとの比較

類似度を計算する方法として，「ひずみ・重なり尺度」DIMを用いた場合と全テストデータによる平均量子化ひずみを用いた場合について比較した．表6に話者識別実験，話者照合実験の結果を示す．両者とも，特徴量としてケプストラム，Δケプストラム，ピッチ，Δピッチを用い，それらを正規化する方法としては特徴量正規化法TVNを用いた．

この実験から，「ひずみ・重なり尺度」DIMは，話者認識において有効な尺度であることが確かめられた．すべての特徴量を用い，DIMを用いたときの誤り率は，平均量子化ひずみを用いる場合の約1/2〜1/4であ

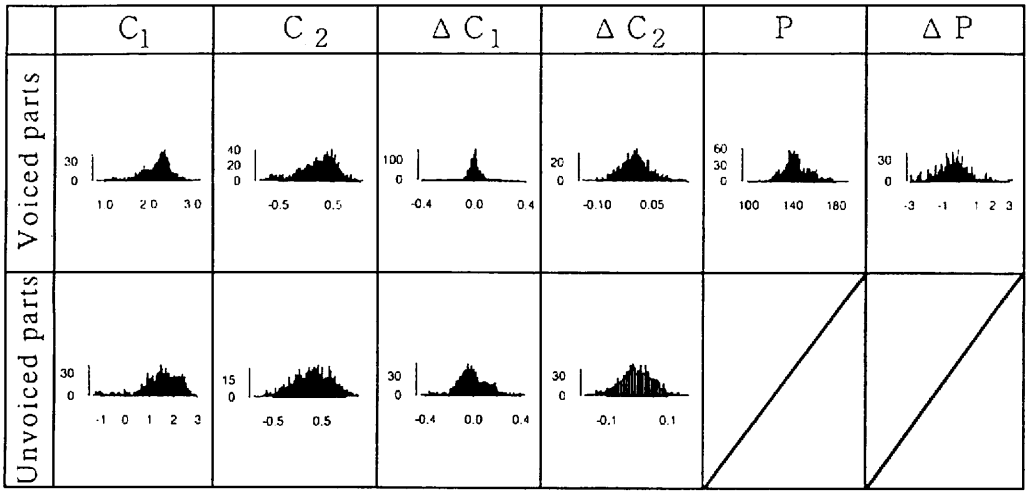

図4 特徴量の有声音・無声音区間別ヒストグラム
Fig. 4 Histogram of each feature parameter for voiced/unvoiced parts.

る．

4.3.2 R_n に関する考案

上記の実験では，2.3 で述べたように，R_n を，各話者の符号帳の全コードワードについて，それをセントロイドとするクラスタを超球で近似したときの半径 r_{nk} を調べ，その最大値に設定していた．この設定の妥当性を確かめるために，R_n^2 の値を 0 から $2 \times R_n^2$ まで動かし，平均識別誤り率の変化を調べた．実験では，特徴量としてケプストラム，Δケプストラム，ピッチ，Δピッチを用い，その正規化法としては特徴量正規化法 TVN を用いた．図 5 に結果を示す．

図 5 より，R_n を 2.3 で述べた方法によって設定する（R_n^2 の係数＝1）のが妥当であることが確認できた．また，R_n^2 の値を小さくするより，大きくする方が平均識別誤り率の増加が極端に少ないことから，テストデータの分布と符号帳の分布の重なり部の大きさが話者認識においては重要であることがわかる．

表 6 平均誤り率(%)による DIM の評価結果

特徴量	実験	DIM	平均量子化歪
ケプストラム	識別	5.2	5.5
	照合	2.2	5.1
ケプストラム Δケプストラム	識別	4.1	6.0
	照合	2.5	4.8
ケプストラム Δケプストラム ピッチ Δピッチ	識別	1.0	1.9
	照合	1.3	5.6

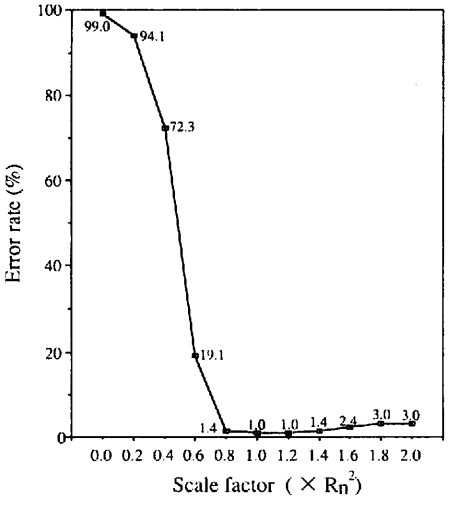

図 5 DIM における R_n^2 の値と平均識別誤り率の関係
Fig. 5 Error rate for different scale factors for R_n^2 in DIM.

4.4 テストデータ長と平均誤り率との関係

上記の実験では，テストデータは発声時間が約 30 秒の長文であった．テストデータ長と平均識別誤り率との関係を調べるために，テストデータを 2～8 等分し，その長さを変えて表 7 に示した 9 通りの話者識別実験を行った．結果を図 6 に示す．

図 6 より，広い範囲のテストデータ長に関して，類似度を計算する方法として「ひずみ・重なり尺度」DIM を用いた実験 C, N, Z は，全テストデータによる平均量子化ひずみを用いた実験 B, M, Y やその他の実験と比較しても，平均話者識別誤り率が低いことがわかる．また，特徴量を正規化する方法として，特徴量正規化法 TVN を用いた実験 B, M, Y とマハラノビス法を用いた実験 A, L, X の結果を比べると，実験 B, M, Y の方がテストデータの長さによる影響を受けにくいことがわかる．

5. むすび

テキスト独立な話者認識において，声道に関する特徴量と音源に関する特徴量の組合せ，変動を考慮した正規化法 TVN，および「ひずみ・重なり尺度」DIM が有効であり，時期的な変動がある場合でも，高い話者

表 7 種々の特徴量，正規化法，類似度の組合せによる話者識別実験の方法

	特徴量	正規化	類似度
A	ケプストラム	マハラノビス法	平均量子化歪
B	ケプストラム	TVN	平均量子化歪
C	ケプストラム	TVN	DIM
L	ケプストラム，Δケプストラム	マハラノビス法	平均量子化歪
M	ケプストラム，Δケプストラム	TVN	平均量子化歪
N	ケプストラム，Δケプストラム	TVN	DIM
X	ケプストラム，Δケプストラム，ピッチ，Δピッチ	マハラノビス法	平均量子化歪
Y	ケプストラム，Δケプストラム，ピッチ，Δピッチ	TVN	平均量子化歪
Z	ケプストラム，Δケプストラム，ピッチ，Δピッチ	TVN	DIM

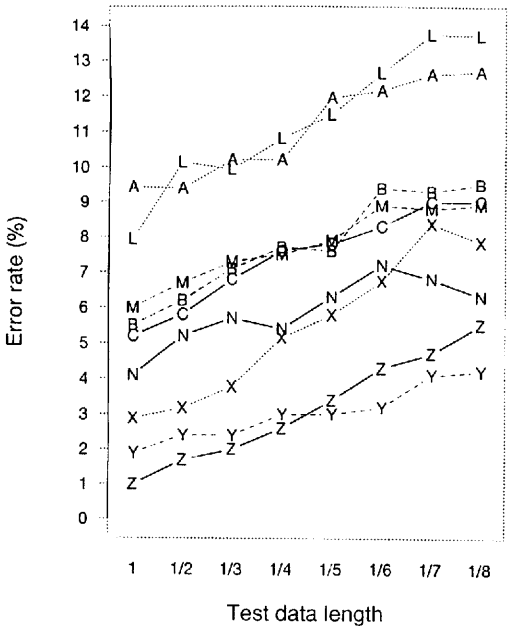

図 6　テストデータ長と平均識別誤り率との関係
Fig. 6　Error rate for different test data lengths.

認識精度が得られることがわかった．声道に関する特徴量と音源に関する特徴量を組み合わせて用いた場合，声道に関する特徴量だけを用いた場合に比べて，平均話者識別誤り率が約 1/3 になった．また，特徴量を正規化する方法として TVN を用いた場合，マハラノビス法を用いた場合と比較して，平均話者認識誤り率が約 2/3 になった．更に，類似度を計算する方法として「ひずみ・重なり尺度」DIM を用いた場合，全テストデータによる平均量子化ひずみを用いた場合に比べ，平均話者認識誤り率が約 1/2 になった．この結果，男性 9 名の約 3 年間にわたって収集されたデータに対して，99.0%の話者識別率，98.7%の話者照合率が得られた．

今後は話者数を増やして実験し，本話者認識法について更に改良を加えていきたい．

謝辞　熱心に討論して頂いた，基礎研究所誉田研究グループ，ヒューマンインタフェース研究所音声情報研究部の方々に感謝致します．

文　献

(1) 古井貞熙：“音声波に含まれる個人性情報の研究”，東京大学学位論文 (1978).
(2) 古井貞熙，板倉文忠，斉藤収三，“長時間平均スペクトルによる話者認識”，信学論 (A)，**55-A**，10，pp. 549-556 (1972).
(3) Furui S.: “Comparison of speaker recognition methods using statistical features and dynamic features”, IEEE Trans. Acoust., Speech & Signal Process., **ASSP-29**,3, pp. 342-350 (1981).
(4) Markel J. D. and Davis S. B.: “Text-independent speaker recognition from a large linguistically unconstrained time-spaced data base”, IEEE Trans. Acoust., Speech & Signal Process., **ASSP-27**,1, pp. 74-82 (1979).
(5) Matsumoto H.: “Text-independent speaker identification from short utterances based on piecewise discriminant analysis”, Computer Speech and Language, **3**, pp. 135-150 (1989).
(6) Soong F. K., Rosenberg A. E., Rabiner L. R. and Juang B. H.: “A vector quantization approach to speaker recognition”, Proc. ICASSP, pp. 387-390 (1985).
(7) Soong F. K. and Rosenberg A. E.: “On the use of instantaneous and transitional spectral information in speaker recognition”, Proc. ICASSP, pp. 877-880 (1986).
(8) Atal B. S.: “Automatic recognition of speakers from their voices”, Proc. IEEE, pp. 460-475 (1976).
(9) Rosenberg A. E.: “Automatic speaker verification: a review”, Proc. IEEE. **64**. pp. 475-487 (1976).
(10) 古井貞熙：“話者識別技術”，計測と制御，**25**，8，pp. 688-693 (1986).
(11) Linde Y., Buzo A. and Gray R. M.: “An algorithm for vector quantizer design”, IEEE Trans. Commun., **COM-28**. 1. pp. 84-95 (1980).

（平成 3 年 3 月 22 日受付，11 月 14 日再受付）

松井　知子

昭 61 東工大・理・情報卒．昭 63 同大大学院修士課程了．同年 NTT(株)入社．以後，NTT ヒューマンインタフェース研究所において，話者認識の基礎研究に従事．IEEE，日本音響学会各会員．

古井　貞熙

昭 43 東大・工・計数卒．昭 45 同大大学院修士課程了．同年 NTT 電気通信研究所入社．以後，同研究所において，音声認識，話者認識，音声知覚などの基礎研究に従事．昭 53～54 ベル研究所客員研究員．現在 NTT ヒューマンインタフェース研究所音声情報研究部長．工博．昭 50 米沢賞，昭 63 論文賞，平 2 著述賞，昭 60，62 音響学会論文賞，平 1 科学技術庁長官賞，IEEE ASSP Society Senior Award 各受賞．著書「ディジタル音声処理」「Digital Speech Processing, Synthesis, and Recognition」「音声情報工学」(分担)，「音の科学」(分担) など．IEEE，日本音響学会各会員．