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)	話者認識の現状と展望
Title(English)	
著者(和文)	古井貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	電子通信学会誌, Vol. 67, No. 5, pp. 537-543
Citation(English)	, Vol. 67, No. 5, pp. 537-543
発行日 / Pub. date	1984, 5
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) 1984 Institute of Electronics, Information and Communication Engineers.

Powered by T2R2 (Science Tokyo Research Repository)

UDC 801.4-051.001.36:[534.784.087.9:681.322.056]

解説B

話者認識の現状と展望

古井 貞熙

古井貞熙：正員　日本電信電話公社武蔵野電気通信研究所情報通信基礎研究部
Review and Perspective of Speaker Recognition. By Sadaoki FURUI, Member (Communication Principles Research Division, Musashino Electrical Communication Laboratory, N.T.T., Musashino-shi).

1. 話者認識と音声認識

音声波に含まれる第一義的な情報は，言葉の意味内容に関連した音韻性情報であるが，第二義的情報として重要なものに，だれが話しているかという個人性情報がある．このうち音韻性情報を自動的に抽出して認識するのがいわゆる音声認識（speech recognition）であり，個人性情報を抽出して認識するのが話者認識（speaker recognition）である[1]．広義には，この両者を合せて音声認識ということもある．

話者認識の研究には，人間が耳でだれの声かを判定するときに，音声のどのような手掛りを用いているかという研究[2]や，スペクトルの時間的変化を濃淡図形に表したスペクトログラム——いわゆる声紋（voice print）——を人間が目で見て判定する手掛りに関する研究[3]もあるが，近年，電子計算機等の発達により，自動的な認識方法の研究が急速に発展してきている．

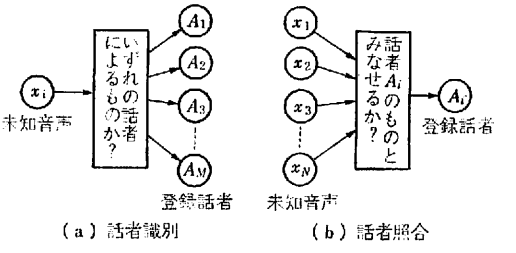
話者の自動認識の研究は，（狭義の）音声認識の研究に比べて歴史が浅く，一般の関心もまだあまり高くないが，犯罪捜査のみならず，情報化社会の発達に伴って，個人のプライバシーを守り，危険防止にも役立つ技術として，需要が増してくるものと考えられる．ここでは，以後話者認識という言葉を自動認識に限定して用いることにする．

話者認識と音声認識の研究は本来表裏一体をなすべきものであるが，音韻性情報と個人性情報は，複雑にからみあった形で音声波（スペクトル）に含まれており[4],[5]，この両者を明確に分離して抽出する技術は，まだ確立されていない．このため，音声認識では個人を固定し，話者認識では音韻性を固定することによって，高い認識率を得ているのが現状である．不特定話者を対象とした音声認識や，発声内容を限定しない話者認識において，高い認識率を得ることは非常に難しい．今までのところ，話者認識と音声認識の研究は，それぞれ独自の立場から行われてきているが，後述するように，音声認識における個人差の正規化という観点から，両者の境界に立った研究も幾つか行われている．

2. 話者認識の分類

2.1 話者識別と話者照合

話者認識は，話者識別（speaker identification）と話者照合（speaker verification）に分けることができる．図1に示すように，話者識別とは，未知音声を，あらかじめ登録してある各話者の音声（標準パターン）と比較して，最も類似している標準パターンの話者が発声した音声と判定するものである．複数カテゴリーの内の一つと判定するという点では，単語音声認識等



（a）話者識別　　（b）話者照合

図 1　話者識別と話者照合

と類似している．一方話者照合とは，未知音声が，その発声者が名乗った本人の音声であるか否かを，登録されている標準パターンとの類似の度合いによって判定するものである．

話者照合の例としては，電話による買物やバンキングサービス（送金，振替，残高照合，振込通知等），個人情報へのアクセス，機密保管場所等への入室者のチェック等において，本人か否かのかぎとして音声を用いることが考えられる．話者識別の例としては，犯罪捜査において，犯行時に録音された音声が，複数容疑者の内のだれの声かを判定するような場合がある．但しこの場合は，容疑者の中に真犯人が含まれていないこともありうるので，実際には話者照合と話者識別の組合せの形をとることになる．

2.2 発声内容依存形と不依存形

話者認識の方法には，認識に用いる音声の発声内容，すなわち言葉をあらかじめ決めておく方法（発声内容依存形，text-dependent）と，どんな言葉を発声してもよいとする方法（発声内容不依存形，text-independent）がある．

発声内容依存形の方法の中には，鼻音などの特定の音素のスペクトルを用いる方法のように，ある特殊な決まった言葉を発声しなければならないものもあるが，現在研究されている方法は，各話者がそれぞれに決めたキーワード（暗証番号，名前等）を一貫して発声すればよい，というものが多い．この場合は，実用的にはそのキーワードの違いも手掛りとして用いることができるので，それだけ高い認識率を得るのが容易となる．研究段階の認識実験では，すべての話者に共通のキーワードを用いて評価を行う場合が多い．

話者照合の応用の多くは，このようにキーワードをそれぞれの人について固定することが可能であるが，犯罪捜査の場合等では，必ずしも同じ言葉どうしを比較することができないケースもある．このような目的のためには，発声内容不依存形の方法を開発する必要があるが，前述のように，今のところ，この方法で高い認識率を得るのはかなり難しい．

2.3 話者認識に用いる音声の特徴パラメータ

音声波に含まれる個人性情報は，声道長，声帯等の先天的な発声器官の個人差に起因するものと，なまり，アクセント等の後天的な発話習慣（話し方のくせ）に起因するものがある．前者は，ホルマント周波数の高低，帯域幅の大小，平均基本周波数，スペクトル概形の傾斜等として現れ，後者は基本周波数やホルマント周波数の時間パターン，単語の時間長等として現れるが，両者を明確に分離して抽出するのは難しい．このため，両者の特徴が同時に含まれる特徴が用いられることが多い．

実際の話者認識の方法としては，スペクトル包絡，基本周波数，パワー等の時間パターンをそのまま用いる場合と，その時間パターンから抽出した統計的特徴を用いる場合がある．発声内容依存形の場合は両者の方法が用いられ，発声内容不依存形の場合は後者の方法，すなわち時間的平均をとることによって音韻的な内容の差による違いを平均し，先天的な個人差をマクロな特徴として抽出しようとする方法を用いることが多い．スペクトル包絡の時間パターンを直接的に用いる方法は，単語音声認識でしばしば用いられる方法と類似しているが，安定した個人の特徴を強調するために，話者認識に特有の配慮が必要である．

声帯模写では，後天的な発話習慣に関連する基本周波数やパワーの時間パターンを真似しているケースが多い．自動的な話者認識においても，このような特徴を基礎とする方法は作り声に対して弱く，スペクトル包絡の特徴を組み合せれば，作り声に対して強くすることができることが示されている[6]．

3. 話者認識系の構成

3.1 認識系の基本的構成

話者認識系の基本的構成を図2に示す．音声波から特徴抽出をしたのち，あらかじめ蓄えられている各登録話者の標準パターンとの距離あるいは類似度を調べその度合いにより認識の判定を行う．

話者識別の場合は，入力音声との距離が最も小さい標準パターンの話者によるものと判定し，話者照合の場合は，入力音声と標準パターンの距離と，あらかじめ定めたしきい値との大小関係によって判定を行う．このしきい値と2種類の誤り率，すなわち本人の音声を棄却してしまう誤り率と，他人（詐称者とよぶ）の音声を受理してしまう誤り率との間には，図3に示すような関係がある．実用的には，2種類の誤り率の重要性に応じて，しきい値を適切な値に設定するのが望

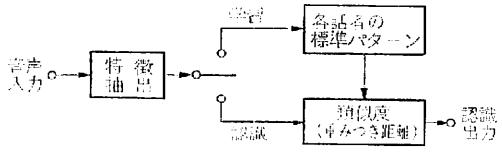
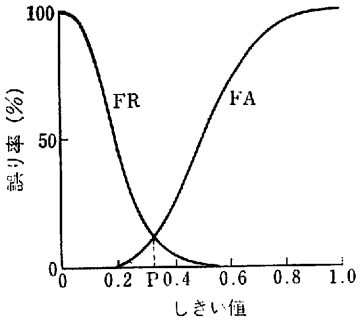


図2 話者認識系の基本的構成



FR：本人棄却率
FA：詐称者受理率

図 3 話者照合における判定のしきい値と2種類の誤り率の関係

ましいが，実験では，2種類の誤り率が等しくなる値(図中の P）にしきい値を事後的に設定し，このときの誤り率によって評価する場合が多い．

3.2 認識誤り率と話者数の関係

登録話者 N 人の集合を Z_N，音声データ（n 次元特徴ベクトル）を $X'=(x_1, x_2, \cdots, x_n)$，話者 i ($i \in Z_N$) の X の確率密度関数を $P_i(X)$ とすると，Z_N に含まれる話者の X の出現確率密度関数は，

$$P_Z(X)=\underset{i\in Z_N}{E}[P_i(X)]$$
$$=\sum_i P_i(X)P_r[i],\ i\in Z_N$$

となる．但し $P_r[i]$ は，話者 i の出現確率である．

話者照合の場合，話者 i 本人の音声として受理すべき X の領域は，

$$R_{Vi}=\{X|P_i(X)>C_iP_Z(X)\}$$

である．ここで C_i は前述の2種類の誤り率を適当なバランスに保つように選ぶ．今，Z_N が話者のランダム集合で $P_r[i]=1/N$ なら，N が大きくなるに従って $P_Z(X)$ は Z_N に無関係な密度関数に近づく．従って，2種類の誤り率は，N が大きくなってもその影響を受けない．

なお，通常 $P_Z(X)$ を正確に推定することは困難なので，一定値を仮定し，

$$R_{Vi}=\{X|P_i(X)>K_i\}$$

とすることが多い．

一方話者識別の場合は，話者 i の音声と判定する領域は，

$$R_{Ii}=\{X|P_i(X)>P_j(X),\ \forall j\neq i\}$$

である．このときの誤り率は，

$$P_{Ei}=1-\prod_{\substack{k=1\\k\neq i}}^{N} P_r(P_i(X)>P_k(X))$$

であるから，Z_N が話者のランダム集合なら，

$$\underset{Z_N}{E}[P_{Ei}]=1-\underset{Z_N}{E}\{\prod_{\substack{k=1\\k\neq i}}^{N} P_r(P_i(X)>P_k(X))\}$$
$$=1-\underset{i}{E}\{\prod_{\substack{k=1\\k\neq i}}^{N} \underset{k}{E}[P_r(P_i(X)>P_k(X))]\}$$
$$=1-\underset{i}{E}\{P_{Ai}{}^{N-1}\}$$

と書ける．ここで P_{Ai} は，i と他の任意の話者との間で誤りを生じない確率の期待値である．従って，誤り率の期待値は，N の増加と共に指数関数的に増大し，極限では1となる．

この結果は，無限個の分布は有限パラメータの空間では別々に分離していることはできないということから当然ともいえる．すなわち，話者数が大きくなれば，話者の集合中の二人，あるいはそれ以上の話者のデータの分布が互いに非常に近い可能性が大きくなるからである．このため話者識別系の評価は，登録話者数の影響を十分考慮して行う必要がある[7],[8]．

3.3 特徴パラメータの長時間変動

話者認識系を構成するときに考慮すべきこととして，特徴パラメータの長期間にわたる変動の影響がある．すなわち，同じ人が同じ言葉を繰返し発声しても，ある程度時期をおくと，発声のし方が変って音声のスペクトルが変化し，認識系の性能に影響を及ぼす．音声認識においてはこのようなことは通常問題にならないが[9]，話者認識においては，標準パターン作成のための学習サンプルと未知入力音声が発声される時期に必ず隔たりがあり，しかも個人性情報は音韻性情報に比べて細かい特徴であるため，この時期差による変動の影響を受けやすい．このように本質的に変動する性質を有する点が，音声の指紋と異なるところである．

特徴パラメータの時期的変動の影響を軽減し，高い認識率を得るためには，次のような方法が有効であることが確認されている[10]．

（1） 音声波にスペクトル等化処理を行う．すなわち，単語または短文章の時間平均スペクトルを1次系または2次臨界制動系で近似し，音声をこの逆フィルタに通す．この処理は近似的に，音声中の声帯波情報を除去し，声道情報のみを保存することに相当している．

（2） 長期間にわたる音声データに基づく統計的評価により，安定した特徴パラメータを選択する．

（3） 長期間にわたる学習サンプルに基づいて，標準パターンと距離尺度を定める．

（4） 標準パターンを適切な周期で更新する．

（5） 異なる複数の単語等から抽出した特徴パラメータを組み合せて用いる．

3.4 特徴パラメータの評価

音声の個人性情報は，比較的細かい情報として音声波のさまざまな性質に分散して含まれているため，種々の特徴パラメータから，長期的に安定しかつ有効性の高いものを見出すのは重要な課題である．この目的のためには，次のような方法を用いることができる．①各パラメータの組合せによる認識実験を行う．②パラメータごとに，F比（話者間/話者内分散比）を計算する[10]．③F比を多次元に拡張した divergence を用いる[11]．④認識誤り率に基づく knock out 法を用いる[12]．

効率的に情報圧縮しパラメータ数を減らす方法としては，判別分析によりF比を最大にする空間へ射影する方法がしばしば用いられる．

4. 話者認識系の具体例

4.1 概　　観

話者認識システムはまだ実用化されるに至っていないが，米国 Texas Instruments 社における，男女計300名の話者による長期間にわたるコンピュータ室の入室管理システムの試行実験[13]，Bell 研究所における男女計100名の電話音声による話者照合実験[14],[15]など，大規模な実験も行われるようになってきた．基礎的な研究も国内外で着実に行われている．

この10年間の主な研究例を表1に示す．認識率の欄には，報告されている値をそのまま載せたが，表中の各条件のほか，学習サンプルの収集期間，入力音声までの時期差等の条件も，各実験で異なっているので，この数値だけで単純に優劣を比較することはできない．全般的には，発声内容不依存形の方式はまだ基礎研究段階にあるが，発声内容依存形の話者照合は，実用化に近づいているといえる．

4.2 発声内容依存形の認識系

統計的特徴を用いる発声内容依存形の代表的方法に，電電公社武蔵野通研の，単語音声スペクトルの統計量を用いる方法[16]がある．この方法では，各話者はあらかじめ定められた4種類の単語を区切って発声し，各単語音声区間での平均相関係数を用いた1次系の逆フィルタによるスペクトル等化が施された後，LPC（線形予測）分析により，LPCケプストラムと基本周波数の時系列に変換される．各単語の有声音区間におけるこれらのパラメータの平均値，標準偏差，フーリエ展開係数，および相互相関係数を特徴量として，認識が行われる．この方法は，マイクロホン音声を用いた実験と，実際の電話音声を用いたオンライン実験により，高い認識率の得られることが確認されている．

スペクトルの動的パターンを直接的に用いる方法には，米国 Bell 研究所，Texas Instruments 社の方法等があり，前述のように，大規模な評価実験が行われている．

4.3 発声内容不依存形の認識系

発声内容不依存形の認識系としては，長期間平均スペクトルを用いる方法[17]が代表的であるが，この方法が平均化された1種類のスペクトルで個人を特徴づけるのに対し，最近米国 ITT から，クラスタ化された複数のスペクトルの組で特徴づける次のような方法が提案されている[18]．すなわち，各話者ごとに100秒の会話音声を学習サンプルとして，短時間ごとにLPC分析し，そのスペクトルパターンの集合をクラスタ化して，代表的な40種類のパターンを用意する．未知音声の短時間ごとのスペクトルと，各登録話者の40パターンとの最小距離の分布を調べ，それが最も小さい値に片寄っている登録話者による音声であると判定する．発声内容を限定しない10, 5, 3秒長の未知音声に対して，男性11名の話者識別でそれぞれ96, 87, 79% の認識率が得られており，学習音声と未知音声に伝送路の違いがある場合についても検討が行われている．

5. 音声認識における個人差の正規化

5.1 不特定話者音声の認識

話者認識が音声の個人差を積極的に利用しようとするのに対し，音声認識における個人差の正規化は，個人差を除去して，不特定多数のだれの声に対しても高い認識性能を実現しようとするものである．このためには，だれの声にも共通に含まれている各音韻の特徴を直接的に抽出するのが正統的なアプローチであるが，音声の動的性質や人間の聴覚の複雑なメカニズムが関連しているため，このような技術はまだ確立されていない．具体的方法として現在試みられている方法には，次のようなものがある．

5.2 音声生成メカニズムに基づく正規化

有声音区間において，発声内容にあまり依存せずに共通に音声スペクトルの個人差を生ずる要因として，

表 1　最近の話者認識の主な研究例

	研究機関	研究者（年代）	発声内容	特徴パラメータ・方法	識別(I) 照合(V)	登録話者数 ( )：照合の詐称者数	認識率 (%)
発声内容依存形	電電公社武蔵野通研	古井・板倉 (1973)	1～4 単語	対数断面積比と基本周波数の統計的特徴（平均値・標準偏差・相互相関係数）の重みつき距離	I	男 9	96.8～99.1
					V	男 9 (111)	97.3～99.3
		古井 (1982)	4 単語	LPC ケプストラムと基本周波数の統計的特徴（平均値・標準偏差・相互相関係数・ブーリエ展開係数）の重みつき距離	I	男 9	100
					V	男 9 (111)	99.9
	日立中研	市川・中島・中田 (1979)	1 単語	Cosh 尺度を用いた DP マッチングと平均基本周波数，帯域選択自己平均逆フィルタ法による伝送ひずみの除去，電話音声	I	男 5	91
	Bell Laboratories (米国)	Rosenberg (1976)	短文章 (2秒)	基本周波数とパワーの DP マッチングの後，種々の距離尺度を適用，電話音声	V	男・女 100 (同)	適応後 FR 4% FA 5%
		Furui (1981)	同上	LPC ケプストラムとその多項式展開係数の DP マッチング，スペクトル等化処理	V	男 10 (49) 女 10 (49) 男 21 (75)	99.8 99.6 99.2 } 電話音声
		Atal (1974)	短文章 (0.5 秒～)	種々の線形予測パラメータを直交線形変換し有効性を評価，LPC ケプストラムが最も有効	V	男 10 (9)	98 (1秒)
					I	男 10	98 (0.5 秒)
	Texas Instruments (米国)	Doddington (1976)	CVC 形 1～4 単語	母音中央部(100 ms) のスペクトルマッチング，スペクトル等化処理，計算機室の入室管理に適用	V	男 180 (同)	4 単語 FR 0.3% FA 1%
発声内容不依存形	電電公社武蔵野通研	古井・板倉・斎藤 (1972)	短文章 (10～30 秒)	長時間平均スペクトルのケプストラムの重みつき距離，時期差の効果の分析	I	男 9	90.8
					V	男 9 (26)	93.5
	東北大	松本・曽根・二村 (1977)	短文章 (0.5 秒～)	ケプストラムと基本周波数の断片的正準判別分析	I	男 10	96
	Bell Laboratories (米国)	Sambur (1976)	短文章 (2秒)	種々の線形予測パラメータを直交線形変換し，固有値の小さい成分を使用，音韻性と個人性を分離	I	男 21	94
	SCRL (米国)	Markel Davis (1979)	会話音声 (39 秒～)	PARCOR 係数と基本周波数の統計量（平均値・標準偏差）の重みつき距離	V	男 11 女 6 (同)	96
					I	男 11 女 6	98
	Philips 研究所 (西独)	Bunge (1977)	短文章 (11 秒)	長時間平均スペクトルに対して種々の距離尺度を適用	V, I	男 41 女 9 (同)	98～99
	ITT (米国)	Li Wrench (1983)	会話音声 (3～10 秒)	LPC 分析した各話者の音声をそれぞれ多次元空間内の 40 点にクラスタ化して登録．未知音声フレームとこれらとの距離の統計的分布を使用	I	男 11	79～96

SCRL：Speech Communications Research Laboratory　FR：本人棄却率　FA：詐称者受理率

声帯波スペクトルと声道長の効果が上げられる．また，第1近似としては，声帯波スペクトルの寄与は，主としてスペクトルパターンの概形として，声道長の寄与は，主としてスペクトルパターンの周波数軸方向の伸縮として効いていると考えられる．そこで，このような単純化された考えに基づき，男女各 30 名が発声した 10 数字音声の認識において個人差の正規化，すなわち自動的除去を試みた結果，この方法が認識率の向上に有効であることが確認されている[19]．但し，音声の個人差のすべてを表現するには，このモデルは簡略化されすぎており，更に複雑なモデル化が必要である．

5.3　少数サンプルによる標準パターンの適応化

この方法は，音声の個人差を除去する代りに，認識装置をできるだけ効率よく個人の声に適応させようとするものである．現在実用化されている単語音声認識装置の多くは，あらかじめ利用する人が認識対象語のすべてを少なくとも1回ずつ発声し，標準パターンと

して蓄えておく必要がある．認識対象語の数が大きくなるとこれは大変な労力となるので，少数の言葉だけを発声して，能率的にその人の声の特徴を抽出し，標準パターンを適応化できることが望ましい．このような考え方から，空港名 67 単語を対象とする認識において，12 単語だけをあらかじめ発声することによって，その人の声に適応した標準パターンを生成する試みが行われ，高い認識率が得られている[20]．

あらかじめ学習サンプルを発声せずに，初めは平均的な標準パターンを用意しておいて，認識動作を繰り返すうちに認識装置が自動的に声の特徴を学習して，その人の声に適応してゆくアルゴリズムの研究も行われている[6]．このような機能は人間にもあると考えられ，個人性情報と音韻性情報を自動的に分離してゆく試みとして，今後の地道な研究が必要な分野である．

5.4 識別関数や複数テンプレートによる認識

不特定話者を対象として，既に実用化されている単語音声認識系においては，積極的に個人差を正規化することをねらわずに，個人差によるスペクトルの分布をカバーするような識別関数や，複数の標準パターンを配置する方法が用いられている[21],[22]．これらの方法は，良い認識率を得るための現実的な手段であるが，システムを構成するために大量の音声を必要とするため，語いの追加が容易でなく，システムが複雑化する難点がある．

6. 話者認識の今後の展望と課題

話者認識の具体的応用としては，**2.1** で述べたように種々の例を上げることができる．今後，音声認識や音声応答の実用化によって，電話を用いた種々のサービスが実現されてくると，それに伴って，音声によって個人を識別する必要性が増して来ると考えられる．声は人間個体と分離することができず，暗証コードや磁気カードと違って盗用することができない上，真似することも難しいから，コードやカードと併用すれば，大幅に確実性が増すと期待される．米国空軍の委託を受けた Mitre 社の調査[23]でも，音声は指紋や筆跡よりも誤り率，処理時間のいずれの点でも優れていることが示されている．

このような将来のニーズに答えるために，今後研究しなければならない課題としては，次のようなものがある．

（1） 長期間にわたって安定で，しかも作り声や風邪などに強い特徴パラメータの開発．

（2） 電話機，伝送路等によるひずみや雑音を除去する対策の開発．

（3） 標準パターンのための記憶容量や，分析・認識のための処理量の削減による経済化．

（4） 数千～数万の大量の被験者による長期間の評価実験．

（5） 発声内容不依存形の認識方法の開発．

上記 (5) の課題は，音声波に含まれる個人性情報と音韻性情報の分離という意味で，不特定話者の音声認識の研究と表裏一体をなしており，難しい課題であるが，今後の地道な努力による進展が期待される．

話者認識で用いる音声分析技術の多くは，単語音声認識等で用いられるものと共通であるから，音声分析用の LSI の開発が進めば，発声内容依存形の話者照合を用いた，一般家庭用あるいは車や金庫等の音声キーの実現は，そう遠い将来のことではないかもしれない．

文　献

（1） 古井貞熙："話者認識"，音学誌，**37**, 5, pp. 232-238 (昭 56-05).

（2） 伊藤，斎藤："音声の音響的特徴パラメータが個人性の知覚に及ぼす影響"，信学論 (A), **J 65-A**, 1, pp. 101-108 (昭 57-01).

（3） Tosi, O., Oyer, H., Lashbrook, W., Pedrey, C., Nicol, J. and Nash, E.: "Experiment on voice identification", J. Acoust. Soc. Am., **51**, 6(pt. 2), pp. 2030-2043 (1972).

（4） 坂井，田畑："VCV 音節の多変量解析"，信学論(D), **56-D**, 1, pp. 63-70 (昭 48-01).

（5） 古井貞熙："音声波に含まれる声質情報と韻質情報の分離"，音響学会音声研資，S 74-17 (昭 49-10).

（6） Rosenberg, A.E. and Sambur, M.R.: "New techniques for automatic speaker verification", IEEE Trans. Acoust., Speech & Signal Process., **ASSP-23**, 2, pp. 169-176 (April 1975).

（7） Doddington, G.R.: "Speaker verification", Rome Air Development Center, Tech Rep., RADC 74-179 (1974).

（8） 古井貞熙："話者認識における登録話者数の効果"，昭 52 信学情報全大，228.

（9） 古井貞熙："話者認識と音声認識におけるスペクトル変動の影響の比較"，音響学会講演論文集，pp. 361-362 (昭 52-04).

（10） 古井貞熙："音声の個人性パラメータの時期的変動と話者認識"，信学論 (A), **57-A**, 12, pp. 880-887 (昭 49-12).

（11） Atal, B.S.: "Automatic speaker recognition based on pitch contours", J. Acoust. Soc. Am., **52**, 6 (pt. 2), pp. 1687-1697 (1972).

（12） Sambur, M.R.: "Selection of acoustic features for speaker identification", IEEE Trans. Acoust., Speech, Signal Process., **ASSP-23**, 2, pp. 176-182 (April 1975).

(13) Doddington, G.R. : "Personal identity verification using voice", Proc. ELECTRO-76, 22-4, pp. 1-5 (May 1976).
(14) Rosenberg, A.E. : "Evaluation of an automatic speaker verification system over telephone lines", Bell Syst. Tech. J., **55**, 6, pp. 723-744 (July-Aug. 1976).
(15) Furui, S. : "Cepstral analysis technique for automatic speaker verification", IEEE Trans. Acoust., Speech & Signal Process., **ASSP-29**, 2, pp. 254-272 (April 1981).
(16) 古井貞熙："ケプストラムの統計的特徴による話者認識"，信学論 (A)，**J 65-A**，2，pp. 183-190 (昭 57-02).
(17) 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，pp. 549-556 (昭 47-10).
(18) Li, K.P. and Wrench, Jr., E.H. : "An approach to text-independent speaker recognition with short utterances", Proc. ICASSP 83, 12.9, pp. 555-558 (April 1983).
(19) 古井貞熙："音声認識における個人差の学習法について"，音響学会音声研資，S75-25 (昭 50-11).
(20) Furui, S. : "A training procedure for isolated word recognition systems", IEEE Trans. Acoust., Speech & Signal Process., **ASSP-28**, 2, pp. 129-136 (April 1980).
(21) 千葉，亘理，渡辺："音声認識システム"，大型プロジェクトパターン情報処理システム研究開発成果発表会論文集，通産省工技院，pp. 157-165 (昭 55).
(22) 古井，管村："擬音韻と単語辞書のクラスタ化による不特定話者単語音声認識"，音響学会春季講演論文集，pp. 29-30 (昭 57-03).
(23) "Test results-Advanced development models of BISS identity verification equipment", MITRE Technical Rep. 1, Rev. 1, MTR-3442 (1977).