

論文 / 著書情報
Article / Book Information

論題(和文)	音声波に含まれる声質情報と韻質情報の分離
Title	
著者(和文)	古井 貞熙
Author	SADAOKI FURUI
出典(和文)	日本音響学会音声研究委員会/聴覚研究委員会, 音声総合研究委員会/音響学会研究委員会 資料番号S74-17/H23-5,, , No. S74-17/H23-5,, pp. 1-8
Journal/Book name	, , No. S74-17/H23-5,, pp. 1-8
発行日 / Issue date	1974, 10

昭和49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 S74-17 (1974-10)
H23-5

音声波に含まれる声質情報と韻質 情報の分離

古 井 貞 熙

(日本電信電話公社 武蔵野電気通信研究所)

1974年10月12日

昭和49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
聴 覚 研 究 委 員 会

音声波に含まれる声質情報と韻質情報の分離

古井 貞熙 (日本電信電話公社 武蔵野電気通信研究所)

昭和49年 10月12日 於 国際電電研究所

内容梗概: 発声者と発声内容が共に未知の音声波から、両者の情報を分離して抽出するための試みについて検討した。まず男性9名が長期間にわたって発声した5母音の音声サンプルを用いて、分散分析などの種々の分析を行ない、その結果にもとづいて考案した方法により、男性110名の5母音の音声サンプルを用いて、音韻と話者の認識実験を行った。ここでは個人性情報を、少数のクラスの間で5母音に共通に線形に変化するものとモデル化する。未知の音声に対して、音韻認識と個人性情報抽出を繰返すことによって、音韻の標準パターンを話者に適応させて音韻認識の誤りをへらし、この結果を用いて話者の認識を行う。実験の結果、この方法の有効性が確認された。

On the Method of Separating the Personal and Phonemic Information
in Speech Wave

Sadaoki FURUI (Musashino ECL, NTT)

October 12, 1974 at K.D.D. Research Laboratory

Abstract: A method of separating the personal and phonemic information in speech wave is examined. Several analyses such as analysis of variance are made using the speech samples of five vowels uttered by 9 male talkers in many sessions over a long period. A method of phoneme and talker recognition is designed based on the results of these analyses, and recognition experiments are made using speech samples uttered by 110 male talkers. Talker-dependent variation of five vowels is assumed to be limited on the tracks among a few classes, and the amounts of variation of five vowels are supposed to be in proportion. For an unknown sample, the phoneme recognition and the extraction of the personal information are iterated, standard pattern of phonemes being adjusted for the talker, and the error rate of the phoneme recognition is reduced. Talker recognition is made using this phonemic information. Results of experiments indicate the effectiveness of this method.

1. まえがき

音声（音響レベル）における認識の問題は、いわゆる音韻認識と話者認識にわけることができる。前者は音声波に含まれる音韻性情報を抽出しようとするものである。後者は個人性情報を抽出しようとするものである。これらの二種の情報は、互いに他方がわかっているればその抽出が容易になるという相補的関係にあり、一般に両者の機械認識においてよい認識率を得るためには、他方の情報が既知であるか、あるいは有効な方法によって、他方の情報をキャンセルする必要がある。

人間は、両者の情報が未知の音声から、同時に二種の情報を抽出することができるが、両者は必ずしも別々の空間で独立になされているわけではなく、互いに相互作用をおよぼし合うことによって、それぞれの認識の効果を高め合っていると考えられる。（図1）

二種の情報が未知の音声波から、このような原理にもとづいて、両方の情報を抽出する方法を検討した。

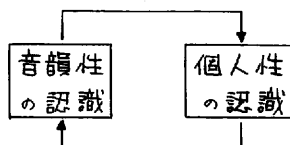


図1. 認識の相互作用

2. 音声分析

次のような方法で、多数の話者（男性）が発声した単姓5母音の音声波の分析を行った。図2の分析部に示すように、音声波を3.4 kHzの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毎に32 msの区間をハミングウィンドウをかけて切り出し、各区間毎にCepstrum または PARCOR 係数を求める。同時に、変形自己相関関数を用いて有声音区間であるかどうかの判定を行う。全有声音区間のCepstrum または PARCOR 係数の平均値を求め、これを各母音音声波の特徴ベクトルとする。特徴ベクトルの次元は、Cepstrum, PARCOR 係数ともに12次元（但次の12成分）とした。

分析に用いたサンプルは、次の2種類である。

〔サンプルα〕：男性9名が3か

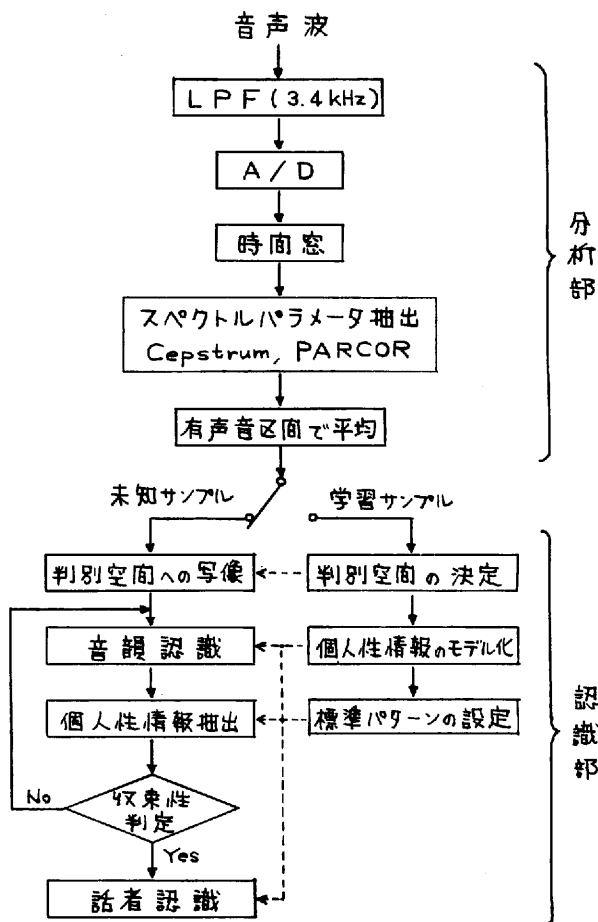


図2. 分析および認識方法

月おきの12時期(ほぼ3年間)に繰返し発声したもの。同じ時期に各母音とも3回発声しているの、サンプルの数は各母音とも $9名 \times 3回 \times 12時期 = 324$ サンプルである。

〔サンプルβ〕: めと異なる男性101名が2回ずつ発声したもの。各母音とも202サンプルである。

3. 多変量分散分析

サンプルαを用い、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%のときの有意水準値で正規化した各効果の大きさを図3に示す。これから次のことがわかる。

1) 3種の効果の中では、音韻効果が圧倒的に大きい。

2) CepstrumとPARCOR係数の結果はよく似ているが、音韻効果はCepstrumベクトルの方が比較的大きく、話者効果はPARCOR係数ベクトルの方が比較的大きい。

3) 音韻-話者交互作用が著しい。このことは、音韻と話者を高い認識率で認識するためには、互いに他方の情報を必要とすることを示している。

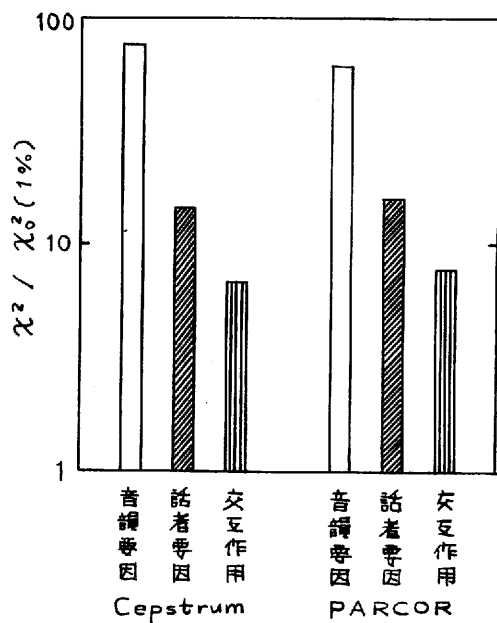


図3. 多変量分散分析結果

4. 主成分分散分析

サンプルαを用いて、CepstrumベクトルとPARCOR係数ベクトルについて主成分分析を行い、各主成分別に音韻要因、話者要因、および音韻-話者交互作用に関する分散分析を行った。危険率1%のときの有意水準値で正規化した各効果の大きさを図4に示す。

この結果から、CepstrumとPARCOR係数のどちらを用いても、音韻性に関する情報の多くは低次の主成分、すなわち特徴ベクトルのマクロ的な特徴として抽出可能であるが、個人性情報は低次から高次までの主成分、すなわち特徴ベクトルのマクロ的な特徴からミクロ的な特徴に分散しており、しかも交互作用に比べてあまり大きくないことがわかる。

5. 判別空間

サンプルαを用いて、判別分析(Discriminant Analysis)の手法により、音韻性空間(〔音韻間分散〕/〔音韻内分散〕を最大にする線形変換)と個人性空間(〔話者間分散〕/〔話者内分散〕を最大にする線形変換)を求めた。

Cepstrumに関して、9名の各話者の5母音の特徴ベクトル(平均値)を、音韻性空間のオ1~オ2主軸からなる平面に射影したものを図5に、個人性空間のオ1~オ2主軸からなる平面に射影したものを図6に示す。図5、6共に、話者Iの5母音を破線で、話者Fの5母音を実線で結んである。これらの結果から、話者Iと

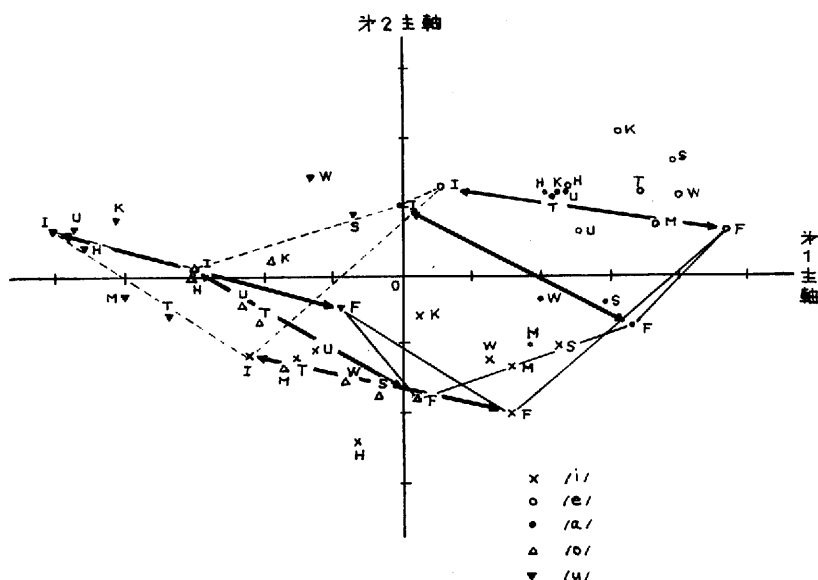
6. 認識方法

以上の分散分析および判別空間へ射影した結果にもとづき、音韻と話者が共に未知の音声波に関して、両者の認識を効率よく行うための方法を検討した。

前に示した図2の認識部は、その具体的方法を示したものである。この方法の特徴は、次の2点にある。

1) まず多数話者に共通の標準パターンを用いて、大ざっぱに音韻認識を行い、その結果を用いて個人性情報を抽出し、さらにその情報を用いて音韻の標準パターンを変更して音韻認識をやり直し、以下結果が収束するまでこの操作を繰返す。最終の音韻認識結果にもとづき、各登録話者の該当する音韻の標準パターンと未知サンプルとの類似度を求め、話者の認識を行う。

2) 個人性情報すなわち個人差を、ケなイクラス (バリエーター) の間で、5 母音に共通に變形に変化するものとしてモデル化する。



四 6. Cepstrum の個人性空間の $\alpha_1 \sim \alpha_2$ 主軸からなる平面に射影した 9 名の話者の 5 母音

6.1. 個人性情情報のモデル化

1) 2クラスを用いる場合

話者（声質）による母音5角形の変化は、両極端に位置する2種類の（多次元）5角形の間で比例的に変化するものとする。この両極端としては、学習サンプルとして9名のサンプルを用いる場合は、話者Fと話者Eの値を用い、学習サンプルとして10名以上の多数の話者のサンプルを用いる場合は、5母音に共通して話者に関する分布の端に位置する計10名を選び、両端にわたるうちに、それぞれ平均化して求める。この2クラスの間の相対的位置として、個人性情報を表現する。

2) 9クラスを用いる場合

話者による母音5角形の変化(バウエティール)として9種類を用意し、このいずれに近いかによつて個人性情報を表現する。学習サンプルとして9名のサンプルを用いる場合は、各話者の平均値をそのまま用い、10名以上の多数の話者のサンプルを用いる場合は、Clusteringを行つて9クラスにまとめる。

図7~8は、40名のサンプルを用いて定めた母音5角形の両極端2クラスと、変化範囲を示したもので、図7はCepstrumの場合、図8はPARCOR係数の場合の結果である。②と③が両極端、①が中央値である。Cepstrumの場合は、②と③の5角

形はほぼ平行移動の関係にあるが、PARCOR係数の場合は回転を伴う変形になっている。

6.2. 認識の手順

1) 学習サンプル

① 多数話者のサンプルを用いて、判別分析により、音韻性空間と個人性空間を設定する。

② 個人性情報のモデル化を行い、声質に関する2クラスまたは9クラスを求める。

③ 話者認識のための登録話者の標準パターンとして、個人性空間で、各話者の各5母音の平均値を求める。

2) 未知サンプル

① 各5母音の標準パターンとして、声質に関する中央値の母音五角形（図7、8の①）を用い、音韻性空間で、未知サンプル5サンプルの音韻認識を行う。標準パターンと未知サンプルの類似度は、ユークリッド距離によって求める。

② 5つの未知サンプルが、それぞれ認識された音韻の個人差に関する変化範囲のうちどこに位置するか（個人性情報）を、音韻性空間と個人性空間で求める。具体的には、図7、8において、該当する（と認識された）音韻の矢印に垂線を下ろし、その足の位置を求める。

③ 5つのサンプルから抽出される個人性情報は、音韻認識の誤りや、サンプルの変動のためにばらつくから、この平均値を求める。

④ こうして求めた個人性情報にもとづいて、音韻性空間で、その値に相当する

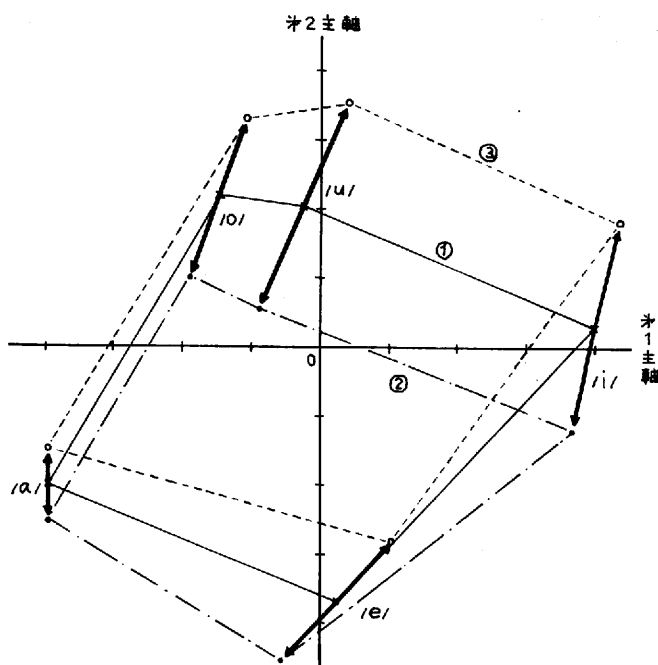


図7. Cepstrumの音韻性空間のF1～F2主軸からなる平面における5母音の変化範囲

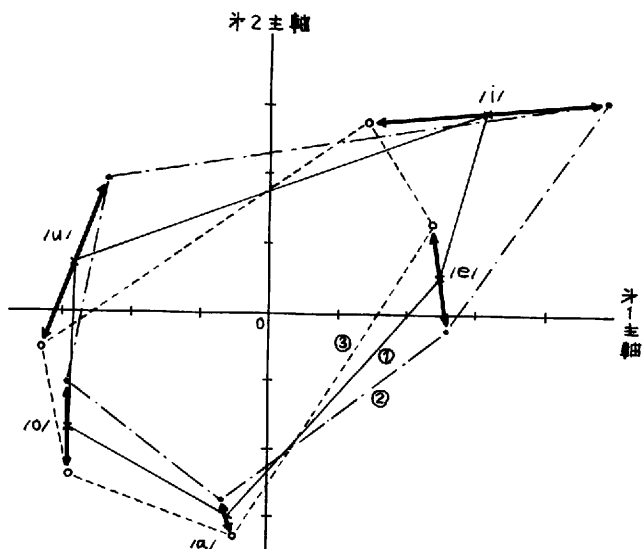


図8. PARCORの音韻性空間のF1～F2主軸からなる平面における5母音の変化範囲

母音5角形を、図7、8の②と③の5角形の間で定める。

⑤この5角形を音韻の標準パターンとして、再び5つの未知サンプルの音韻認識を行う。

⑥この結果にもとづいて個人性情報を再び抽出し、音韻認識結果と抽出された個人性情報が収束するまで以上の操作を繰返す。

⑦収束したら、その最終の音韻認識結果を用いて、5つの各未知サンプル毎に、個人性空間において、該当する音韻の各登録話者の標準パターンとの類似度（ユークリッド距離）を求める。この値を5つの未知サンプルについて平均し、この値にもとづいて話者識別および話者照合の判定を行う。

7. 認識実験

7.1. サンプルと方法

1) 音韻認識

・学習サンプルとして9名のサンプルを用いる場合

{ 学習サンプル：サンプルαのうちの1/2
未知サンプル：サンプルαの残り1/2 とサンプルβ

・学習サンプルとして60名のサンプルを用いる場合

{ 学習サンプル：サンプルαの1/2 とサンプルβの1/2
未知サンプル：サンプルαおよびβの残り1/2

学習サンプルと未知サンプルは、αについては発声時期が3か月以上異なり、βについては発声者が異なる。

2) 話者認識

{ 学習サンプル：サンプルαのうちの1/2

{ 話者識別の未知サンプル：サンプルαの残り1/2

{ 話者照合の未知サンプル：受け入れべきサンプルとして、サンプルαの残り1/2のうちの本人のサンプル。棄却すべきサンプルとして、サンプルαの残り1/2のうちの本人以外のサンプルとサンプルβ。

話者識別は9名から1名を選ぶ形をとり、話者照合の詐称者としては、1回の実験に約60名のサンプルを用いた。

7.2. 認識結果

音韻認識の結果を図9に、話者認識の結果を図10に示す。比較のために、音韻認識については、話者に共通の標準パターン（図7、8の②の5角形）を用いる場合の結果を、話者認識については、音韻情報を用いずに個人性空間における5母音の平均値を登録話者の標準パターンとして用いる場合の結果をあわせて示した。この結果から次のことがわかる。

1) 音韻認識

ここで検討した方法によって話者に適応した標準パターンを用いることにより、学習サンプルと同じ話者(α)の未知サンプルだけでなく、学習サンプルと異なる話者(β)の未知サンプルについても、認識誤り率を低下させることができる。全体的に、個人性情報のモデル化は9クラスよりも2クラスの方がよい。特に学習サンプルと異なる話者(β)に対しては、2クラスの方が有効である。学習サンプルの話者の数は9名でも60名でも結果にはあまり差がない。なお標準パターンを話者に適応

して変更する回数は、1回でほぼ結果が収束することがわかった。

2) 話者認識

音韻認識結果を用いることにより、話者識別、話者照合ともに認識誤り率を大幅に低下させることができる。個人性情報に関して、9クラスのモデル化を行う場合と2クラスのモデル化を行う場合の差は小さいが、2クラスのモデル化を行った方が比較的良い結果を示している。音韻認識に關係する学習サンプルの話者の数は、9名でも60名でも、話者認識結果におよぼす影響はほとんどない。

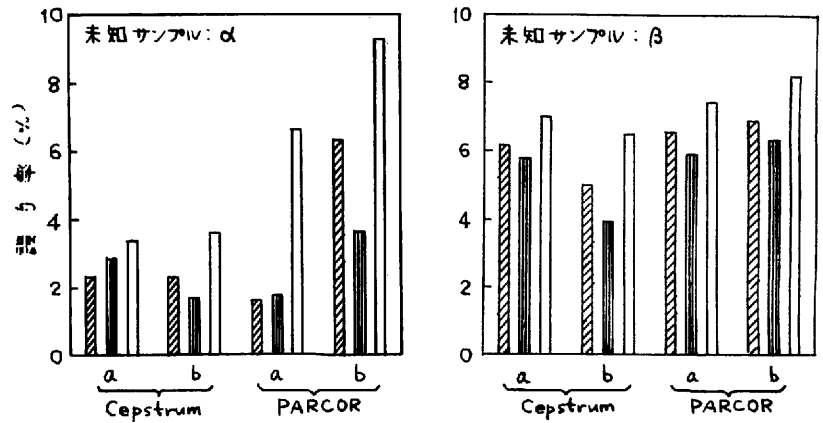


図9. 音韻認識の結果

{ a : 学習サンプルとして9名のサンプルを用いる場合
 { b : " 60名 " "
 □ 話者に適した標準パターン(9クラスモデル化) □ 話者に共通の標準パターン(2クラスモデル化)

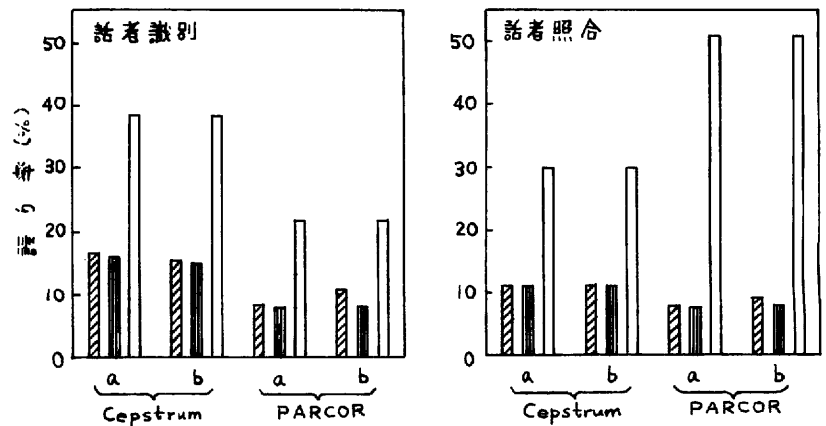


図10. 話者認識の結果

a, b : 音韻認識の学習サンプル数 (図9と同じ)
 □ 音韻情報使用(9クラスモデル化) □ 同(2クラスモデル化) □ 音韻認識せず

8. むすび

発声者と発声内容が共に未知の音声波から、両者の情報を分離して抽出するための簡単な試みについて検討した。まず男性9名が長期間にわたって発声した5母音の音声サンプルを用いて、分散分析

などのデータの分析を行った。次にその結果にもとづいて考案した方法により、男性110名の5母音の音声サンプルを用いて、音韻と話者の認識実験を行った。

その結果、次のような方法が有効であることがわかった。まず個人性情報を極端な2クラスの内で、5母音に共通に線形に変化するものとモデル化する。未知の音声に対しては、初めに多数話者に共通の標準パターンを用いて音韻認識を行い、その結果を用いて個人性情報を抽出する。この情報を用いて標準パターンを変更して音韻認識をやり直し、この音韻情報を用いて話者の認識を行う。

今後は個人性情報のモデル化の改良、音韻の種類の大、単語音声への適用などを検討して行く予定である。

[謝辞] 日通御指導頂く、当研究所野田基穂研究部長、春藤孝4研究室長、ならびに研究室の諸氏に感謝する。