

論文 / 著書情報
Article / Book Information

論題(和文)	単語音声における声質情報と韻質情報の分離
Title(English)	
著者(和文)	古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会1975年講演論文集, , , pp. 399-340
Citation(English)	, , , pp. 399-340
発行日 / Pub. date	1975, 5

古井 貞熙 (日本電信電話公社 武蔵野電気通信研究所)

1. まえがき 音声波に含まれる二種の情報、すなわち韻質情報と声質情報とを分離して抽出することができれば、音声認識および話者認識の高度化のうえで、極めて有用であると考えられる。筆者は、そのような試みの第一歩として、声質情報を少数のクラスの固で5母音に共通に線形に変化するとモデル化し、多数の話者が発声した単独5母音の音声サンプルを用いた認識実験によって、その有効性を確認した⁽¹⁾⁽²⁾。音韻の種類に依存しない声質情報を分離除去したのち、音韻に依存した声質情報のモデル化の方法としては、① 1～多次元の声質空間を構成して、各音韻の特徴空間と声質空間との写像関係を求める方法と、② すべての音韻に共通して距離の近い話者をひとつとめにして、声質情報のクラスタ化を行う方法、の2つが考えられる。

2. 声質空間の設定 各音韻の特徴空間と声質空間との写像関係を線形に限るとすると、この写像方法は、多数の話者が発声した異なる音韻サンプルの間の相関をできるだけ大きくするような変換方法として求めることができる。そこで約100名の男性話者が発声した5単語および、それぞれ30名の男・女性話者が発声した10数字音声から抽出した7～9音韻(5母音と/m,n/, /N/, /R/, /r/)のサンプルを用い、音韻性空間⁽³⁾に写像したのち、異なる二種の音韻のすべての組み合わせについて、正準相関分析を行った。

2.1. 正準相関分析 多数のサンプルについて得られた $X_1 = (x_1, x_2, \dots, x_s)$ と $X_2 = (x_{s+1}, x_{s+2}, \dots, x_{s+t})$ の二組の特性値について、 $U = (u_1, u_2, \dots, u_s)$ と $V = (v_1, v_2, \dots, v_t)$ への線形変換を行って、 $u_k - u_{k'}$ 間、 $v_k - v_{k'}$ 間、 $u_k - v_{k'}$ 間 ($k \neq k'$) が無相関になるようにし、 u_k と v_k の相関 $r_{[k]}$ は大きさの順に並べることにする。このとき第1正準相関 $r_{[1]}$ は、二組のサンプルの間の最大の相関となる。

2.2. 正準相関分析の結果 異なる二種類の音韻間の第1正準相関の値を図1に示す。

すべての音韻の組み合わせに、同程度の相関のあることがわかる。第1正準変量 u_1, v_1 に変換するためのベクトルは、ほとんどの音韻で

/e/	.598
/a/	.516 .698
/o/	.621 .703 .694
/u/	.536 .628 .636 .765
/m,n/	.550 .590 .606 .643 .588
/N/	.620 .653 .619 .732 .567 .734
/l/	/e/ /a/ /o/ /u/ /m,n/

図1. 第1正準相関の値
(男性100名の5単語)

組み合わせる相手の音韻には無関係によく似

ており、これらを平均して、1つの音韻に対して1つのベクトルを定め、このベクトルで変換しても相関係数はあまり低下しないことがわかった。このことは、これらの音韻に共通した声質情報が存在することを示している。図2に、こうして決定した、各音韻に共通した声質情報を表現する第1の軸と、その軸に射影したサンプルの広がりを出す。これらの軸の方向は、5母音ではよく似ているが、/m,n/ と /N/ の軸はやや異なっており、またこれらは、各音韻サンプルの主成分すなわち話者による最大のばらつき方向とは異なっているものが多い。

次に上と同様にして、各音韻に共通した声質情報を表現する第2の軸を設定し、

* On the Separation of Personal and Phonemic Information in Spoken Words. By Sadaoki Furui (Musashino ECL, NTT)

その軸に射影した異なる音韻サンプル間の相関を求めた。その結果、このときの相関の値は、 σ_1 の軸に射影した場合の約 $1/2$ になることがわかった。

2.3. 声質空間への写像 上のようにして各音韻ごとに σ_1 , σ_2 の軸からなる平面に写像したのちに、共通の声質平面に写像し、音韻および繰り返し発声によるばらつきを調べた。話者要因に関する分散分析を行ったところ、極めて大きな有意差を示し、 σ_1 の軸に関しては、検定統計量Fの値は危険率1%の有意水準値の約25倍であった。また各話者の母平均の95%信頼区間幅は、全サンプルの標準偏差の約 $1/2$ であった。

3. 声質情報のクラスタ化

3.1. クラスタ化の可能性 音韻性空間における約100名の話者相互間の距離を各音韻について求め、各音韻における平均距離を μ として分布の大きさの相違を正規化したのち、すべての話者の組み合わせについて、全音韻に関する平均値と標準偏差の関係を求めた。図3は、5単語中の5母音に関する結果を示したものである。同図には、組み合わせの数の分布もあわせて示してある。

これからわかるように、 μ の値すなわち話者間の平均距離がある程度以上になると、 σ の値すなわち音韻による距離のばらつきの大きさがある一定値に収束する。このことは、各音韻に共通した声質情報（話者相互間の距離関係）の存在、すなわちクラスタ化の可能性を示している。

3.2. クラスタ化 前述のようにして正規化した距離に関して、平均距離がある閾値以下の話者の組み合わせについては同一のクラスにするという方法で、約100名の話者のクラスタ化を行った。その結果、閾値を0.25程度にすれば、クラスタ数はもとの $1/2$ 程度になることがわかった。

4. むすび 種々の音韻に共通に含まれる声質情報を抽出する方法として、声質空間への写像方法と、声質情報のクラスタ化について検討した。前者の方法で表現される声質情報は、各音韻の話者によるばらつきを表現する上ではまだ不十分であるが、声質情報を抽象化し、話者による各音韻の特徴パラメータの変動を適応的に正規化する方法として役に立つと考えられる。声質情報を表現するためには、前者と後者の方法のほかに、前処理として分離した音韻の種類に依存しない声質情報などを組み合わせることが必要であろう。

[謝辞] 日頃御指導御鞭撻を頂く 野田基礎研究部長、斎藤特別研究室長ならびに研究室の諸氏に感謝する。
[文献] (1) 古井：音響学全講演論文集 2-2-10 (昭49-10) (2) 同：音響学会音声研究会資料 S74-17 (昭49-10)

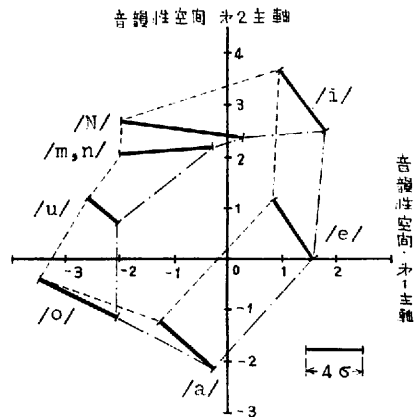


図2. 音韻に共通した声質情報を表現する σ_1 の軸と、その軸に射影したサンプルの広がり

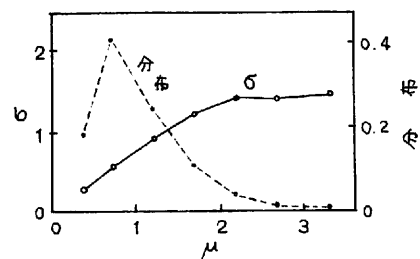


図3. 話者相互間の距離の全音韻に関する平均値と標準偏差