TR-IT-0216

音声の個人性知覚に影響を及ぼす音響的特徴の定量的分析

Quantitative Analysis of Acoustic Features Affecting Speaker Identification

橋本 誠 Makoto Hashimoto
北川 敏 Satoshi Kitagawa
樋口 宜男 Norio Higuchi

1997.3.28

音響特徴の個人性知覚への寄与度と話者間の各音響特徴の差との関係を分析するために、基本周波数、スペクトル、音素継続時間に対して、ABX 聴取実験により聴覚的寄与度を求めると共に、寄与度の予測モデルを提案した。さらに、話者の既知／未知、背景雑音の影響を、統計的手法により分析した。分析の結果、(1) スペクトル、基本周波数は個人性知覚に対する顕著な寄与が認められること、(2) 寄与度は音響特徴の差に依存すること、(3) 既知話者の場合は、未知話者の場合に比べてスペクトルの影響が大きくなること、(4) 雑音重畳音声の場合は、無雑音音声の場合に比べて基本周波数の影響が大きくなる傾向があること、(5) 予測モデルは、6.2 ～ 13.4 % (RMS) の誤差で推定が可能であること、などが明らかとなった。

もくじ

1 はじめに

声質制御機能を備え、多様な合成音声を生成できる音声合成システムの実現は、合成音声の高品質化や音声合成システム自体の普及のために非常に重要であると同時に、多話者間での音声翻訳システムにおける話者識別のためにも重要である。合成音声の多様化の研究は、近年、声質変換や種々の発話様式音声の分析・合成などを中心に盛んになってきている [1][2][3][4][5][6]。特に声質変換など合成音声の話者性制御を精度良く行うためには、音声の持つ個人性情報を把握することが非常に重要であり、そこで得られた知見は、合成音声の多様化だけでなく、話者認識や音声知覚などの分野にも貢献をもたらす。

音声の個人性知覚に関しては、伊藤らが発話者を熟知した被験者による聴取実験を行っており、基本周波数、音韻継続時間長よりもスペクトルが支配的であるという結果を示している [7]。また、北村らは単母音のスペクトル包絡成分に注目した個人性の分析を行っており、個人性情報は高域により多く含まれ、1,740Hz 付近に存在する peak 以上の帯域を利用して話者変換が可能であると述べている [8]。桑原らは、ホルマント周波数とバンド幅に着目した分析を行い、F_3 までのホルマントに個人性情報が含まれると報告している [9]。ただし、これらの報告では、比較する話者間の音響特徴の差（以下、音響特徴距離と呼ぶ）の大小が個人性知覚に及ぼす影響を定量的にモデル化するまでには至っていない。一般に、比較する話者間の基本周波数の差が他の音響特徴距離よりも顕著に大きければ、基本周波数が個人性知覚に支配的となり、スペクトル距離が顕著に大きければ、スペクトルが支配的となることが予想される。つまり、ある特定の話者の組合せで分析を行うのではなく、複数の話者セットを用意し、話者間の音響特徴距離の影響も考慮しながら定量的に分析することによって、個人性知覚に寄与する音響特徴の全体の傾向を把握する必要がある。

このような理由から筆者らは、音響特徴が音声の個人性知覚に及ぼす影響と話者間の音響特徴距離との関係に注目しており、聴取実験による分析を行うと共に、個人性知覚に与える寄与度を予測するモデルを構築することによって個人性判断のメカニズムの解明を試みている [10][11][12][13][14]。

本報告では、聴取実験によって音響特徴の個人性知覚に及ぼす寄与度の聴覚的な分析を行うとともに、音響特徴距離とそれぞれの音響特徴に対する重み係数から寄与度を予測するモデルの提案を行い、聴取実験結果とモデル予測値との両側面から音響特徴の個人性知覚への寄与について考察する。さらに、既知話者／未知話者の違い、無雑音音声／雑音重畳音声の違いなど種々の要因が寄与度に与える影響を明らかにする。

2 実験概要

2.1 分析手順

まず、個人性知覚の分析手順について述べる。本実験は、大きく分けて

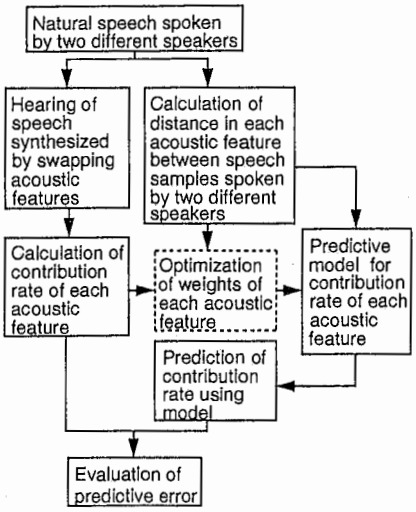

図 1: 実験手順

- 聴取実験によって聴覚的な寄与を定量的に求める
- 話者間の音響特徴距離を利用する寄与の予測モデルを作成する

という 2 つの処理に分けて考えることができる。

聴覚的寄与は、音声の各音響特徴を 2 話者間で個別に置換することにより、各音響特徴の個人性知覚への寄与度が求められるという前提を置き、ABX 聴取実験によって各音響特徴の個人性知覚に及ぼす寄与度を求めた。また、予測モデルは、聴取実験に用いた話者間の音響特徴距離を算出し、これを入力とするものとし、各音響特徴距離に重み係数をかけたものを聴覚的差異と見なして、全体に対する割合で寄与度を出力するものとした。予測モデルの評価は聴取実験から得られた結果と比較することにより行った。音響特徴の寄与度は、予測モデルの最適化を行った際の各音響特徴に対する重み係数を調べることにより得られる。図 1に実験手順を示す。

なお、音響特徴は、基本周波数、スペクトル、音素継続時間の 3 種類に分離した。スペクトルパラメータは、30 次の FFT ケプストラムである。

2.2　聴取実験による線形寄与率の算出

(2.2.1)　聴取実験音声

ABX 聴取実験では、図 2に示すように、A,B はそれぞれ話者 A、話者 B の分析合成音とし、X は話者 A,B 間で基本周波数、スペクトル、音素継続時間の各音響特徴を入れ換えた合成音声とした。被験者には、「X の話者は、A,B のどちらの話者だと思えるか」を強制判断させた。音響特徴を入れ換えた合成音声は、図 3に示すように 6 種類となる。図中、●印は話者 A、

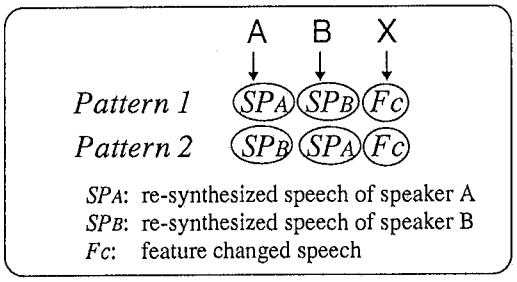

図 2: ABX 聴取実験の呈示音声パターン

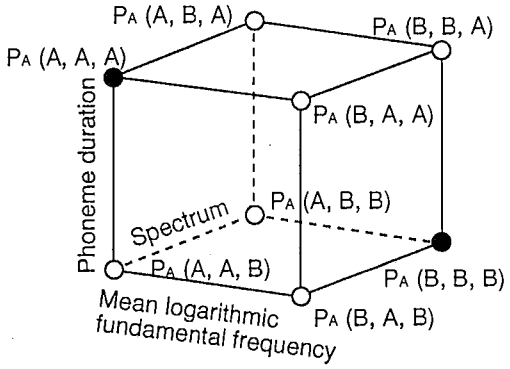

図 3: 各音響特徴を入れ換えた合成音声

話者 B のオリジナル分析合成音を表し、○印は各音響特徴を入れ換えた合成音声を表す。なお、表 1に、本実験で用いた音声の分析条件を示す。

(2.2.2) 聴覚的な線形寄与率の算出方法

本報告では、聴覚的寄与に対して、以下の仮定 1 を置いた。なお、本報告で用いた個人性知覚に対する音響特徴の寄与の度合の定式化は、通常統計学で用いる「寄与率」と異なるために、特に「線形寄与率」と呼ぶこととする。(詳細は 3.3 参照)

仮定 1 話者 B の音響特徴の 1 つを話者 A の音響特徴で置換することにより話者 A という判断が増加する率は相加的で、他の音響特徴に依存しない。

仮定 1 に基づき、聴取実験結果から求める基本周波数の線形寄与率 C_{f0}、スペクトルの線形寄与率 C_{spec}、音素継続時間情報の線形寄与率 C_{dur} を次式で定義した。

$$C_{f0} = C'_{f0}/(C'_{f0} + C'_{spec} + C'_{dur})$$

$$C_{spec} = C'_{spec}/(C'_{f0} + C'_{spec} + C'_{dur})$$

$$C_{dur} = C'_{dur}/(C'_{f0} + C'_{spec} + C'_{dur})$$

表 1: 分析条件

Sampling frequency	12kHz
Window	Blackman window
Window length	21.3ms
Frame shift	5ms
Spectal feature	30-dimensional FFT cepstrum

$$C'_{f0} = \frac{1}{4}\sum\{P_A(A,Y,Z) - P_A(B,Y,Z)\}$$

$$C'_{spec} = \frac{1}{4}\sum\{P_A(X,A,Z) - P_A(X,B,Z)\}$$

$$C'_{dur} = \frac{1}{4}\sum\{P_A(X,Y,A) - P_A(X,Y,B)\}$$

ここで、$P_A(X,Y,Z)$ は、話者 X の基本周波数、話者 Y のスペクトル、話者 Z の音素継続時間を用いて合成された音声試料が話者 A に近いと判定された率で、X,Y,Z は A,B のいずれかである。

2.3　音響特徴寄与のモデル化

さらに、話者の判別における各音響特徴の寄与に関するモデルを作成するに当たって、以下の 2 つの仮定を置いた。

仮定 2 異なる話者間の聴覚的な差は各音響特徴距離に重み係数をかけたものの総和として記述できる。ただし、重み係数は話者に対する熟知度、背景雑音などの諸条件により変化し得る。

仮定 3 各音響特徴距離に重み係数をかけたものの大小は聴取実験によって求められる各音響特徴の線形寄与率に比例する。

これらの仮定に基づいて、各音響特徴の聴覚的な寄与の大きさを表す重み係数を求めるとともに、各音響特徴距離の大小から聴覚的な寄与の大きさを予測するモデルを作成した。

これらの仮定に基づいた 2 種類の音響特徴を用いた場合の音響特徴別線形寄与率予測モデルを図 4に示す。前述した仮定に基づき、本モデルでは、図 4(a) に示す通りそれぞれの音響特徴距離に重み係数を乗じたものの和が全体の聴覚的差異を形成する。更にそれぞれの音響特徴毎の聴覚的差異の大小に比例して話者判別の線形寄与率が決まる（図 4(b)）。つまり、音響特徴

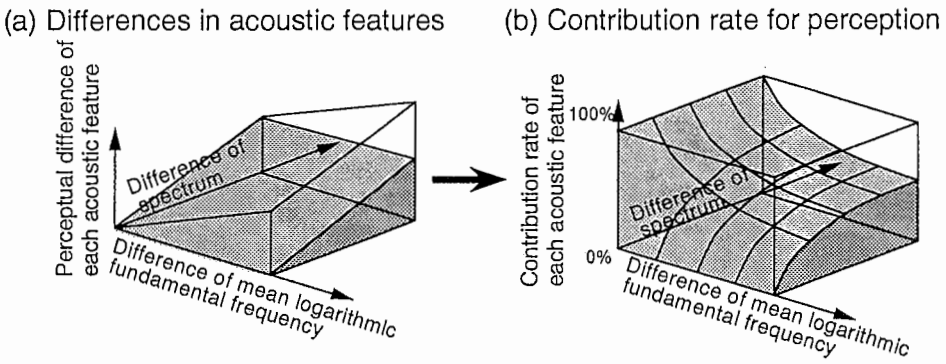

図 4: 線形寄与率予測モデルの概念図（基本周波数とスペクトルの関係）

距離に重み係数をかけたものがその音響特徴に対する聴覚的差異であると見なし、全聴覚的差異に対する割合によって各音響特徴の寄与度を表現する。従って、音響特徴の線形寄与率は、重み係数の比に反映される。

本モデルにおける線形寄与率予測式を以下に示す。

$$\hat{C}_{f0} = W_f \cdot d_{f0}/(W_f \cdot d_{f0} + W_s \cdot d_{spec} + W_d \cdot d_{dur})$$

$$\hat{C}_{spec} = W_s \cdot d_{spec}/(W_f \cdot d_{f0} + W_s \cdot d_{spec} + W_d \cdot d_{dur})$$

$$\hat{C}_{dur} = W_d \cdot d_{dur}/(W_f \cdot d_{f0} + W_s \cdot d_{spec} + W_d \cdot d_{dur})$$

ここで、$\hat{C}_{f0}, \hat{C}_{spec}, \hat{C}_{dur}$ は、それぞれ、基本周波数、スペクトル、音素継続時間の線形寄与率予測値であり、$d_{f0}, d_{spec}, d_{dur}$ は、それぞれ基本周波数、スペクトル、音素継続時間の音響特徴距離である。また、W_f, W_s, W_d は各音響特徴距離の重み係数である。重み係数の最適化には、Analysis-by-Synthesis 法を用いた。

3 線形寄与率の音響特徴距離に対する依存性の分析

3.1 実験条件

まず予備実験として、ABX 聴取実験によって個人性知覚に影響を及ぼす線形寄与率の音響特徴距離に対する依存性を調べた。音声試料は、表 3.1に示す ATR 連続音声データベース [15] 中の 3 文で、男性アナウンサまたはナレータである 6 名 (*MHT,MHO,MMY,MSH,MTK,MYI*) の発声した音声を用いた。また、呈示音声は、話者の中の 1 名を基準話者と定め、基準話者と他の 5 名の同一文とを対とした。被験者は 9 名である。なお、同一音声試料の聴取回数は 4 回とし、順序効果を排除するために A,B の順序を半分ずつ入れ換えた。

呈示音声 A と B との間の音響特徴距離は、基本周波数については平均対数基本周波数の差、スペクトルについてはケプストラム距離とし、音素継続時間については、次式で定義する距離尺度とした。

表 2: 音声試料

ATR 音声データベース中の 3 文
あらゆる現実をすべて自分のほうへねじ曲げたのだ
一週間ばかりニューヨークを取材した
テレビゲームやパソコンでゲームをして遊ぶ

$$d_{dur} = \sum_{i=0}^{S} |d_A(i) - d_B(i)|$$

ここで、$d_a(i), d_b(i)$ は話者 A・話者 B の第 i セグメントの音声の継続時間長、S はセグメント数である。セグメントは、ラベル付けによって始終端が決定されたポーズ以外の音声区間であり、最小単位は音素である。

音声 A と音声 B とのケプストラム距離は、音素毎の時間整合を行ったあと、次式に示す距離尺度 $d(A,B)$ により算出した。

$$d(A,B) = \frac{1}{f} \sum_{ij} (c_{ij}^A - c_{ij}^B)^2$$

ここで、c_{ij}^A は音声 A の i 番めフレーム j 次ケプストラム係数、c_{ij}^B は音声 B の i 番めフレーム j 次ケプストラム係数であり、f はフレーム数である。

3.2　聴覚的な線形寄与率の分析

聴取実験結果から聴覚的な線形寄与率を求め、各音響特徴距離に対する変化を調べた。その結果、

(1) 基本周波数とスペクトルについては寄与があることが認められるが音素継続時間については寄与がほとんど認められないこと

(2) 基本周波数とスペクトルの寄与は、それぞれ、音響特徴距離に依存した傾向があること

が示された。音素継続時間の寄与がほとんど認められない理由として、アナウンサまたはナレータの音声を用いているため個人差が小さかったということが考えられる。

そこで、平均対数基本周波数の差とケプストラム距離からなる 2 次元平面上に、基本周波数およびスペクトルの線形寄与率 C_{f0}、C_{spec} の関係をプロットした。図 5 に結果を示す。図 5 のそれぞれのボックスは、聴取実験に使用された 1 音声サンプル対を示しており、ボックスの中心が音声サンプル間の平均対数基本周波数の差とケプストラム距離を示している。また、各ボッ

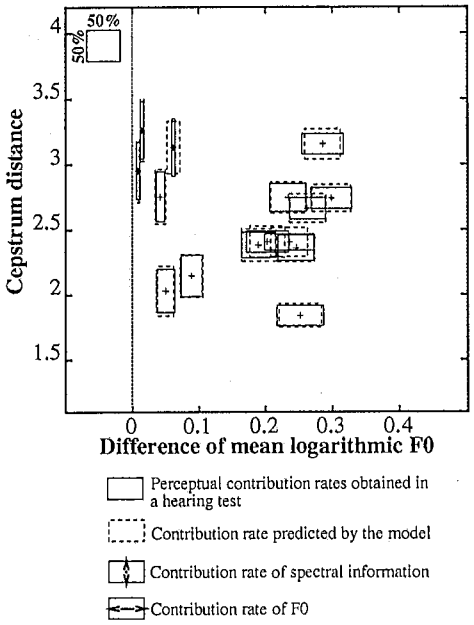

図 5: 音響特徴距離と線形寄与率との関係（実測値とモデル予測値）

クスの横幅・縦幅が、それぞれ C_{f0}・C_{spec} を表す。なお、実線は聴取実験結果を表し、破線は後述する予測値を示す。

図 5から基本周波数とスペクトルの個人性知覚に対する線形寄与率は、それぞれの音響特徴距離が増すほど大きくなる傾向のあることが明らかになった。

このような結果は、ある特定の話者の組合せで分析を行っても個人性知覚に寄与する音響特徴の本質を得ることは難しく、複数の話者セットを用いることによって、話者間の音響特徴距離の影響も考慮しながら定量的に分析することが重要であることを示唆するものといえる。

3.3 線形寄与率予測モデルの評価

次に、音響特徴の線形寄与率予測モデルによる予測を行い、評価した。

線形寄与率予測モデルの評価は、予測値と聴取実験で得られた実測値との誤差により行った。

線形寄与率の大小は、音響特徴距離にかかる重み係数の比から判断できるため、基本周波数の差にかかる重み係数 W_f を 1.0 として、各重み係数の比を求めた。

重み係数の最適化の結果、表 3に示す通り、[平均対数基本周波数の差・ケプストラム距離・音素継続時間の差] の重み係数を、それぞれ [1.000・ 0.062・ 0.102] としたとき、予測誤差が最小値 6.2 %となった。本モデルの予測結果を図 5に破線で示す。この結果から、本モデルにより線形寄与率の予測が良好に行えることが分かる。

なお、通常、統計学では、ある音響特徴として話者 A のものを用いるか否かを説明変数、話者 A と判断された率を目的変数とした場合、相関係数の 2 乗に比例した尺度を寄与率として用いるが、本報告で定義した線形寄与率の方がモデルの予測精度が良いことを予備実験で確認し

表 3: 線形寄与率予測モデルの予測結果

	Difference		Contribution		Prediction Error		
No.	d_{f0}	d_{spec}	C_{f0}	C_{spec}	E_{f0}	E_{spec}	E_{dur}
1	0.089	2.144	0.319	0.638	0.056	-0.080	0.025
2	0.042	2.753	0.132	0.765	0.054	-0.010	-0.044
3	0.050	2.029	0.277	0.662	-0.008	0.014	-0.006
4	0.263	2.663	0.537	0.333	0.058	0.040	-0.098
5	0.286	3.155	0.626	0.309	-0.048	0.085	-0.037
6	0.299	2.742	0.611	0.319	0.005	0.030	-0.034
7	0.190	2.381	0.504	0.390	0.039	0.030	-0.069
8	0.234	2.746	0.542	0.431	0.201	-0.023	0.003
9	0.246	2.360	0.535	0.394	0.074	-0.033	-0.040
10	0.062	3.130	0.048	0.876	0.178	-0.170	-0.009
11	0.015	3.261	0.046	0.881	0.018	-0.012	-0.007
12	0.009	2.951	0.049	0.874	-0.007	-0.015	0.021
13	0.203	2.407	0.622	0.329	-0.074	0.073	0.001
14	0.236	2.405	0.743	0.243	-0.147	0.132	0.015
15	0.252	1.839	0.701	0.313	-0.040	-0.015	0.055
				Mean error	0.012	0.003	-0.015
				RMS error	0.073	0.069	0.041
$W_s = 0.062, W_d = 0.102$ -> RMS error = 6.249 %							

たため、本報告では統計学でいう寄与率ではなく、本報告で定義した線形寄与率を用いた。

4　聴覚的寄与に対する聴取条件の影響

4.1　話者の熟知度の影響

音声の音響特徴のうち、F_0 は比較的容易に変えられるが、スペクトルは話者依存性が強く、話者認識にも広く使われている。前節で述べたモデルを用いることにより、各音響特徴の寄与の大小を定量的に求められることを述べたが、話者の熟知度が寄与の大小に影響を与えることが予想される。

前節の実験は、データベース音声を用いたものであり、言い換えると未知話者音声に対するものであった。そこで次に、聴取する音声の話者が未知話者である場合と既知話者である場合とで線形寄与率および重み係数の比較を行い、話者に対する熟知度の影響を調べた。

(4.1.1) 実験条件

前節の実験では、ABX聴取実験におけるABXの3つの音声は同じ発話内容のものとしていたが、発話毎の局所的な特徴に影響を受ける可能性があったため、本実験では、XをA,Bとは異なる発話内容とした。また、パワーについてもピーク値で正規化した。

既知話者音声は、同じ職場の男性7名の音声で、いずれも無響室にて収録したものである。未知話者音声については、前述の実験で使用したATR音声データベースの男性話者6名で、基準話者を変えた。なお、同一音声試料の聴取回数は2回とした。被験者は前節とは一部異なる9名で、被験者には既知話者として誰の音声を用いたかを事前に知らせた。

(4.1.2) 聴取実験結果

聴覚的な線形寄与率を求めた結果、既知話者／未知話者に拘らず基本周波数とスペクトルについては顕著な寄与があることが認められたが、音素継続時間に関しては前節と同様に寄与は認められなかった。そこで、本実験でも平均対数基本周波数の差、ケプストラム距離に対するC_{f0}, C_{spec}の関係をグラフ化した。図6に既知話者音声に対する結果を、図7に未知話者音声に対する結果を、それぞれ実線で示す。

この結果からも、前節と同様、線形寄与率は音響特徴距離に依存することが示された。また、図6と図7で、各音響特徴距離が類似したサンプルに着目して、その線形寄与率を比較すると、未知話者音声よりも既知話者音声に対する方が、スペクトルの線形寄与率が大きいという傾向が見られた。

(4.1.3) モデルによる線形寄与率予測結果

次に、線形寄与率予測モデルによって、線形寄与率の予測を行い、各音響特徴距離に乗じる重み係数の比に着目して、既知話者音声の場合と未知話者音声の場合を比較した。重み係数の最適化の結果、[平均対数基本周波数の差・ケプストラム距離・音素継続時間の差]の重み係数を、既知話者の場合[1.000・0.122・0.027]に、未知話者の場合[1.000・0.079・0.056]にした時、予測誤差が最小となった。この時のモデルの予測誤差(RMS error)は、既知話者音声に対しては10.6％、未知話者に対しては13.4％であった。本モデルの予測結果を、図6、図7に、それぞれ破線で示す。

(4.1.4) 既知話者 / 未知話者の影響の分析

線形寄与率予測モデルでは、各音響特徴距離に対する重み係数の比によって、各音響特徴の寄与の大小を表現できるが、重み係数の次元がそれぞれ異なるため、単一条件の実験結果で数値の大小を論ずることに意味はない。しかしながら、異なる聴取条件間における同一音響特徴の重み係数の比は、同一の次元を持っており、そのまま寄与の大小の比と考えることができる。

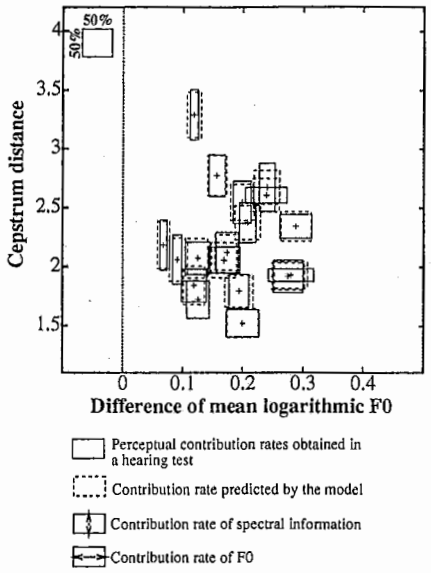

図 6: 既知話者音声に対する音響特徴距離と線形寄与率との関係（実測値とモデル予測値）

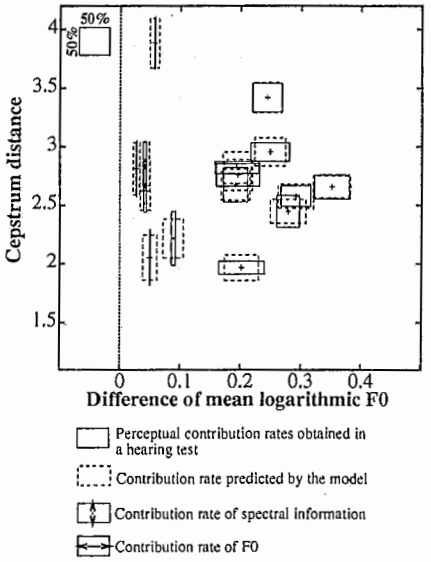

図 7: 未知話者音声に対する音響特徴距離と線形寄与率との関係（実測値とモデル予測値）

すなわち、4.1.3 において既知話者および未知話者に対するスペクトルの重み係数がそれぞれ 0.122、0.079 であったことから、既知話者音声のケプストラム距離に対する重みが未知話者のそれよりも約 1.5 倍（= 0.122/0.079）大きくなっており、音声の個人性を知覚する際、既知話者音声に対する場合の方が、未知話者音声に対する場合よりも、スペクトルの影響が大きくなることが示された。

この結果が有意差を有するかどうかを調べるために統計的手法による分析を行った。結果を図 8に示す。図 8は、被験者毎の各実験サンプルに対して線形寄与率予測モデルの重み係数を求め、基本周波数に対する重み係数とスペクトルに対する重み係数との比の分布を示したものである。図 8のグラフは、ノッチの部分が 95 %信頼区間を表し、上方向ほど基本周波数の線形寄与率がスペクトルの線形寄与率よりも高いことを示す。この結果から、既知話者音声に対す

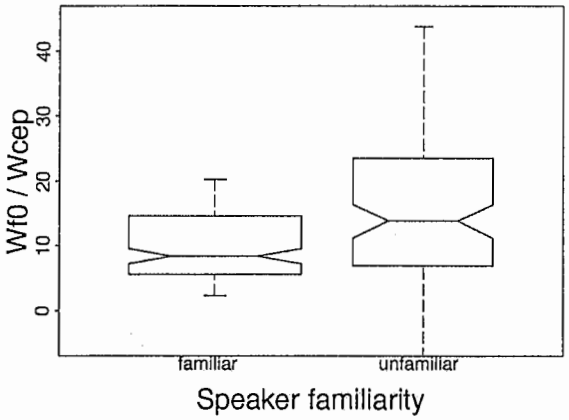

図 8: 既知話者音声／未知話者音声に対する重み係数の分布

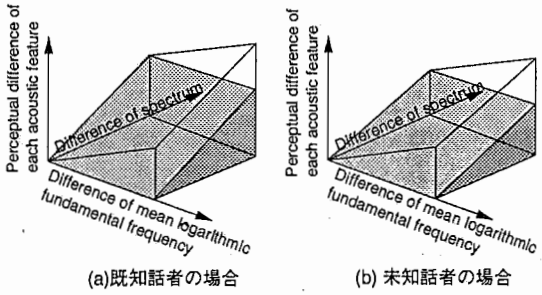

図 9: 既知話者／未知話者における線形寄与率存在空間の違い

る場合の方が、未知話者音声に対する場合に比べてスペクトルの影響が危険率５％で有意な差を持って大きいということが示された。

これは、模式的に書くと図9のように聴覚的寄与の大小が既知話者／未知話者で異なっているということを示しており、前述したように図6、図7の類似した位置におけるボックスの形状の違いも間接的にこれを表しているものと考えられる。

このように、音響特徴が音声の個人性知覚に及ぼす影響は、音響特徴の話者間距離に依存すると同時に、話者が既知か未知かにも依存し、既知話者の場合はスペクトルが支配的になるということが示された。これは、被験者にとって既知である発話者に対してはスペクトルが個人性知覚に支配的であるとする文献 [7] の報告とも一致するものであるとともに、話者の既知／未知の違いや話者間距離などの状況によって、支配的となる音響特徴やその線形寄与率の大きさが変化するということも示している。

表 4: 音声試料

ATR 音声データベース中の 3 文
朝の光線を逆光気味に受けてススキが美しく光る
私がふとつぶやくと父は大きく首をふった
お金を入れボタンを押すと切符が出てくる

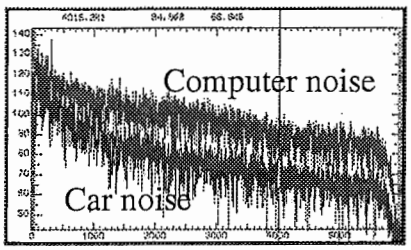

図 10: 各雑音の周波数特性

4.2　背景雑音の影響

4.1 の実験では、未知話者音声／既知話者音声とも、背景雑音のない状態で収録した音声を用いていたが、雑音環境下ではスペクトル情報の一部がマスクされ、線形寄与率が変化することが考えられる。そこで次に、未知話者データに雑音を重畳させて、線形寄与率にどのような変化が生じるかを調べた。

(4.2.1)　実験条件

これまでの実験の影響を避けるため、音声試料を表 4に示す 3 文に変えた。被験者は 8 名である。

重畳雑音としては，背景雑音の違いの影響を調べるため、電子協騒音データベースの「走行自動車内 (2000cc クラス)」、「計算機室 (ワークステーション)」の 2 種類を用い、S/N 比を変えて分析した。図 10に両雑音の周波数特性を示す。

(4.2.2)　実験結果

線形寄与率を求めた結果、雑音を重畳した場合においても、基本周波数とスペクトルについては寄与があることが認められたが、音韻継続時間に関しては顕著な寄与は認められなかった。そこで、ここでも基本周波数に対する線形寄与率とスペクトルに対する線形寄与率について分析を行うこととした。図 11に、平均対数基本周波数の差、ケプストラム距離に対する各線形寄与率の関係を示す。

図 11(a) ～ (e) は、それぞれ S/N 比∞ dB （雑音非重畳） , 5dB(自動車内), 5dB(計算機室),

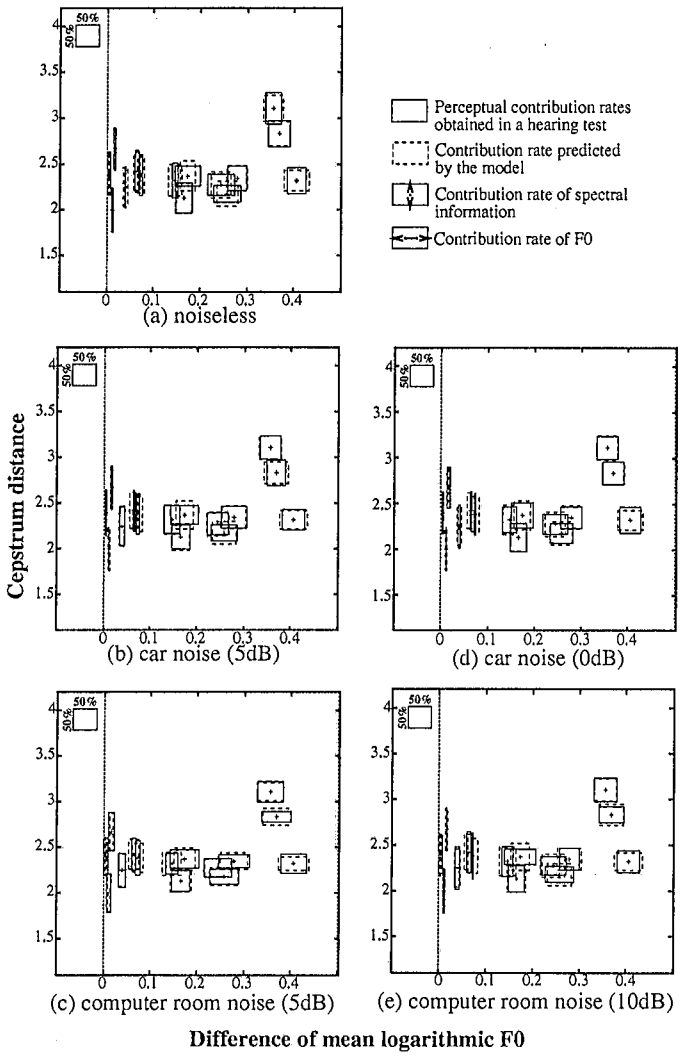

図 11: 雑音重畳音声の音響特徴距離と線形寄与率との関係（実測値とモデル予測値）

0dB(自動車内), 10dB(計算機室) における結果を示したものであり、それぞれ、実線は聴取実験結果から求めた線形寄与率、破線はモデル予測値である。

線形寄与率予測モデルにおける最適重み係数 $[W_f \cdot W_s \cdot W_d]$ は、それぞれ、

[1.0 • 0.149 • 0.077] (S/N 比 $= \infty$ dB)

[1.0 • 0.118 • 0.074] (5dB, 自動車内)

[1.0 • 0.080 • 0.177] (5dB, 計算機室)

[1.0 • 0.133 • 0.136] (0dB, 自動車内)

[1.0 • 0.112 • 0.050] (10dB, 計算機室)

であった。この時の予測誤差 (RMS error) は、それぞれ 9.0 %, 7.5 %, 8.2 %, 7.0 %, 9.9 % であり、提案モデルは、雑音重畳音声に対しても良好に線形寄与率を予測できることが示された。

次に、雑音非重畳音声と雑音重畳音声との間で線形寄与率がどのように変化したかを調べるため、変化の様子をグラフ化した。図 12に結果を示す。図 12は、縦軸がスペクトルの線形寄

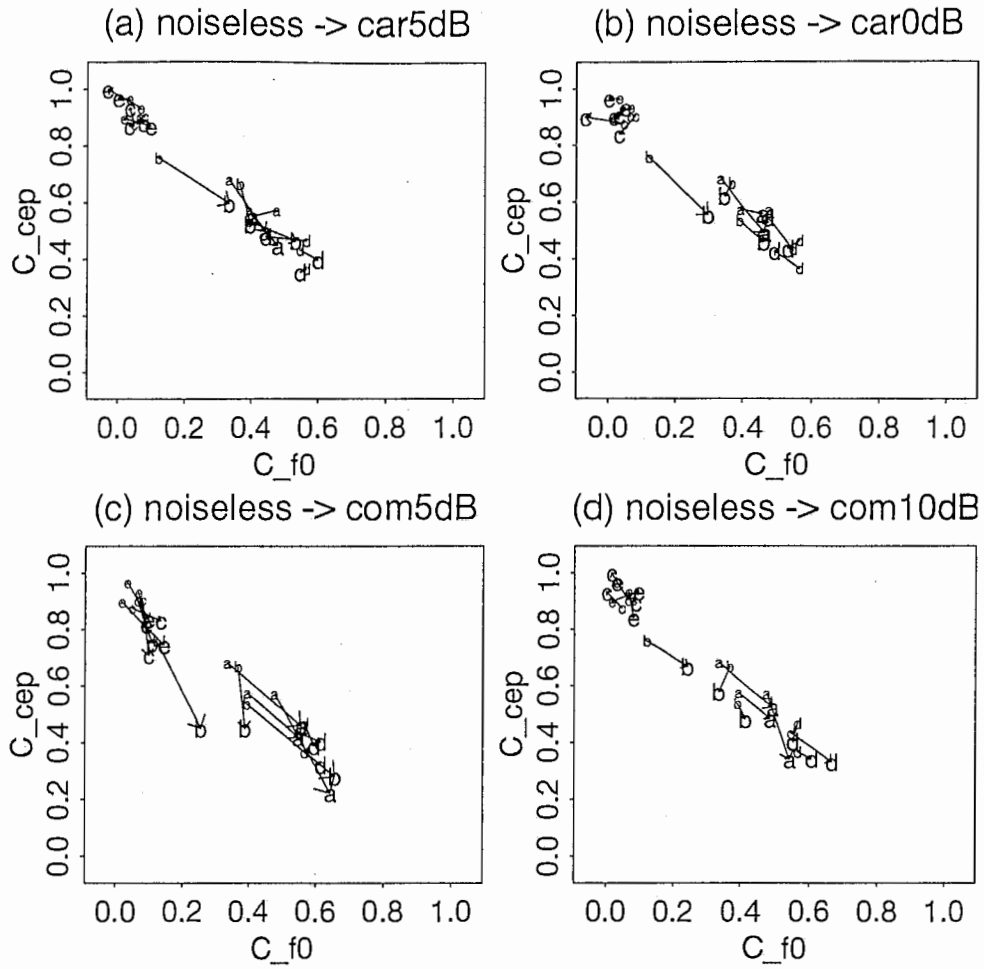

図 12: 背景雑音による線形寄与率の変化

与率、横軸が基本周波数の線形寄与率であり、各プロットに対して雑音非重畳音声に対する線形寄与率から雑音重畳音声に対する線形寄与率への変化を矢印で表現したものである。この図から、雑音重畳音声に対しては、基本周波数の線形寄与率が大きくなる傾向が伺える。

そこで、4.1.4 と同様に基本周波数とスペクトルそれぞれの線形寄与率にかかる重み係数比の分布によって、有意差を調べた。分布図を図 13に示す。図 8と同様に図 13も上方向ほど基本周波数の線形寄与率の方が大きいということを表す。

(4.2.3)　背景雑音の影響の分析

以上の結果から、

(1) 雑音環境下音声に対しても、各音響特徴の線形寄与率は音響特徴距離に依存すること

(2) 同じ S/N 比の場合、計算機室雑音の場合の方が自動車内雑音の場合よりも基本周波数の寄与が増加すること

(3) S/N 比が低下するほど基本周波数の寄与が増加する傾向があり、しかも、自動車内雑音よりも計算機室雑音の場合の方が基本周波数の寄与の増加傾向が速まること

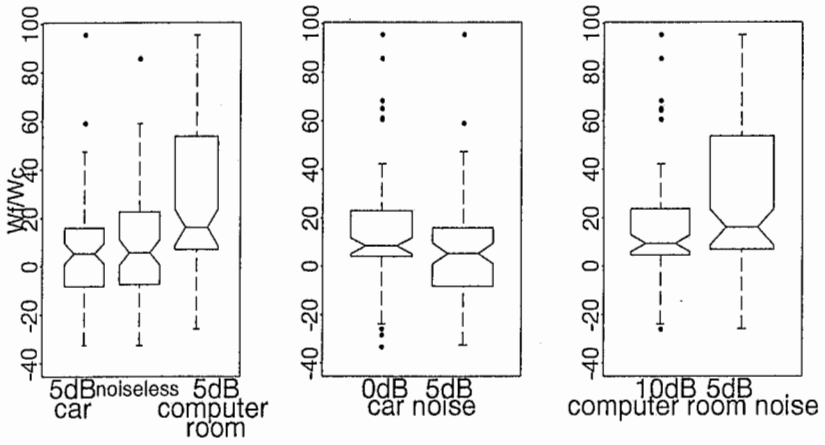

図 13: 各雑音重畳音声に対する重み係数の分布

などが明らかとなった。計算機室雑音の場合の方が基本周波数の寄与が大きくなる原因としては、図10に示すように自動車内雑音よりも計算機室雑音の方が音声をより広い帯域にわたってスペクトル情報をマスクするためであると考えられる。

5 むすび

個人性知覚に影響を与える音響特徴を、音響特徴の差（音響特徴距離）との関係から分析した。ABX 法による聴取実験から、基本周波数、スペクトル、音素継続時間の 3 つの要因について、個人性知覚への寄与度（線形寄与率）を求めた。また、聴覚的な線形寄与率を予測するモデルの構築を試み、聴取実験から得られた結果との予測誤差により、これを評価した。さらに、既知話者音声と未知話者音声、無雑音音声と雑音環境下音声における線形寄与率の変化について、聴取実験結果と線形寄与率予測モデルの予測結果から統計的手法によって分析した。この結果、

- 個人性知覚に関しては、基本周波数とスペクトルに、顕著な寄与が認められる。
- 音響特徴の線形寄与率は、音響特徴距離が増すほど大きくなる。従って、特定の音声サンプルの線形寄与率のみを見ても寄与の度合は把握できず、線形寄与率の音響特徴距離に対する変化を見ることが重要である。また、音響特徴距離に重み係数をかけたものをその音響特徴に起因する聴覚的な差異と見なし、それが全聴覚的差異に対してどの程度の割合であるかを調べれば、各音響特徴の個人性知覚に対する振舞いを知ることができる。
- 音響特徴の線形寄与率は、話者などの条件によって異なり、既知話者音声と未知話者音声を比較すると、既知話者音声の方がスペクトルの寄与が大きくなる傾向がある。その

差は、線形寄与率予測モデルのケプストラム距離にかかる重み係数から、既知話者音声の方が未知話者音声の場合に比べて約1.5倍大きい。

- 音響特徴の線形寄与率は、背景雑音などの条件によっても異なり、雑音を重畳した場合と雑音を重畳しない場合とを比較すると、S/N比の低下に伴って基本周波数の寄与が増加するが、自動車内雑音よりも計算機室雑音の場合の方が基本周波数の寄与の増加が速まる。

- 線形寄与率予測モデルの予測誤差は6.2～13.4％となり、本モデルにより聴覚的線形寄与率の推定が良好に行えることが示された。これは言い換えると、本モデルで表される空間が聴覚的線形寄与率の存在する空間に非常に近いものであるといえる。

ということが示された。

以上のように、本報告では日本語の朗読調音声に対する個人性知覚についての分析結果を示した。本報告で示した結果が、他の言語や他の発話様式の音声にも適用できるかどうかは明確ではない。従って今後は、他言語に対する分析、自然発声音声・電話音声など発話様式や発話環境が異なる場合の音声に対する分析などを行っていきたい。

謝辞

研究の機会を与えて頂いた、ATR 音声翻訳通信研究所山崎泰弘社長に感謝致します。また、聴取実験に御協力を頂いた皆様、および日頃から熱心に御討論頂く ATR 諸氏、特にニック・キャンベル主幹研究員に感謝致します。

参考文献

[1] M.Abe, S.Nakamura, K.Shikano and H.Kuwabara : "Voice conversion through vector quantization", Proc. ICASSP'88, pp.565–568 (1988).

[2] 松本 弘, 丸山靖史, 井上博夫 : "教師あり / 教師なしスペクトル写像による声質変換", 音響学会誌, **50**, pp.549–555 (1994).

[3] N.Iwahashi and Y.Sagisaka : "Speech spectrum conversion based on speaker interpolation and multi-functional representation with weighting by radial basis function networks", SPEECH COMMUNICATION, **16**, pp.139–151 (1995).

[4] M. Hashimoto and N. Higuchi : "Spectral Mapping for Voice Conversion Using Speaker Selection and Vector Field Smoothing", Proc. EUROSPEECH'95, pp.431–434 (1995).

[5] 阿部匡伸 : "異なる発話様式の特徴分析とその制御", 音響学会誌, **51**, pp.929–937 (1995).

[6] N. Higuchi, T. Hirai and Y. Sagisaka : "Effect of speaking style on parameters of fundamental frequency contour", Proc. The Second ESCA/IEEE Workshop on Speech Synthesis, pp.135–138 (1994).

[7] 伊藤憲三, 斉藤収三 : "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, **J65-A**, pp.101–108 (1982).

[8] 北村達也, 赤木正人 : "単母音の話者識別に寄与するスペクトル包絡成分", 音響学会誌, **53**, No.3, pp.185–191, 1997.

[9] H. Kuwabara and Y. Sagisaka : "Acoustic characteristics of speaker individuality : Control and conversion", Speech Communication, **16**, pp.165–173, 1995.

[10] 橋本 誠, 樋口宜男 : "個人性の知覚に影響を及ぼす音響的特徴の分析", 音講論集, pp.323–324 (1995.3).

[11] N. Higuchi and M. Hashimoto : "Analysis of acoustic features affecting speaker identification", Proc. EUROSPEECH'95, pp.435–438 (1995).

[12] N. Higuchi and M. Hashimoto : "Analysis of acoustic features affecting speaker identification", J. Acoust. Soc. Japan(E), **17**, pp.33–35 (1996.1).

[13] 橋本 誠, 樋口宜男 : "音声の個人性知覚に影響を及ぼす音響的特徴の分析", 信学技報, SP96-59, pp.29–36 (1996.10).

[14] 北川 敏, 橋本 誠, 樋口 宜男：“雑音環境下における音声の個人性知覚の分析”, 音講論集, pp.263–264 (1997.3).

[15] 阿部匡伸, 匂坂芳典, 梅田哲夫, 桑原尚夫：“研究用日本語音声データベース利用解説書”, Tech. report of ATR, TR-I-0166 (1990).