日本音響学会誌53巻3号（1997），pp. 185-191

論　　文

43.70.Gr；43.71.Bp；43.72.Ja

単母音の話者識別に寄与するスペクトル包絡成分*

北村達也*1　赤木正人*1

（1996年5月2日受付）

［要旨］単母音のスペクトル包絡において個人性が顕著に現れる帯域とその帯域において話者識別に寄与する成分についての検討を行った。スペクトル包絡の特定の帯域を変形させた刺激音を用いた聴覚実験により，スペクトル包絡の変形と個人性知覚との定量的な関係を求めた。その結果，以下のことが明らかになった。（1）個人性はスペクトル包絡全体に現れるが，高域により多く現れる。（2）話者識別にはスペクトル包絡のdipよりもpeakが重要な意味を持っている。（3）個人性は音韻によらずスペクトル包絡の20 ERB rate（1,740 Hz）付近に存在するpeak以上の帯域に顕著に現れる可能性が高く，この帯域を利用して話者変換が可能である。（4）この帯域のpeakを3角形で近似しても個人性が保存される。

キーワード　話者識別，個人性情報，スペクトル包絡，話者変換
Speaker identification, Speaker individuality, Spectral envelope, Voice quality control

1. はじめに

近年，音声合成の分野における個人性の重要性が注目されるに伴い，話者変換（声質変換）に関する研究が盛んに行われている[1),2)]。その中で，個人性を制御することを目的として，個人性を表している物理量の調査も行われるようになってきた[3),4)]。個人性を表す物理量に関する研究は，話者変換への応用のみならず，話者認識への応用や人間の知覚過程の解明へとつながるものとして重要な意義がある。

音声における個人性は基本周波数とスペクトル包絡に多く含まれると言われている。また，個人性は母音に顕著に現れると言われている。本研究では，人間が話者を識別する際に利用している物理量が個人性を表す重要な物理量であるという作業仮説のもとで，単母音のスペクトル包絡において個人性が顕著に現れる帯域と，その帯域において話者識別に寄与する成分についての検討を行う。

従来より，個人性とスペクトルの周波数帯域との関係が調べられている。桑原ら[5)]は5母音のみを含む無意味音声を対象として，ホルマント周波数とバンド幅を独立に制御する方法により声道特性と個人性の関係を調べた。その結果，F4以上よりもF3までのホルマントにより多くの個人性情報が含まれ，特にF3が最も重要であることを示した。阿部[6)]は連続音声を対象とした音声モーフィングに関する研究の中で，低域のスペクトルが個人性知覚に重要であると述べている。これらの研究は連続音声を対象にしている。連続音声のスペクトル包絡にはその定常部に含まれる個人性と，動的な時間変化に含まれる個人性が混在していると考えられる。後者が個人性知覚に重要な意味を持っていることに疑いの余地はないが，我々は定常部に含まれる個人性に関する研究が連続音声における個人性の研究の基礎となると考えている。

Furuiら[7)]は電話帯域の単語音声を対象として個人性知覚と種々の物理量との関係を調べた。そして，スペクトル包絡に定常的に現れる個人性を捉えていると考えられる時間平滑スペクトル包絡の2.5～3.5 kHzの帯域と心理的距離の相関が高いことを示した。しかし，この研究では単語全体のスペクトル包絡を平均しているため，音韻によるスペクトル包絡の違いが考慮されていない。この違いは個人性知覚に影響を与える可能性があるため，スペクトル包絡に関する個人性は音韻ごとに調べる必要がある。

以上の点から，本研究では単母音を対象として，スペクトル包絡の変化と個人性知覚との定量的な関係を求めることを目的とする。

我々は，従来より単母音のスペクトル包絡における個人性に関する研究を行ってきた[8)]。そして，話者間で平均したスペクトル包絡のF3以上の帯域をある話者のもので置換した音声を用いて話者識別実験を行い，個人性はこの帯域に顕著に現れることを示した。

しかし，従来の我々の研究では話者が3名と少な

* Significant cues in spectral envelope of isolated vowels for speaker identification, by Tatsuya Kitamura and Masato Akagi.

*1 北陸先端科学技術大学院大学情報科学研究科
（問合先：北村達也　〒923-12 石川県能美郡辰口町旭台1-1 北陸先端科学技術大学院大学情報科学研究科）

く，得られた結果が話者セットに依存している可能性がある。そこで，本研究の実験 1 では話者 9 名の単母音を用いて，スペクトル包絡において個人性が顕著に現れる帯域を調査する。

合成音声における個人性の制御や人間の話者識別過程の解明を目的とするとき，個人性が顕著に現れる帯域を調べるだけでは不十分である。この帯域のどの成分が話者識別の手がかりとなっているのかを明らかにする必要がある。そこで，実験 2 では F 3 以上の帯域において話者識別に寄与する成分について調査する。特に，スペクトル包絡の全体的な形状を決定しているスペクトルの peak と dip が知覚的に重要であろうという推測から，これらが話者識別に与える影響について検討を行う。

実験 3 では，ここまでの結果を整理して，再び個人性が顕著に現れる帯域を調査する。更に，スペクトル包絡における個人性情報の表現を簡略化し，制御を容易にすることを目的として，この帯域の peak を 3 角形で近似することを試みる。

2. 実験 1 個人性が顕著に現れる帯域の検討 1

実験 1 では単母音のスペクトル包絡において個人性が顕著に現れる帯域を ABX 法により調査する。スペクトル包絡を 0～10，10～20，20～30 ERB rate の 3 帯域に分割し，個人性がどの帯域により多く現れるかを調査する。

ERB rate は等価矩形帯域幅 (Equivalent Rectangular Bandwidth: ERB) を幅 1 として周波数軸を変形したものである[9),10)]。ERB と ERB rate はそれぞれ以下の式で求められる。

$$ERB = 24.7(4.37F + 1) \quad (1)$$

$$ERB\ rate = 21.4 \log_{10}(4.37F + 1) \quad (2)$$

ここで，F は周波数 (kHz) である。式 (2) は基底膜上の周波数マッピングを近似的に求めている[11)]。スペクトル包絡を ERB rate 上で等間隔に分割したのは，基底膜上の周波数の表現に対応させるためである。

2.1 実験条件

2.1.1 音声データ

ATR 音声データベース[12)]の男性話者 9 名 (mau, mht, mmy, mnm, msh, mtk, mtm, mtt, mxm) による標本化周波数 20 kHz の日本語 5 母音 (タスクコード SY)。

2.1.2 刺激音

刺激音は LMA 分析合成系[13)]により合成した。LMA フィルタの作成に用いるケプストラムは，改良ケプストラム法[14)]により求めた。分析条件はフレーム長 25.6 ms，フレーム周期 6.4 ms，加速係数 1.0，近似回数 3 である。刺激音 A，B，X には以下のものを用いた。

A: 話者 9 名中 1 名のスペクトル包絡を持つ合成音声

B: 話者間で加算平均したスペクトル包絡を持つ合成音声

X: B のスペクトル包絡の以下の 4 帯域を A のもので置換した合成音声

X 1. 全帯域 (刺激音 A と同じ)

X 2. 0～10 ERB rate (0～442 Hz)

X 3. 10～20 ERB rate (442～1,740 Hz)

X 4. 20～30 ERB rate (1,740～5,544 Hz)

刺激音 A は各話者の各音韻を有声区間で時間平均した 60 次までのケプストラム c_A から合成し，刺激音 B は c_A を音韻ごとに話者間で加算平均したケプストラム c_B から合成した。刺激音 X は c_A と c_B を用いて合成した。刺激音 X 2 を例に作成方法を説明する。初めに，c_A と c_B に 512 点 DFT をかけて対数スペクトラム s_A, s_B を得る。次に，s_B の 0～10 ERB rate を s_A の 0～10 ERB rate で置換する。置換した対数スペクトラムに 512 点 IDFT をかけ，再びケプストラムを得る。このケプストラムから LMA フィルタを作成し，合成音声を得る。変形を加えた対数スペクトルには不連続点が生じることがある。しかし，LMA フィルタの作成には 60 次までのケプストラムを用いるため，合成音声のスペクトル包絡に不連続点はほぼ存在しなくなる。

実験 2，3 で用いるスペクトル包絡を変形した合成音声の作成も，これと同様の方法で行う。すなわち，ケプストラムをいったん対数スペクトラムに変換し，この領域で変形を加えた後，再びケプストラムに変換し，LMA フィルタを作成して合成音声を得る。

合成音声の平均基本周波数は 130 Hz である。刺激音の長さは 0.5 s で，振幅を正規化し，更に刺激音の前後部 0.05 s を sin 関数で重み付けした。

本研究で用いる刺激音における違いはスペクトル包絡のみである。よって，実験結果はスペクトル包絡の違いのみに起因する。

2.1.3 被験者

正常聴力を有する 23～25 歳の大学院生 10 名 (男性 9 名，女性 1 名)。

2.1.4 実験方法

ABX 法により行った。同じ音韻の刺激音 A，B，X を約 2 s の間隔で呈示し，X の話者が A と B の話者のどちらに似ているかを強制判断させた。継時効果

Table 1 聴取実験に使用した機器

WS	Sun S-4/IX
PC	Macintosh PowerBook Duo
Headphone	STAX SR-λ pro. (exp. 1) STAX λ Nova Signature (exp. 2 and 3)
Amplifier	STAX SRAM-1/MK-2 pro.
DSP	M.I. Systems VMEDSP 56 K Engine
LPF	NF CORPORATION P-86

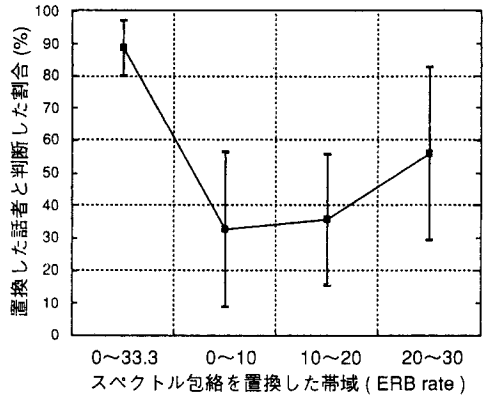

Fig.1 置換した話者の音声であると判断した割合と標準偏差

を打ち消すために，BAX の順についても実験を行った。A，B，X の三つの刺激音の組を 1 刺激とし，1 刺激につき ABX，BAX を各 3 回，計 6 回呈示し，1 回の実験では 150 刺激を呈示した。

被験者は防音室内でヘッドフォンにより受聴した。受聴は各被験者の聞き易いレベルによる両耳受聴である。被験者には聞き直しを許し，パーソナルコンピュータ（PC）を用いて回答させた。なお，刺激音の呈示中は PC の HDD を停止させるため，PC によるノイズは発生しない。

刺激音は防音室の外に設置されたワークステーション（WS）内に保存されており，被験者の応答に応じて呈示される。WS から出力された刺激音は D/A 変換され，更に 8 kHz（33.3 ERB rate）の LPF を通過させることにより高域に発生するノイズを除去した。聴取実験に使用した機器を **Table 1** に示す。

2.2 実験結果と考察

スペクトル包絡の一部の帯域を置換した話者の音声であると判断した割合を **Fig.1** に示す。100%の場合は置換により完全に話者が変換されたことを意味し，0%の場合は置換が話者識別に全く影響を与えないことを意味している。置換した帯域が 0～33.3 ERB rate の場合はスペクトル包絡の全帯域を置換した場合に相当する。

Fig.1 から置換する帯域が高くなるに従い，値が増加する傾向があることが分かる。このことは音声の個人性はスペクトル包絡の全帯域に現れるが，高域により多く現れ，この部分を置換することによって話者変換の効果が得られることを示している。

実験 1 では ABX 法により実験を行ったため，被験者が音色の違いによる判断を行っていた可能性もあり，結果が個人性のみの影響を反映したものかについては疑問が残る。そこで，以下では被験者が知っている話者の音声を用いて話者の名前を回答させる naming 法により実験を行う。

3. 実験 2 peak と dip が話者識別に与える影響の検討

実験 2 では，スペクトル包絡の F 3 以上の帯域において話者識別に寄与する成分について調査する。スペクトル包絡の全体的な形状を決定している peak と dip を除去した音声を用いて話者識別実験を行い，これらが話者識別に与える影響について調査する。

3.1 実験条件

3.1.1 音声データ

音声データは，基本周波数が 125 Hz 前後である 24～26 歳の男性 5 名による日本語 5 母音である。本研究では，スペクトル包絡における個人性に関する実験を行うため，話者ごとの基本周波数の違いが話者識別に与える影響を極力抑える必要がある。そこで，録音の際，話者に 125 Hz の純音をヘッドフォンにより呈示し，それに声の高さを合わせるよう指示した。

録音は騒音レベル 22.7 dB(A)の防音室にて行った。マイクロフォンからの距離を約 15 cm に保って発声させた音声を防音室の外の DAT レコーダに入力し，標本化周波数 48 kHz で録音した。この音声を 20 kHz に変換して WS に保存し，更に定常部約 200 ms を切り出して音声データとした。録音に使用した機器を **Table 2** に示す。

3.1.2 刺激音

刺激音は音声データから LMA 分析合成系を用いて合成した。刺激音の平均基本周波数は 125 Hz である。これ以外の分析合成に関する条件は実験 1 と同じである。

実験 2 で用いた刺激音は下記の 4 種類である。F 3 は目視により決定した。回帰直線は ERB rate における F 3 以上の帯域の対数スペクトル包絡に関するものである。なお，これらの刺激音の音韻性が保存されていることは実験前に確認してある。

2 a. LMA 分析合成音声

Table 2 録音に使用した機器

Microphone	SONY C-536 P
DAT recorder	SONY TDC-D 10 PRO II
Headphone	STAX SR-λ pro.
Amplifier	STAX SRAM-1/MK-2 pro.

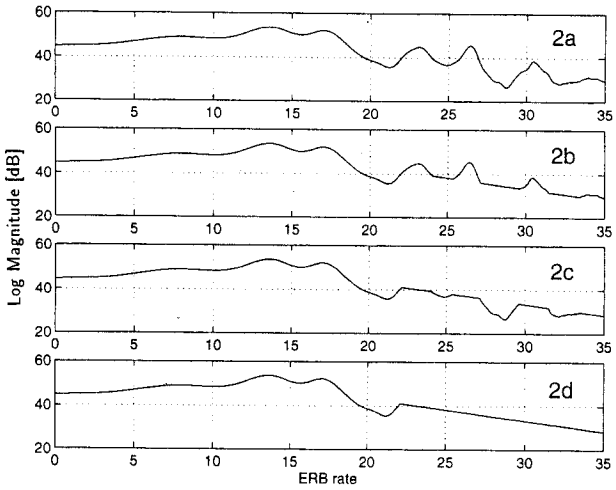

Fig.2 刺激音2a, 2b, 2c, 2dのスペクトル包絡
（1段目：2a，2段目：2b，3段目：2c，4段目：2d）

2b. F3以上の帯域の回帰直線より小さい成分を回帰直線によって置換した合成音声

2c. F3以上の帯域の回帰直線より大きい成分を回帰直線によって置換した合成音声

2d. F3以上の帯域を回帰直線によって置換した合成音声

全刺激音のF3未満の帯域は各話者のスペクトル包絡を用いている。**Fig.2**に1話者の/a/の音声データをもとにした各刺激音のスペクトル包絡を示す。

3.1.3 被験者

正常聴力を有し，音声データの話者と日頃接している24〜29歳の男性6名。

3.1.4 実験方法

上述の4種類の刺激音をランダムに並べ変え，4等分したものを1セッションとした。1セッションは125個の刺激音から成っている。一つの刺激音は4セッションのうちに5回現れる。被験者には実験1と同じ条件で受聴させ，刺激音の話者を強制判断させた。回答はPCのディスプレイ上の話者の名前が書いてあるボタンをクリックすることにより行わせた。

3.2 実験結果と考察

2aの話者識別率に関する被験者間の平均値と標準偏差を**Table 3**に示す。標準偏差が比較的大きいことから，短時間のスペクトル包絡のみを利用して話者を識別する能力には個人差があることが分かる。そこで，実験2では各被験者の2aの話者識別率から2b，2c，2dの話者識別率を減じた値（減少値）により評価を行う。被験者間で平均した減少値を**Fig.3**に示す。

Table 3 刺激音2aに対する話者識別率と標準偏差

	/a/	/i/	/u/	/e/	/o/
話者識別率（%）	91.3	92.0	83.3	94.0	67.3
標準偏差	6.3	7.7	11.2	11.7	11.2

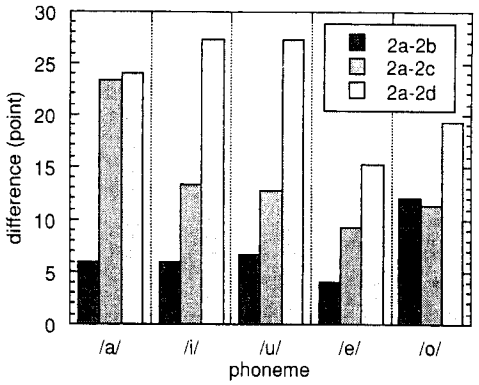

Fig.3 実験2の減少値

これらの結果について有意水準5%の分散分析を行ったところ（$F(1,58:0.05)=4.01$），2aと2bの減少値の間（$F(1,58)=21.1$），2bと2cの減少値の間（$F(1,58)=8.45$），2cと2dの減少値の間（$F(1,58)=7.27$）に有意差があった。これより，各刺激音の話者識別率が2a>2b>2c>2dという関係にあることが分かる。

この結果から，F3以上の帯域におけるpeakやdipは話者識別に寄与し，特にpeakが重要であることが分かる。これは，人間の聴覚ではスペクトルのpeakが重要であるという従来からの知見と矛盾しない。また，2cはF3以上の帯域におけるpeakの上部を除去したものと見なすことができ，上述の結果はpeakの周波数やパワーが話者識別に寄与することを示している。しかし，話者識別率に2a>2bという関係もあることから，人間が話者識別に利用しているのはpeakの周波数やパワーの情報だけではないことが分かる。これらの結果は，peakとdipのパワー差も話者識別に利用されている可能性を示唆するものであると考える。

4. 実験3 個人性が顕著に現れる帯域の検討2

我々は先の研究[8]で，単母音のスペクトル包絡にお

ける個人性はF3以上の帯域に顕著に現れることを示した。この実験結果を音韻ごとに分析したところ，/a/, /u/, /o/ではF3以上の帯域に個人性が顕著に現れるが，/i/と/e/では顕著ではないことが分かった。

一方，実験1によりスペクトル包絡における個人性が20 ERB rate以上の帯域により多く現れることが示されている。以上の点と/i/と/e/のF2が20 ERB rate付近に現れることを併せて考えると，単母音のスペクトル包絡における個人性は音韻によらず20 ERB rate付近に存在するpeak以上の帯域に顕著に現れることが推察される。

実験3ではこの推察を確認するため，この帯域における情報で話者識別が可能か否かを調べる。更に，スペクトル包絡における個人性情報の表現を簡略化し，制御を容易にすることを目的として，この帯域のpeakを3角形で近似することを試みる。

4.1 実験条件

4.1.1 音声データ

実験2と同じ，男性5名による日本語5母音の定常部約200 ms。

4.1.2 刺激音

刺激音はLMA分析合成系を用いて合成した。これらの刺激音の音韻性が保存されていることは実験前に確認してある。実験3で用いた刺激音は以下の2種類である。

3a. 以下のスペクトル包絡を持つ合成音声
　低域…話者間で平均したスペクトル包絡
　高域…回帰直線より小さい成分を回帰直線によって置換

3b. 3aにおいて高域のpeakを3角形で近似した合成音声

ここで，下線付きの「高域」は「20 ERB rate付近に存在するpeak以上の帯域」を表し，「低域」は「20 ERB rate付近に存在するpeak未満の帯域」を表す。簡単のため以降ではこの表記を用いる。**Fig.4**にFig. 2と同じ話者による/a/, /i/, /u/の20 ERB rate付近に存在するpeakと高域の範囲を図示する。

北村ら[15]は0～10 ERB rate（442 Hz）の帯域におけるパワーの違いも話者識別に寄与することを報告した。低域を話者間で平均する際にはこの点を考慮し，話者を0～10 ERB rateに大きなパワーを持つ2名とそれ以外の3名のグループに分けた。そして，スペクトル包絡の0～10 ERB rateの帯域は各グループ内で平均したスペクトル包絡により置換し，10 ERB rate以上の帯域は5名間で平均したスペクトル包絡により置換した。

peakを3角形で近似する方法は以下のとおりである。ここでpeakとはスペクトル包絡の高域において回帰直線よりも大きい値を持つ成分のことを意味している。3角形の頂点となるpeakの頂点を目視により決定し，その頂点と，スペクトル包絡と回帰直線の交点を直線で結ぶ。これによりpeakの周波数とパワーとバンド幅が大まかに近似される。本研究で用いた音声データに対しては，この方法で決定される3角形の個数が4個以内におさまった。**Fig.5**にFig. 2と同じ音声データから作成した3aと3bのスペクトル包絡を示す。

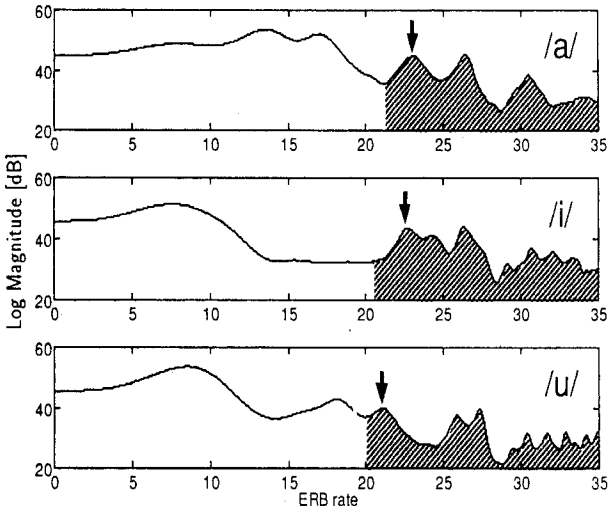

Fig.4　20 ERB rate付近に存在するpeak（矢印）と高域の範囲（斜線）

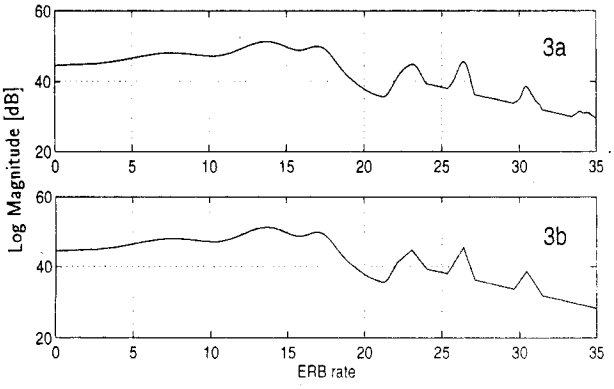

Fig.5　刺激音3aと3bのスペクトル包絡（1段目：3a, 2段目：3b）

4.1.3 被験者

実験2と同じ男性6名。

4.1.4 実験方法

上述の刺激音をそれぞれ1セッションとして実験を行った。呈示順序はランダムであり，一つの刺激音は5回呈示される。呈示条件や回答方法は実験2と同じである。

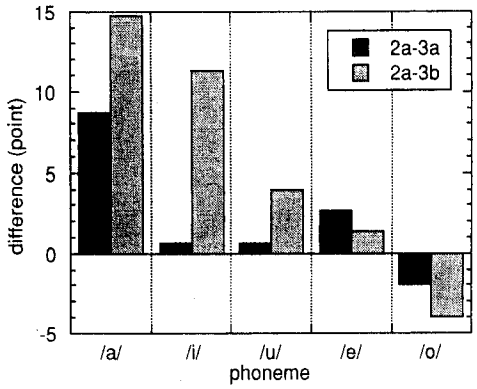

Fig.6 実験3の減少値

4.2 実験結果と考察

実験3の結果も実験2と同様に各被験者のLMA分析合成音声(2a)の話者識別率から3a,3bの話者識別率を減じた値（減少値）により評価を行う。被験者間で平均した減少値を**Fig.6**に示す。

3aと2b(F3以上の帯域において回帰直線より大きい値を持つ成分を回帰直線により置換した音声）の減少値について有意水準5%の分散分析を行ったところ，有意差は見られなかった($F(1,58)=3.76$)。2bと3aにおける大きな違いは低域を話者間で平均しているか否かである。低域を話者間で平均したことによる影響がないことから，話者識別には高域がより重要であることが分かる。また，実験に用いた刺激音の音韻性が保存されていることから，話者に関して平均したスペクトル包絡の高域を変換したい話者のもので置換することにより，音韻識別に影響を与えずに話者変換の効果が得られることが分かる。

次に，3aについて音韻により減少値に有意差があるか否かを調べたところ($F(4,25:0.05)=2.76$)，音韻間には有意差が見られなかった($F(4,25)=0.91$)。この結果と，先の研究で得られた/a/,/u/,/o/ではスペクトル包絡のF3以上の帯域に個人性が顕著に現れ，/i/と/e/では顕著ではないという結果[8]を併せて考えると，単母音のスペクトル包絡における個人性は音韻によらず20 ERB rate付近に存在するpeak以上の帯域（高域）に顕著に現れる可能性が高いことが分かる。

一方，3aと3bの減少値の間には有意差が見られなかった($F(1,58)=1.26$)。これはスペクトル包絡における個人性情報の表現を簡略化できる可能性を示唆している。また，話者識別には高域におけるpeakの周波数とパワーとバンド幅の情報が重要であることを示唆している。

しかし，ほとんどの被験者から3bに対する話者識別の困難さを指摘された。これは，peakを一つの3角形により近似することは個人性の劣化を引き起こすことを示している。しかし，3角形を複数用いることにより個人性の劣化を抑えつつ個人性情報の表現を簡略化することができる可能性が高く，音声合成等への応用が期待できる。

5. 総合考察

以上三つの実験から，単母音のスペクトル包絡における個人性に関して以下の結果が得られた。

1. スペクトル包絡における個人性はスペクトル包絡全体に現れるが，高域により多く現れる。
2. 話者識別にはスペクトル包絡のpeakが重要な意味を持っている。peakとdipのパワー差が重要である可能性もある。
3. 個人性は音韻によらず高域に顕著に現れる可能性が高い。また，高域を利用して話者変換が可能である。
4. 話者識別にはpeakの周波数とパワーとバンド幅が重要であることが示唆された。

以上のことから，単母音の話者識別にはスペクトル包絡の高域におけるpeakが重要な意味を持つ可能性が高いという結論が得られる。従って，人間が話者を識別する過程では，ホルマントの順番を数えて何番目かのホルマントから個人性を抽出するような処理は行われておらず，高域から個人性を抽出するような処理が行われている可能性が高いと考えられる。

また，実験3により得られたスペクトル包絡における個人性は高域に顕著に現れる可能性が高いという結果は，刺激音2dと3aの結果と矛盾していない。刺激音2dはF3以上の帯域を回帰直線で置換しているため，被験者は主にF3未満の帯域における個人性を手がかりに話者識別を行っていると考えられる。一方，3aは低域を話者間で平均しているため，高域における個人性を手がかりにしていると考えられる。刺激音2dと3aの音韻間で平均した話者識別率は，それぞれ62.9%，83.4%である。この結果も，スペクトル包絡における個人性は高域により多く現れることを示している。

本研究で個人性が顕著に現れるとした高域は，Furuiら[7]が個人性知覚の心理的距離との相関が高いとした時間平滑スペクトル包絡の2.5～3.5 kHzの帯域を含む帯域である。よって，本研究の結果はFuruiらの結果を支持するものであると言える。

更に，党らの結果[16]は本研究の結果を生成系から支持するものである可能性がある。党らによれば，喉頭部における声道の分岐である梨状窩（pyriform fassa）は2～6 kHzの音声スペクトルに大きな影響

を与える。梨状窩は声道内で相対的に不変な部分であるため，その音響特性は個人性の要因の一つである可能性があるとしている。この帯域は，本研究で高域と呼んだ帯域とほぼ一致している。これは，音声の個人性に対する梨状窩の影響を示唆するものである。

6. おわりに

本研究では単母音のスペクトル包絡において話者識別に寄与する成分に関する検討を行った。その結果，単母音の話者識別にはスペクトル包絡の20 ERB rate付近に存在するpeak以上の帯域（高域）におけるpeakが重要な意味を持つ可能性が高いことを示した。これは，人間が話者を識別する際にはこの帯域から個人性を抽出していることを示唆するものである。

更に，これらのpeakを3角形で近似しても音声の個人性が保存されることを示した。この方法によりスペクトル包絡における個人性情報の表現を大幅に簡略化できる可能性があり，音声合成等への応用が期待できる。

しかし，本研究の実験結果が話者セットに未だ依存している可能性は否定できない。そのため，大規模な話者セットを用いた実験を行う必要がある。更に，連続音声ではスペクトル包絡や基本周波数の時間変化が話者識別へ与える影響が大きくなることが予想される。今後，連続音声における個人性に関する検討を行う必要がある。

謝　辞

本研究の一部は文部省科学研究費補助金（課題番号07680388）及び日本学術振興会特別研究員奨励費によって行われたものである。

文　献

1) 小坂直敏, “Sinusoidal modelを用いた母音の声質補間,” 音講論集, 263-264 (1995.9).
2) H. Mizuno and M. Abe, “Voice conversion algorithm based on piecewise linear conversion rules of formant frequency and spectrum tilt,” Speech Commun. **16**, 153-164 (1995).
3) H. Kuwabara and Y. Sagisaka, “Acoustic characteristics of speaker individuality : Control and conversion,” Speech Commun. **16**, 165-173 (1995).
4) N. Higuchi and M. Hashimoto, “Analysis of acoustic features affecting speaker identification,” Proc. EUROSPEECH '95, Vol. 1, 435-438 (1995).
5) 桑原尚夫, 大串健吾, “ホルマント周波数・バンド幅の独立制御と個人性判断,” 信学論 **J69-A**, 509-517(1986).
6) 阿部匡伸, “基本周波数とスペクトルの漸次変形による音声モーフィング,” 音講論集, 259-260 (1995.9).
7) S. Furui and M. Akagi, “Perception of voice individuality and physical correlates,” 音響学会聴覚研資 H 85-18 (1985).
8) T. Kitamura and M. Akagi, “Speaker individualities in speech spectral envelopes,” J. Acoust. Soc. Jpn. (E) **16**, 283-289 (1995).
9) B.R. Glasberg and B.C.J. Moore, “Derivation of auditory filter shapes from notched-noise data,” Hear. Res. **47**, 103-138 (1990).
10) 赤木正人, “聴覚フィルタとそのモデル,” 信学会誌 **77**, 948-956 (1994).
11) D. Greenwood, “A cochlear frequency-position function for several species—29 years later,” J. Acoust. Soc. Am. **87**, 2592-2605 (1990).
12) 武田一哉, 匂坂芳典, 片桐　滋, 阿部匡伸, 桑原尚夫, “研究用日本語音声データベース利用解説書,” ATR Tech. Rep. TR-I-0028 (1988).
13) 今井　聖, 北村　正, “対数振幅特性近似フィルタを用いた音声の分析合成系,” 信学論 **J61-A**, 527-534 (1978).
14) 今井　聖, 阿部芳春, “改良ケプストラム法によるスペクトル包絡の抽出,” 信学論 **J62-A**, 217-223 (1979).
15) 北村達也, 高木直子, 赤木正人, “個人性情報を含む周波数帯域について,” 信学技報 SP 95-37 (1995).
16) 党　建武, 本多清志, “母音発声時の音声スペクトルに対する梨状窩の影響,” 信学技報 SP 95-10 (1995).

北村　達也

平4年山形大・工・情報工卒。平6年北陸先端科学技術大学院大学情報科学研究科博士前期課程修了。音声の個人性情報に関する研究に従事。現在，同大学博士後期課程在学中。日本音響学会，電子情報通信学会各会員。平7年より日本学術振興会特別研究員。

赤木　正人

昭54年名工大・工・電子卒。昭59年東工大大学院博士課程情報工学専攻了。工博。同年電電公社（現NTT）研究所入社。以来，ATR視聴覚機構研究所，NTT基礎研究所を経て，現在，北陸先端科学技術大学院大学情報科学研究科助教授。この間，昭63年米国MIT客員研究員，平5年オランダアムステルダム大学客員研究員。音声信号処理，聴覚機構のモデル化の研究に従事。日本音響学会，電子情報通信学会，IEEE, ASA, ESCA各会員。