

## 論文

## 音声における個人性の知覚と生成について

北村 達也

甲南大学 知能情報学部 知能情報学科  
神戸市東灘区岡本 8-9-1, 658-8501

(受理日 2008 年 11 月 3 日)

## 概要

音声の個人性（話者の特徴）に関して，著者らが音声知覚と音声生成の両方の観点から行ってきた研究について概説する．前者は，心理物理学的手法によって個人性知覚に寄与する特徴量を明らかにすることを目的としている．後者は，MRI による生体観測や電気回路モデルを用いたシミュレーションによって，声道形状の個人差と伝達特性との関係を明らかにすることを目的としている．これらの研究の結果，母音のスペクトル包絡の高周波数帯域における個人差が個人性知覚に顕著に寄与し，この周波数帯域の個人差が下咽頭腔の形状の個人差に由来することを明らかにした．この周波数帯域は話者認識にも有効である．また，プロの物真似タレントが他人の個人性を模倣した音声を分析した結果についても述べる．

キーワード：個人性，母音，スペクトル包絡，磁気共鳴画像法（MRI），声道，下咽頭腔，物真似

## 1 はじめに

音声には文字として書き起こすことができる言語情報以外にも，性別，話者，年齢，体調，感情などの情報が含まれている．これらは我々が顔から視覚的に得ているような情報と類似しているため，“auditory face”と呼ばれることがある．また，言語情報（linguistic information）に対して，“パラ言語（周辺言語）情報（para-linguistic information）”，“非言語情報（non-linguistic information）”と呼ぶこともある．Fujisaki [1] の定義に従えば，パラ言語情報とは発話の意図や話者の心的態度を意味し，非言語情報とは話者の身体性に関する情報である．著者らは，これら言語情報以外の情報のうち特に話者の特徴，すなわち個人性に関して，主に母音を対象にその知覚要因と生成要因に関する研究を行ってきた．それらの研究で得られた結果は，双方からの知見を補完しあうものになっている．本論文ではその研究について概説するとともに，残された課題についても述べる．

顔が一人ひとり異なるように，音声も人それぞれの特徴，すなわち個人性を持っている．音声の個人性に対応する特徴は，時間的に静的か動的か，生得的か後天的かという 2 つの軸で分類することができる．時間的に静的な特徴には，母音のスペクトルや声帯音源特性などがあり，いずれも身体の生得的な構造に由来する．一方，時間的に動的な特徴には話し方や訛りなどがあり，これらは後天的に獲得する個人性である．

音声の個人性知覚メカニズムを明らかにすることができれば、話者の個人性に適応して音声から言語情報を抽出する人間の能力の解明や、その技術的実現などにつながっていく。一方、音声の個人性生成メカニズムを明らかにすることができれば、個人性を自在に制御した音声的合成できるようになる。

本論文では、まずスペクトル包絡において個人性を顕著に含む周波数帯域の分析結果を示し、その結果にもとづいて行われた聴取実験について述べる。続いて、聴取実験により個人性知覚に寄与することが明らかになった周波数帯域が話者認識技術にも有効であることを示す。次に、磁気共鳴画像法 (MRI) を用いて音声の個人性生成要因に関して行った研究について解説する。さらに、音声の個人性を極めてよく模倣することができるプロの物真似タレントの音声の分析結果を示し、最後に、今後の課題について議論する。

## 2 母音のスペクトル包絡の話者間、話者内分散 [2]

母音のスペクトル包絡において個人性を多く含む周波数帯域を明らかにするため、男女各 10 名の単母音のスペクトル包絡を対象にして話者間、話者内分散を求めた。話者間分散の大きい周波数帯域は個人差が大きく、話者内分散の大きい周波数帯域は母音間の差が大きいことになる。スペクトル包絡とはスペクトルの概形であり、声道の音響特性に対応する。音声データは ATR 音声データベースの単母音を用いた。標本化周波数は 20 kHz で、スペクトル包絡を求める際のケプストラム<sup>1</sup> 次数は 60 である。

男性話者を対象とした分析結果を図 1 に示す。話者間分散は 22 ERB rate<sup>2</sup> (2,212 Hz) 以上の周波数帯域で大きく、この周波数帯域において個人性が大きいことがわかる。一方、話者内分散は 12~22 ERB rate (603~2,212 Hz) の周波数帯域で大きく、母音間の差異はこの周波数帯域で大きいといえる。女性話者でも同様の傾向が見られた。Li and Hughes [3] および Mokhtari and Clermont [4] も同様の結果を報告している。

## 3 母音の個人性の知覚要因 [2]

前節の分析結果にもとづいて、母音のスペクトル包絡における周波数帯域と個人性知覚および母音知覚との関係について聴取実験を行った。

### 3.1 刺激音

成人男性 5 名の日本語 5 母音を標本化周波数 20 kHz、量子化 16 bit にて収録した。聴取実験における基本周波数の影響を排除するため、収録の際に話者に 120 Hz の純音をヘッドフォンにより聴かせ、それに基本周波数を合わせて発話するよう指示した。

<sup>1</sup> スペクトルの絶対値の対数を取り、さらに逆フーリエ変換したもの。次元は時間。

<sup>2</sup> ERB (equivalent rectangular bandwidth) rate とは、ヒトの聴覚フィルタの特性を考慮して作成された周波数スケールである。低周波数帯域では分解能が高く、周波数帯域が高くなるほど分解能は低くなる。

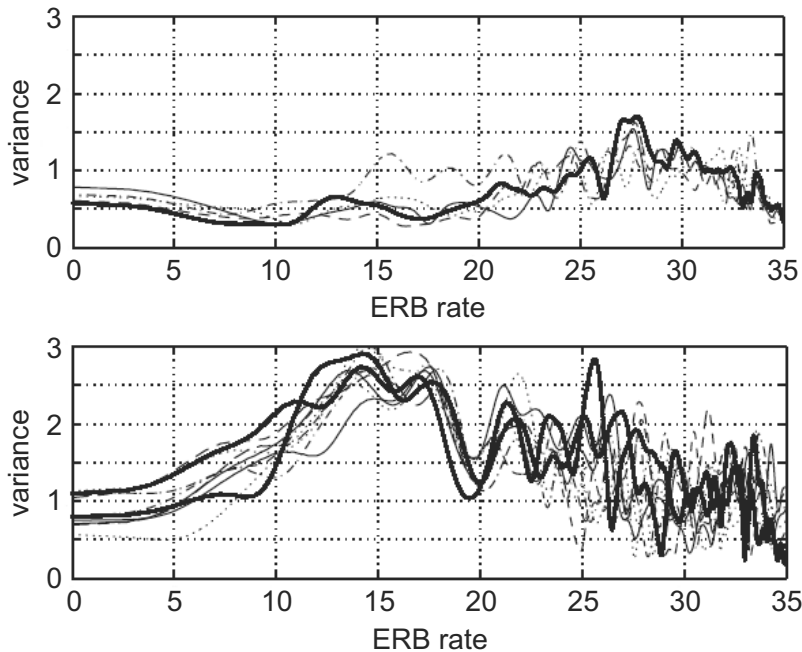


図 1: 男性話者 10 名により発話された単母音のスペクトル包絡の話者間分散 (上段) と話者内分散 (下段)

収録した音声データから LMA (log magnitude approximation) 分析合成系 [5] を用いて以下の 4 種類の刺激音を合成した。刺激音の時間長は約 0.5 s である。

**1a** LMA 分析合成音声

**1b** 12~22 ERB rate の周波数帯域のスペクトル包絡をその自己回帰直線に対して反転させた音声

**1c** 22 ERB rate 以上の周波数帯域のスペクトル包絡をその自己回帰直線に対して反転させた音声

**1d** 22 ERB rate 以上の周波数帯域のスペクトル包絡をその自己回帰直線で置換した音声

刺激音 **1a**, **1c**, **1d** のスペクトル包絡の例を図 2 に示す。

刺激音の合成にあたっては、基本周波数パターンを固定し、さらにフレームをランダム化した。これによって、スペクトル包絡の時間的に静的な成分のみが刺激音に残るようにした。また、LMA フィルタの作成には 60 次のケプストラムを用いた。

### 3.2 被験者

被験者は刺激音の話者の声をよく知っている成人 8 名 (男性 7 名, 女性 1 名) である。いずれも聴覚に問題はない。

### 3.3 実験方法

実験は防音室にて行われた。刺激音はヘッドフォンにより両耳呈示された。音量は各自の聴きやすいレベルとした。各刺激音はランダムに並べられ、実験全体で3回呈示された。被験者には、刺激音の母音と話者を回答し、それらが識別できない場合には“X”を回答するよう依頼した。

### 3.4 実験結果

被験者間で平均した母音識別率と話者識別率を図3に示す。この結果から以下のことが明らかになった ( $F(1, 14) = 4.60, p < 0.05$ )。

1. 刺激音 **1a** と **1c** の母音識別率に有意差がなく ( $F(1, 14) = 4.51$ )、話者識別率には有意差がある ( $F(1, 14) = 88.90$ )。従って、スペクトル包絡の 22 ERB rate 以上の周波数帯域に対する操作は、母音識別に影響を与えないが、話者識別には影響を与える。
2. 刺激音 **1a** と **1b** の母音識別率に有意差があり ( $F(1, 14) = 342.85$ )、刺激音 **1b** と **1c** の母音識別率にも有意差がある ( $F(1, 14) = 11.84$ )。従って、スペクトル包絡の 12~22 ERB rate の周波数帯域に対する操作は母音識別に影響を与え、その影響は 22 ERB rate 以上の周波数帯域に対する操作よりも影響が大きい。
3. 刺激音 **1c** と **1d** の話者識別率に有意差がある ( $F(1, 14) = 14.32$ )。従って、刺激音 **1c** に加えた操作の方が話者識別への影響が大きい。

第1の点は、母音の個人性は 22 ERB rate 以上の周波数帯域に顕著に現れることを示している。第1、第2の点から、この周波数帯域を操作することによって、母音性とは独立に個人性を制御しうることを示された。また、第3の点は、話者識別にはスペクトル包絡のピークと谷が重要であることを示唆している。

## 4 話者認識への応用 [6]

前節の聴取実験によって、スペクトル包絡の高周波数帯域には個人性知覚に重要な要素が含まれることが明らかになった。そこで、話者認識に有効な周波数帯域を調査し、聴取実験の結果と比較することにした。

### 4.1 音声データ

基本周波数が 125 Hz 前後である成人男性 5 名による日本語 5 母音を用いた。話者毎の基本周波数の違いが話者認識に与える影響を抑えるため、前節の音声データと同様に、録音の際に話者に 125 Hz の純音をヘッドフォンから呈示し、それに声の高さを合わせるよう指示した。

録音した音声の定常部約 200 ms を切り出して音声データとした。標本化周波数は 20 kHz、量子化レベルは 16 bit である。音声データは各母音につき 5 個ずつ作成した。

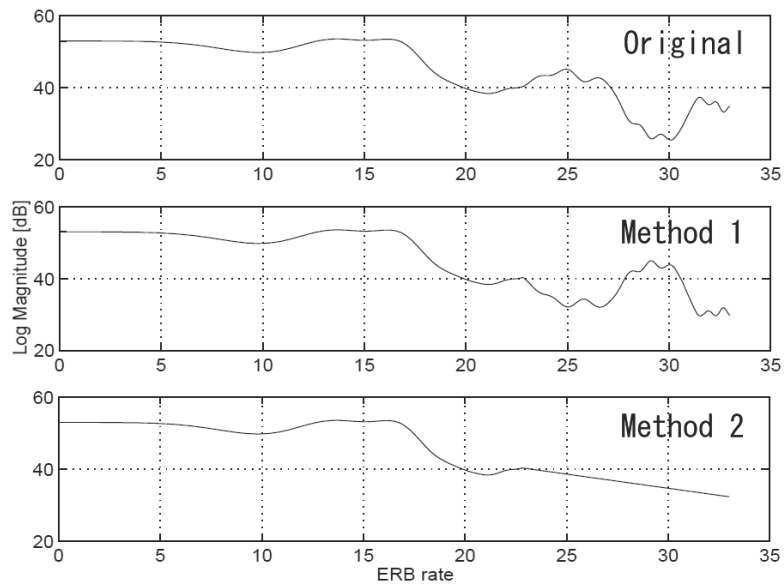


図 2: 刺激音 1a (上段), 刺激音 1c (中段), 刺激音 1d (下段) のスペクトル包絡の例

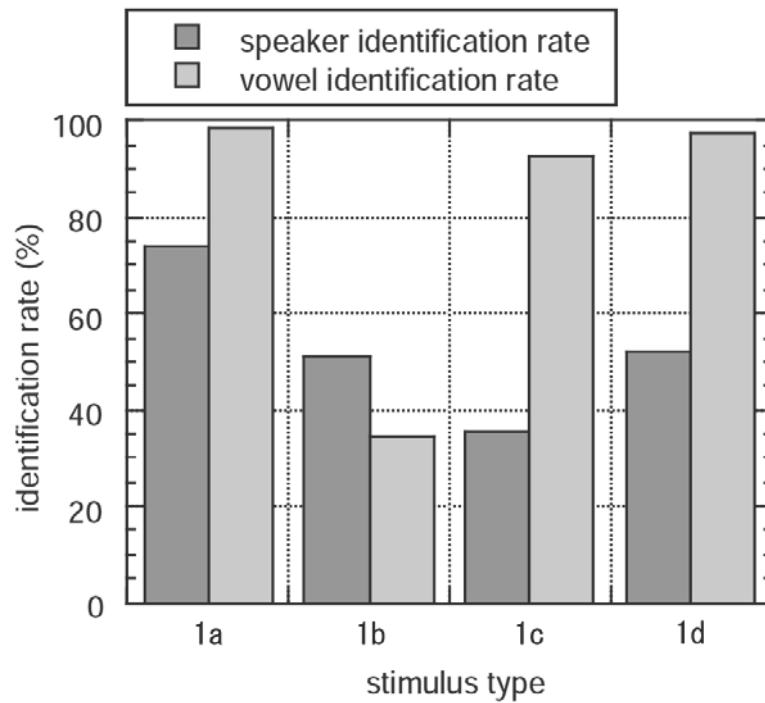


図 3: 被験者間で平均した母音識別率と話者識別率

## 4.2 特徴量

認識の特徴量として、有声区間で加算平均した 60 次のケプストラムから作成したスペクトル包絡を用いた。スペクトル包絡のパワー成分は除去した。分析条件は、フレーム長 25.6 ms, フレーム周期 6.4 ms である。テストパターンは各音声データから求めたスペクトル包絡で、標準パターンは母音ごとにテストパターンをケプストラム上で加算平均して求めたスペクトル包絡である。標準パターンとテストパターンの音韻は同じものを用いた。

## 4.3 識別手法

単純類似度法 [7] を用いた。単純類似度法は標準パターンとの単純類似度が最大となるテストパターンの話者を認識結果とする。標準パターン  $\mathbf{r}$  とテストパターン  $\mathbf{t}$  の間の単純類似度  $S_s$  は以下の式で求められる。

$$S_s = \frac{(\mathbf{r}, \mathbf{t})^2}{\|\mathbf{r}\|^2 \|\mathbf{t}\|^2}, \quad (1)$$

ここで、 $(\mathbf{r}, \mathbf{t})$  は  $\mathbf{r}$  と  $\mathbf{t}$  の内積であり、 $\|\mathbf{r}\|$  は  $\mathbf{r}$  のノルムである ( $= \sqrt{(\mathbf{r}, \mathbf{r})}$ )。

## 4.4 Averaged Distance

話者認識の弁別性能の評価には北村ら [8] により提案された averaged distance (AD) を用いる。単純類似度を距離尺度としたとき、AD は以下の式で求められる。

$$AD = \frac{\sum_i^{N_{sp}} \sum_{j \neq i}^{N_{sp}} \sum_k^{N_{set}} \{S_s(\mathbf{r}_i, \mathbf{t}_{ik}) - S_s(\mathbf{r}_i, \mathbf{t}_{jk})\}}{N_{sp}(N_{sp} - 1)N_{set}}, \quad (2)$$

ここで、 $N_{sp}$  は話者数 (5 名),  $N_{set}$  はテストパターンの数 (5 個),  $\mathbf{r}_i$  は話者  $i$  ( $1, \dots, N_{sp}$ ) の標準パターン,  $\mathbf{t}_{ik}$  は話者  $i$  の  $k$  番目のテストパターンである。

AD は  $N_{sp}$  人の話者における同話者内の単純類似度と異話者間の単純類似度の差の平均を求めている。つまり、AD 値が高いほど弁別性能の良い話者認識法もしくは特徴量であるといえる。また、単純類似度は 2 つのパターンの形状の近さを測るので、AD 値が高い周波数帯域には話者特有の形状が現れていることを意味する。

## 4.5 結果

話者認識に用いるスペクトル包絡の周波数帯域として、0~33 ERB rate (0~8,000 Hz), 0~20 ERB rate (0~1,740 Hz), 20~33 ERB rate (1,740~8,000 Hz) の 3 条件を設定し、AD 値を求めた。第 3 の周波数帯域は同じ音声データを用いた聴取実験により、個人性が顕著に現れるとされた周波数帯域である [9]。

得られた AD 値を表 1 に示す。この結果から、スペクトル包絡の 20~33 ERB rate を用いた話者認識の弁別性能が最も高いことがわかる。つまり、スペクトル包絡のこの周波数帯域は話者特有の形状を有していることを意味している。また、聴取実験で個人性が顕著に現れるとされた周波数帯域が話者認識に有効であることも意味している。

表 1: 3 つの周波数帯域における averaged distance (AD) 値

frequency region (ERB rate)	vowel				
	/a/	/e/	/i/	/o/	/u/
0 – 33	0.077	0.103	0.110	0.046	0.068
0 – 20	0.033	0.045	0.035	0.017	0.031
20 – 33	0.510	0.351	0.639	0.514	0.579

## 5 母音の個人性の生成要因

前節までは、音声の個人性知覚や話者認識に寄与する成分に関して説明した。では、その個人性はヒトの音声生成系のどこから生み出されるのであろうか。本節では、磁気共鳴画像法 (MRI) を用いた生体観測にもとづいて、個人性の生成要因に関して行った研究について概説する。

### 5.1 下咽頭腔の形状の個人差 [10]

音声の個人性を表す特徴量は、(1) 音韻に共通し、かつ (2) 話者ごとに異なると考えられる。もし、この 2 条件を満たす特徴量の生成要因が声道に存在するならば、(1) 発話する音韻が変わっても形状変化が小さく、かつ (2) そのサイズや形状に個人差がある部位であると考えられる。この考えにもとづいて、MRI を用いて母音発話時の声道形状を分析した。

声道形状の MRI 動画による観測 [11] から、連続母音発話時に下咽頭腔の変動が小さいことが明らかになってきたため、特にこの部位に着目し検討した。下咽頭腔は声道下部に位置し、喉頭腔と両側の梨状窩から成る (図 4)。梨状窩は声道の分岐管であり、反共鳴によって音声の 4~5 kHz に零点を生じさせる。そして、その周波数は発話中もほとんど変化しないことから、個人性への寄与が示唆されていた [16]。

#### 5.1.1 MRI データの分析

成人男性 4 名の日本語 5 母音発話時の正中矢状断面を分析したところ、母音の違いによって声門の位置が上下するものの、喉頭腔の形状変化は小さいことが明らかになった<sup>3</sup>。そこで、この部位に関して横断面上の声道断面形状を調査した。男性 1 名についての結果を図 5 に示す。声門から 30 mm 前

<sup>3</sup>この観測には、歯列補填 [12]、発話同期撮像 [13]、喉頭コイル [14]、骨伝導によるガイド音提示 [15] などの MRI 撮像に関する要素技術が極めて重要な役割を果たした。

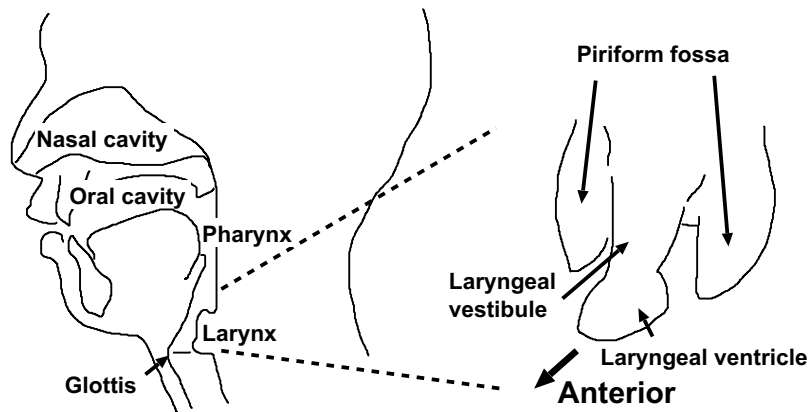


図 4: 声道下部の構造

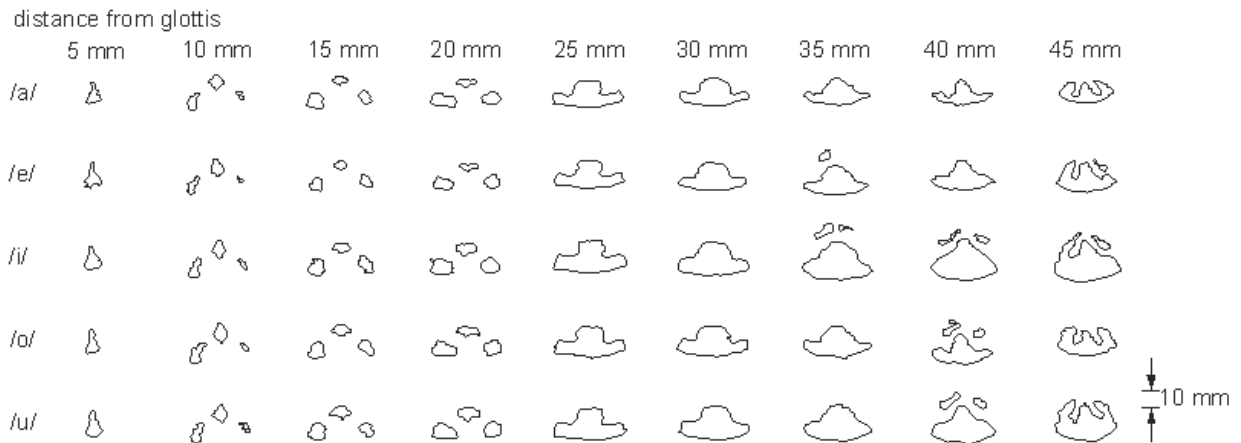


図 5: 成人男性話者の日本語 5 母音発話時における横断面上の声道断面形状。声門から頭頂方向 45 mm の面まで 5 mm ごとに横断面を抽出

後の部位（下咽頭腔）の声道断面形状が被験者内ではよく似ていることが見てとれる。その一方で、話者間の比較により、この部分の形状は被験者間では異なっていることが明らかになった。

### 5.1.2 電気回路モデルによるシミュレーション

声道内で音が平面波伝搬する周波数帯域では、同心円を連結させた 1 次元の音響管として声道を近似することができる（図 6）。各同心円の半径はその位置に対応する声道の断面積から決まる。なお、声道中心線に直交する断面の面積を 1 次元にならべたものを声道断面積関数という。この 1 次元の音響管を電気回路で近似することによって、声道伝達特性を計算することができる。声道伝達特性は声道の共鳴、反共鳴などに関する情報が得られる重要なものである。

この電気回路モデルを用いて、下咽頭腔の形状の個人差が音声の個人差に与える影響を調べた。下



咽頭腔のみが話者ごとに異なり、それより上の部分は話者間で共通にした声道断面積関数を作成し、声道伝達特性を求めた。母音/e/に関する結果を図7に示す。この結果から、下咽頭腔の形状の個人差は音声スペクトルの約 2.5 kHz 以上の周波数帯域に影響を与えることがわかる。

このシミュレーションから、下咽頭腔の形状の個人差は、音声スペクトルの約 2.5 kHz 以上の周波数帯域に影響を与えることが明らかになった。この周波数帯域は、Furui and Akagi [18] や著者らによって個人性知覚に寄与するとされた周波数帯域とほぼ一致する。従って、下咽頭腔は音声の個人性の生成要因の1つであると結論できる。この研究は、聴取実験により得られた結果を音声生成系の研究によって裏付けたものといえる。

## 5.2 新しい音声生成モデル

発話状態のMRI観測技術の向上によって、MRIデータの空間分解能やS/Nが向上した。その結果得られた声道形状にもとづいて検討を重ねるうちに、喉頭腔が声道全体に対してほぼ独立な音響管とみなせること、喉頭腔の共鳴（喉頭腔共鳴）によって男性の場合3~4 kHz付近にホルマント<sup>4</sup>（多くの場合第4ホルマント）が生じることなどが明らかになった [19]-[21]。

Hondaら [22] はこの知見にもとづいて新しい音声生成モデルを提案した。このモデルでは、声道は下咽頭腔（喉頭腔と梨状窩）と主声道に分離され、独立なフィルタとして働く。すなわち、声道全体のフィルタ特性は、下咽頭腔フィルタと主声道フィルタの縦続接続によって表される。発話時に、下咽頭腔フィルタの特性は音韻によってほとんど変化しないが（第5.1.1節参照）、主声道フィルタの特性は大きく変化する。そして、この縦続接続されたフィルタに声帯音源波が入力されることによって、音声が生産されると考える。Chiba and Kajiyama [23] 以来、声道は1つの音響管として全てのホルマントを生成するとされてきたので、大きな進歩である。

このモデルは個人性生成にも適用できる。つまり、下咽頭腔の形状の個人差が高周波数領域（喉頭腔共鳴によるピークと梨状窩の反共鳴による零）の個人差を生み出し、主声道の長さや形状の個人差がそれぞれ音声スペクトルの全体的なシフトと低次ホルマントの個人差を生み出すと考えることができる。

## 6 物真似音声の分析 [24]

何人もの声色を本人そっくりに真似る物真似タレントは、個人性研究の観点から大変興味深い対象である。彼らが音声中のどの特徴量をいかにして対象話者に近づけているのかが明らかになれば、音声の個人性の生成と知覚に関する重要な知見が得られるだろう。

また、我々が「同じ」と感じるものと「似ている」と感じることとの関係も興味深い。「似ている」程度がある閾値を越えると「同じ」と感じるのだろうか。物真似に関する研究は、この疑問についても示唆を与えるものと期待できる。

物真似音声の研究には権利等の問題でデータ入手が困難という難点があるが、幸いにしてプロの物真似タレントの音声进行分析する機会を得たので、その結果を紹介する。

<sup>4</sup>音声スペクトルにおいてパワーの集中した部分。音韻や個人性の知覚にとって重要。

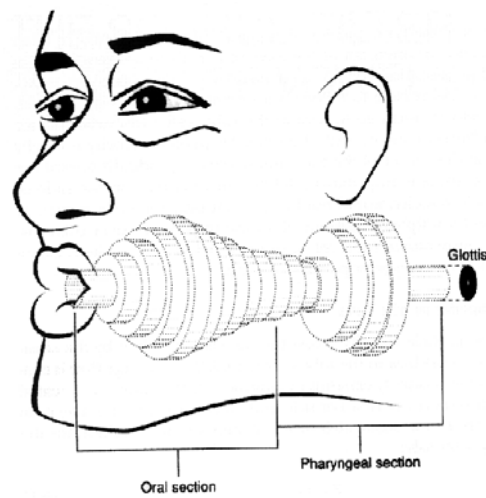


図 6: 声道を 1 次元の音響管として近似. 文献 [17] より転載

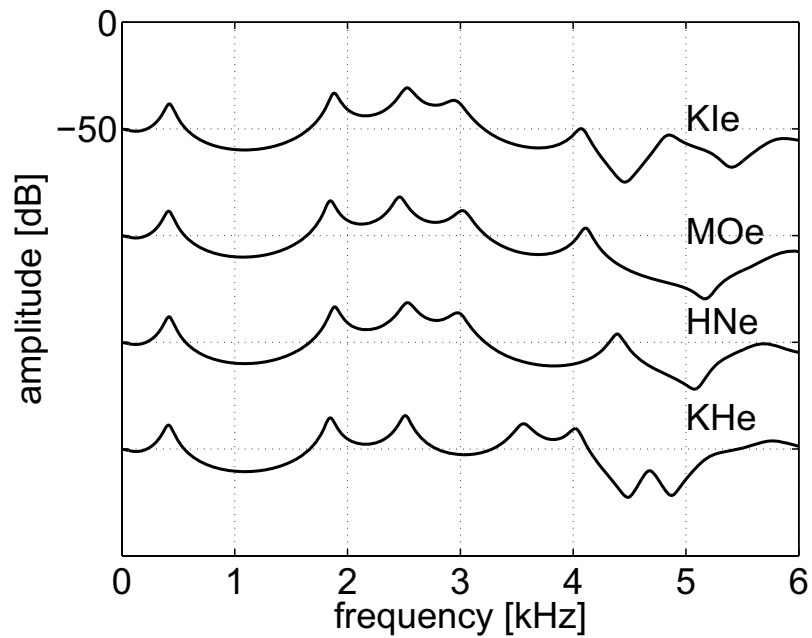


図 7: 下咽頭腔のみが話者ごとに異なる母音/e/の声道断面積関数から求めた声道伝達特性. 上から被験者 KI, MO, HN, KH の声道伝達特性

## 6.1 音声データ

音声データは、物真似の対象話者 A と物真似タレントの話者 B が次の 2 つの文章（都々逸）を発話したものである。なお、括弧内の文字は話者 A の発話の仕方を記載したものである。

**文章 1** 一度でいいから見てみたい、女房（にようぼ）がへソクリ隠すところ。

**文章 2** 出かける猫に行（ゆ）き先聞けば、旅行が好きでまた旅だ。

音声データは、標準化周波数 16 kHz、量子化 16 bit にて収録した。話者 B の音声は物真似発話と地声発話したものを収録した。

## 6.2 分析結果と考察

今回対象にした物真似音声では以下のような傾向が見られた。

1. 物真似音声の平均基本周波数は話者 A のものより高い。
2. 基本周波数の変化パターンが類似している区間が多い。
3. 第 1, 第 3, 第 4 ホルマント周波数が近い。
4. スペクトル包絡間の距離が小さい。
5. 声帯音源特性が近い。

以下、いくつかの点について説明する。

### 6.2.1 基本周波数パターン

文章 1 の基本周波数パターンを図 8 に示す。ただし、矢印と数字は傾きを表す。話者 A の音声と話者 B の物真似音声を比較すると、変化のタイミングはずれているものの、傾きの値が近い部分が多いことがわかる。Akagi and Ienaga [25] は、3 モーラの単語を用いた実験によって基本周波数の平均値のみならず動的成分も個人性知覚に寄与することを示し、後者の影響がより強いことを報告している。物真似発話の際に基本周波数の変動パターンを似せることは、彼らの結果から見ても理にかなっているといえる。

### 6.2.2 スペクトル

文章 1 中「ら」の母音区間のスペクトルをフレーム方向に加算平均した結果を図 9 に示す。話者 A の音声と話者 B の物真似音声のスペクトルの概形がよく似ており、一方で、話者 B の地声音声と物真似音声のスペクトルの概形は異なっている。この結果から、話者 B は物真似発話の際に声道を変形させていることがわかる。

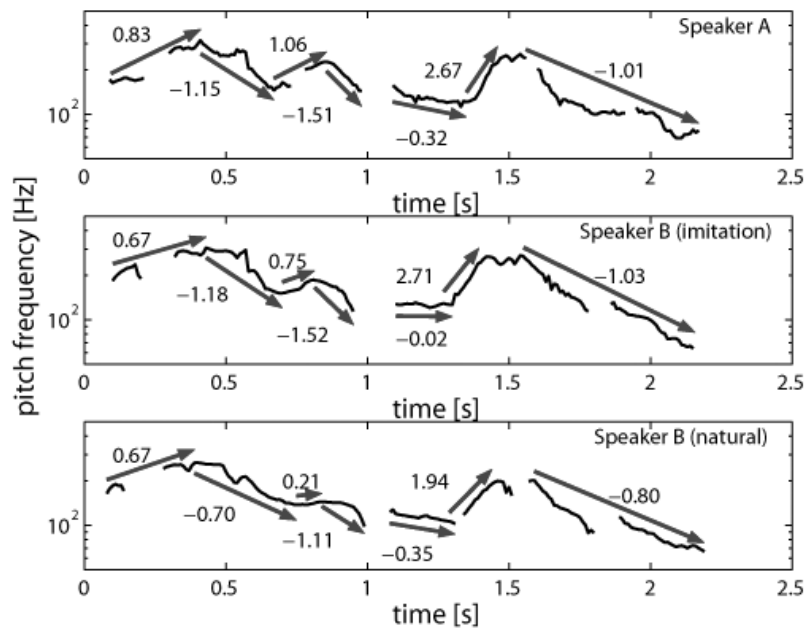


図 8: 文章 1 の前半部 (一度でいいから見てみたい) の基本周波数パターン. 話者 A (上段), 話者 B の物真似音声 (中段), 話者 B の地声 (下段) の基本周波数パターン

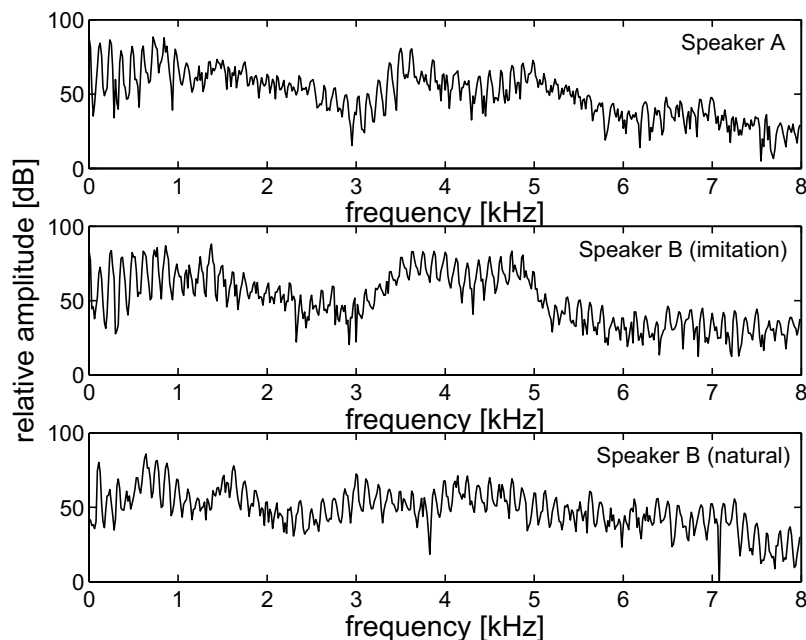


図 9: 文章 1 中の「ら」の母音区間の平均スペクトル. 話者 A (上段), 話者 B の物真似音声 (中段), 話者 B の地声 (下段) の平均スペクトル

話者 A の音声と話者 B の物真似音声のスペクトルでは、ピークだけでなく、3 kHz 前後の谷の周波数と形状も一致している。地声音声にはこの谷が見られないので、話者 B は物真似の際に意図的にこの谷を生み出していることがわかる。スペクトルの谷は聴感上の影響こそ小さいが、その周辺のスペクトルに影響を与えるため、ホルマント周波数やスペクトルの概形の類似性に寄与していると考えられる。

## 7 おわりに

本論文では、音声の個人性の生成と知覚に関する著者らによる研究を概説した。音声生成の研究と音声知覚の研究は密接に関連している。どちらの研究でも、ある知見が得られた背景にはその基礎となった知見や要素技術がある。このような階層構造が多層につながることによって、研究は少しずつ進展している。本論文によって音声の個人性や音声生成の研究に少しでも興味を持っていただけたら望外の喜びである。

最後に今後の課題を挙げておく。

1. **鼻腔が音声の個人性に及ぼす影響** 発話する音韻が変わっても形状変化が小さく、かつそのサイズや形状に個人差がある部位としては、下咽頭腔の他に鼻腔が挙げられる。現在の MRI の空間分解能では鼻腔の微細構造まで観測できないが<sup>5</sup>、何らかの方法で音声の個人性への鼻腔の寄与について調べる必要がある。
2. **声道の微細構造が音声の個人性に及ぼす影響** 声道形状を段階的にスムージングし、その伝達特性を求めることによって、声道の微細構造が音声の個人性に及ぼす影響を調査する必要がある。声道形状をスムージングしても個人性が保たれるのであれば、声道形状を少ないパラメータで表現できる。
3. **音声の個人性の学習過程** 初対面の人の声がどのように学習されていくのかに興味がある。第 1 印象を得る際に使われる特徴量は何か？その人の声と接する時間が増えることで、使われる特徴量が増えるのか、それとも分解能が上がっていくのか？また、人による学習過程の違いも興味深い。
4. **似た声を持つ話者の音声の分析** 個人間の差異を研究する方法では、個人性知覚に寄与する特徴量のリストがいつ終わるのかわからない。似た声を持つ話者の音声を比較し、値の近い特徴量を明らかにできれば、個人性知覚に寄与する特徴量必要十分なリストが得られるはずである。
5. **物真似音声の生成メカニズム** 前節で紹介した物真似音声の研究では、残念ながら物真似発話中の MRI データを得ることができなかった。このデータを得ることができれば、個人性の生成メカニズムの解明に役立つはずである。
6. **連続音声における個人性の生成と知覚** 発話同期撮像やリアルタイム MRI などの技術によって、発話時の運動が観測できるようになったので、これにもとづいて時間的に動的な個人性の生成要因および知覚要因を調査する必要がある。

<sup>5</sup>X 線 CT を用いれば鼻腔の微細構造まで観測可能であるが、被曝の問題が生じる。

## 謝辞

本研究の一部は、NICTからの研究委託「人間情報コミュニケーションの研究開発」により実施され、平成16–18年度科学研究費補助金(60293594)、平成16–18年度総務省SCOPE-R(042107002)、平成20年度総務省SCOPE(071705001)、文部科学省ORC整備事業(2004–2008)の援助を受けた。これまでご教授、ご議論いただいた皆様、実験にご協力いただいた皆様に感謝します。なお、本論文は文献[26]にもとづいている。

## 参考文献

- [1] H. Fujisaki, “Prosody, models, and spontaneous speech,” in Sagisaka *et al.* eds., *Computing Prosody*. Springer, pp. 27–42, 1997.
- [2] T. Kitamura and M. Akagi, “Speaker individualities in speech spectral envelopes,” *J. Acoust. Soc. Jpn. (E)*, vol. 16, no. 5, pp. 283–289, 1995.
- [3] K.-P. Li and G.W. Hughes, “Talker differences as they appear in correlation matrices of continuous speech spectra,” *J. Acoust. Soc. Am.*, vol. 55, pp. 833–873, 1974.
- [4] P. Mokhtari and F. Clermont, “Contributions of selected spectral regions to vowel classification accuracy,” in *Proc. ICSLP’94*, pp. 1923–1926, 1994.
- [5] 今井, “対数振幅近似(LMA)フィルタ,” 信学論, vol. J63-A, no. 12, pp. 886–893, 1980.
- [6] 北村, 赤木, “単純類似度法による話者識別に適した周波数帯域の検討,” 日本音響学会研究発表会講演論文集, no. 2, pp. 341–342, 1996.
- [7] 飯島, パターン認識理論. 森北出版, 1989.
- [8] 北村, 岩城, 飯島, “不特定話者母音の認識における類似度の多重化法,” 信学技報(SP), vol. 95, no. 565, pp. 47–54, 1996.
- [9] 北村, 赤木, “単母音の話者識別に寄与するスペクトル包絡成分,” 音響誌, vol. 53, no. 3, pp. 185–191, 1997.
- [10] T. Kitamura, K. Honda, and H. Takemoto, “Individual variation of the hypopharyngeal cavities and its acoustic effects,” *Acoust. Sci. & Tech.*, vol. 26, no. 1, pp. 16–26, 2005.
- [11] H. Takemoto, K. Honda, S. Masaki, Y. Shimada, and I. Fujimoto, “Measurement of temporal changes in vocal tract area function during a continuous vowel sequence using a 3D cine-MRI technique,” in *Proc. the 6th International Seminar on Speech Production, Sydney*, pp. 284–289, 2003.
- [12] H. Takemoto, T. Kitamura, H. Nishimoto, and K. Honda, “A method of tooth superimposition of MRI data for accurate measurement of vocal tract shape and dimensions,” *Acoust. Sci. and Tech.*, vol. 25, no. 6, pp. 468–474, 2003.

- [13] S. Masaki, M. Tiede, and K. Honda, "MRI-based speech production study using a synchronized sampling method," *J. Acoust. Soc. Jpn. (E)*, vol. 20, no. 5, pp. 375–379, 1999.
- [14] 高野, 本多, "高分解能 MRI による輪状甲状関節の観測," 音声言語医学, vol. 46, no. 3, pp. 174–178, 2005.
- [15] Y. Nota, T. Kitamura, H. Takemoto, H. Hirata, K. Honda, Y. Shimada, I. Fujimoto, Y. Syakudo, and S. Masaki, "A bone-conduction system for auditory stimulation in MRI," *Acoust. Sci. & Tech.*, vol. 28, no. 1, pp. 33–38, 2007.
- [16] J. Dang and K. Honda, "Acoustic characteristics of the piriform fossa in models and humans," *J. Acoust. Soc. Am.*, vol. 101, pp. 456–465, 1997.
- [17] I.R. Titze, *Principles of voice production*. NJ: Prentice-Hall, 1994.
- [18] S. Furui and M. Akagi, "Perception of voice individuality and physical correlates," *Trans. Tech. Com. Psycho. Physio. Acoust.*, vol. H85-18, 1985.
- [19] 本多, 北村, 竹本, 能田, 平田, 正木, 藤田, "母音発話 MRI データに基づく声道模型の製作と音響分析," 日本音響学会研究発表会講演論文集, no. 2, pp. 313–314, 2005.
- [20] H. Takemoto, S. Adachi, T. Kitamura, P. Mokhtari, and K. Honda, "Acoustic roles of the laryngeal cavity in vocal tract resonance," *J. Acoust. Soc. Am.*, vol. 120, pp. 2228–2238, 2006.
- [21] T. Kitamura, H. Takemoto, S. Adachi, P. Mokhtari, and K. Honda, "Cyclicality of laryngeal cavity resonance due to vocal fold vibration," *J. Acoust. Soc. Am.*, vol. 120, pp. 2239–2249, 2006.
- [22] K. Honda, H. Takemoto, T. Kitamura, S. Fujita, and S. Takano, "Exploring human speech production mechanisms by MRI," *IEICE Trans. Inf. and Syst.*, vol. E87-D, no. 5, pp. 1050–1058, 2004.
- [23] T. Chiba and M. Kajiyama, *The Vowel*. Tokyo: Tokyo-Kaiseikan, 1942. (邦訳: 杉藤, 本多訳, 母音: その性質と構造, 岩波書店)
- [24] 北村, "物真似タレントによる物真似音声の分析," 信学技報 (SP), vol. 107, no. 282, pp. 49–54, 2007.
- [25] M. Akagi and T. Ienaga, "Speaker individuality in fundamental frequency contours and its control," *J. Acoust. Soc. Jpn. (E)*, vol. 18, no. 2, pp. 73–80, 1997.
- [26] 北村, "音声の個人性の生成と知覚," 日本音響学会聴覚研究会資料, vol. 38, no. 6, pp. 653–658, 2008.