

物真似タレントによる物真似音声の分析

北村 達也[†]

[†] 甲南大学 理工学部 情報システム工学科
658-8501 兵庫県神戸市東灘区岡本 8-9-1
E-mail: †t-kitamu@konan-u.ac.jp

あらまし 物真似音声に似ていると知覚される要因を調べるため、プロの物真似タレントによる物真似音声と対象話者の音声を比較した。2つの文章音声を分析した結果、基本周波数の平均値は対象話者の音声より 20 Hz 前後高く、その変動パターンはこれらの話者の間でよく似ていた。スペクトルに関しても、その概形や第1, 第3, 第4ホルムント周波数が対象話者のものに近付けられていた。さらに、声帯音源特性を表す、DFT スペクトルの第1, 第2調波の振幅差 (H1-H2) もこれらの話者間で近い値を示した。一方、第2ホルムント周波数は話者間で異なっており、音節継続時間長は対象話者のものに近付ける傾向が見られなかった。

キーワード 物真似音声, 基本周波数, 音声スペクトル, 声道音響特性, 声帯音源特性

Analysis of imitated voice produced by a professional impersonator

Tatsuya KITAMURA[†]

[†] Department of Information Science and Systems Engineering, Faculty of Science and Engineering, Konan University
8-9-1 Okamoto, Higashinada, Kobe, Hyogo, 658-8501, Japan
E-mail: †t-kitamu@konan-u.ac.jp

Abstract This study is a comparative survey of voice produced by a professional impersonator imitating a target speaker in order to explore possible perceptual factors of similarity of speaker characteristics. The results show that the mean pitch frequency (F0) of the impersonator is approximately 20 Hz higher than the target speaker and the dynamics of the F0 contour of the two speakers is closely resemble. The DFT spectra of the speakers are quite similar in its shape and the first, third, and fourth formant frequencies. Moreover, the difference between the amplitude levels of the first harmonic (H1) and the second harmonic (H2), a measure of the glottal source characteristics, are close between the speakers. In contrast, the second formant frequency and syllable duration of the imitated voice differ from the target voice.

Key words imitated voice, pitch frequency, speech spectrum, vocal tract acoustic characteristics, glottal source characteristics

1. はじめに

一人で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、音声科学の観点から見ても大変興味深い。彼らは音声の中の音響特徴量を対象話者のものに近付けているのだろうか。また、物真似発話時の音声生成系の制御はどのようなものだろうか。そして、私たちが物真似音声を似ていると知覚する要因は何だろうか。これらのことが明らかになれば、個人性知覚や声質制御に関する有益な知見が得られることだろう。

さらに、我々が「同じ」と感じることと「似ている」と感じるものの関係も興味深い。音声から話者を同定するときの処理

過程と、その音声がある人の声と似ていると感じるときの処理過程とは同じなのだろうか。「似ている」程度が無限大になるか、もしくはある閾値を越えたところに「同じ」があるのだろうか。物真似に関する研究は、これらの問題についても示唆を与えるものと期待できる。

このように、物真似音声には科学的興味をそそる点が多いが、データ入手が困難という難点がある。また、物真似というものは本質的に対象話者に依存するため、体系的に研究しにくい面もある。しかしながら、これまでいくつかの報告がなされている。鈴木 [1] は、桜井長一郎氏が宇野重吉氏と柳家金語楼氏の声を物真似 (声帯模写) した音声を分析した。そして、聞いた

限りではよく似ている物真似音声であっても、種々の音響特徴量が対象話者と異なることを示している。しかし、当時(1968年)の技術的制約もあって、物真似音声に似ていると知覚される要因を定量的に明らかにできていない。

また、Zetterholm はプロの物真似タレントによる物真似音声の基本周波数の分析 [3]、基本周波数、継続時間長、ホルマント周波数の分析と音素ラベリング [4] による評価を行っている。Zetterholm はプロの物真似タレント 2 名が 1 人の対象話者の物真似をした音声の分析も行っている [2]。さらに、Clermont と Zetterholm [5] は、方言を真似て発話された単語音声に関してホルマント周波数の変化パターンを報告している。

本研究では、物真似音声に似ていると知覚される要因を解明するために、プロの物真似タレントによる物真似音声进行分析した。物真似の対象話者の音声と物真似タレントによる物真似音声を対象にして、音節継続時間長、基本周波数、DFT スペクトル、ホルマント周波数、声帯音源特性を比較することによって、音声の中のどのような成分を似せているのかを定量的に分析した。さらに、物真似タレントによる物真似音声と地声で発話した音声とを比較し、物真似の際に音声生成系をいかに制御しているのかを推察した。

2. 音声データ

音声データは、静かな部屋で物真似の対象話者 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。話者 A は落語家である。なお、括弧内の文字は話者 A の発音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)がヘソクリ隠すところ。

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。

本研究では、「一度でいいから見てみたい」を文章 1 の前半部、「女房がヘソクリ隠すところ」を文章 1 の後半部、「出かける猫に行き先聞けば」を文章 2 の前半部、「旅行が好きでまた旅だ」を文章 2 に後半部と表す。

話者 B の音声は、物真似したものと地声で発話したものを収録した。物真似音声の収録の際、話者 B は話者 A の発話を聞いた直後に発話した。すなわち、手本とすべき音声を聞いた直後に、それを真似て発話した。話者 A と B が 2 回発話した中からノイズや飽和がない、または少ない音声データを 1 つずつ選択した。選択した音声データのうち、話者 B が発話した文章 2 の「が」の部分に飽和がある。また、話者 B は文章 2 の「行(ゆ)き先」の部分で「行(ゆ)く先」と発音している。なお、これらのデータには収録場所の影響で暗騒音や残響が含まれている。

一方、話者 B の地声音声は、物真似音声の収録とは別の日に話者 A のいない場所で 1 回発話したものを収録した。この音声データには文章 1 の「ぼ」の部分に飽和がある。文章 2 の「行(ゆ)き先」の部分は「行(ゆ)き先」と発音されている。なお、このデータには暗騒音や残響がほとんど含まれていない。

上記いずれの音声データも標本化周波数 16 kHz、量子化 16

bit にて保存した。

3. 分析方法

上記の音声データの音節継続時間長、基本周波数、DFT スペクトル、ホルマント周波数を求め比較した。DFT スペクトルの比較には後述するスペクトル包絡間距離を用いた。さらに、声帯音源特性についても比較した。音声データが暗騒音や残響の影響を受けていたため、パワーは分析対象から除外した。

音節継続時間長は、まず音声処理ソフトウェア WaveSurfer [6] 上でスペクトログラムを参考にしつつ音節区間を決定し、次にそれぞれの区間長を求めた。基本周波数は、WaveSurfer の基本周波数抽出機能を用いて抽出した。基本周波数抽出の際のパラメータは、WaveSurfer の既定値を用いた。すなわち、抽出手法は ESPSP method [8]、フレーム長は 7.5 ms、フレーム周期は 10 ms を用いた。基本周波数の抽出後、明らかな抽出誤りを目視で修正した。DFT スペクトルはフレーム長 64 ms、フレーム周期 8 ms にて Matlab [7] 上で求めた。窓関数はハニング窓を用いた。ホルマント周波数は、ケプストラム次数 40 で求めた DFT スペクトル包絡をフレーム方向に加算平均した平均 DFT スペクトル包絡からピークピッキングにより求めた。

さらに、スペクトルの概形の類似性を評価するため、2 つの DFT スペクトル包絡間の距離を求めた。まず、フレーム長 64 ms、フレーム周期 8 ms、ケプストラム次数 40 という条件で、分析対象区間の平均 DFT スペクトル包絡を求めた。窓関数はハニング窓を用いた。なお、ケプストラムの 0 次項(直流成分)は 0 にした。2 つの平均 DFT スペクトル包絡を $E_1(f)$ 、 $E_2(f)$ とすると、周波数 $f = f_{N1}, \dots, f_{N2}$ におけるスペクトル包絡間距離 D を次の式で求めた。

$$D = \frac{1}{N2 - N1 + 1} \sum_{f=f_{N1}}^{f_{N2}} |E_1(f) - E_2(f)| \quad (1)$$

D の値が大きいくほど 2 つの平均 DFT スペクトル包絡の差が大きいくことを意味する。

さらに、音声データの声帯音源特性を比較するために、DFT スペクトルの第 1、第 2 調波の振幅差 ($H1-H2$) を求めた [9]。第 1、第 2 調波の周波数はピークピッキングにより求めた。

4. 結果

4.1 音節継続時間長

文章 1 および 2 の各音節の継続時間長をそれぞれ図 1、2 に示す。これらの音節継続時間長について音声データ間の相関係数 r を求めた。その結果、文章 1 では、話者 A の音声と話者 B の物真似音声との間に 0.758、話者 A の音声と話者 B の地声音声との間に 0.847、話者 B の物真似音声と地声音声との間に 0.836 の相関があった。一方、文章 2 では、話者 A の音声と話者 B の物真似音声との間に 0.629、話者 A の音声と話者 B の地声音声との間に 0.755、話者 B の物真似音声と地声音声との間に 0.799 の相関があった。いずれの文章でも、話者 A の音声に対する相関は地声音声の方が高い。すなわち、この話者 B が話者 A の物真似をする場合、音節継続時間長を話者 A のもの

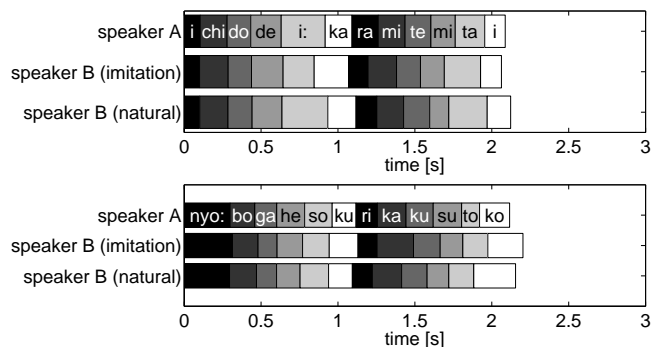


図1 文章1の音節継続時間長．上段：文章1前半部（一度でいいから見てみたい），下段：文章1後半部（女房がへソクリ隠すところ）．各図の上から話者A，話者Bの物真似音声，話者Bの地声音声．

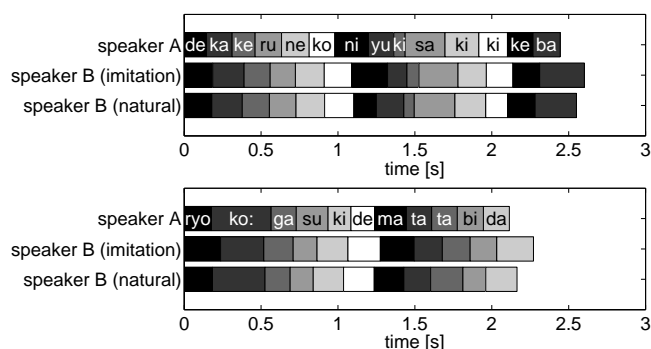


図2 文章2の音節継続時間長．上段：文章2前半部（出かける猫に行き先聞けば），下段：文章2後半部（旅行が好きでまた旅だ）．各図の上から話者A，話者Bの物真似音声，話者Bの地声音声．

に近付ける制御をしていないことがわかる．

また、いわゆる「間(ま)」に対応する、前半部と後半部の間の無音区間長を比較した．文章1では、話者Aで0.342 s、話者Bの物真似音声で0.528 s、話者Bの地声音声で0.620 sであった．一方、文章2では、話者Aで0.388 s、話者Bの物真似音声で0.299 s、話者Bの地声音声で0.515 sであった．物真似音声の無音区間長は地声音声のものに対して話者Bのものに近づく傾向を示してはいるものの、これらの無音区間長の値が極めて近いとは言えない．データ数の不足もあり、話者Bが発話の「間」を模倣しているかについては判断できない．

4.2 基本周波数

文章1の平均基本周波数は、話者Aが167.2 Hz、話者Bの物真似音声で185.1 Hz、地声音声で152.0 Hzであった．また、文章2の平均基本周波数は、話者Aが162.2 Hz、話者Bの物真似音声で187.3 Hz、地声音声で154.8 Hzであった．文章1と文章2の平均基本周波数の間に大きな差異は見られない．話者Bの物真似音声の平均基本周波数は話者Aのものに対して文章1で17.9 Hz、文章2で25.1 Hz高かった．話者Bは基本周波数を高く誇張することによって、平均的な成人男性よりも高い基本周波数を持つ話者Aの声に似て聞こえるようにしていると推察される．

次に、文章1前半部、文章1後半部、文章2前半部、文章2後半部の基本周波数をそれぞれ図3、4、5、6に示す．話者A

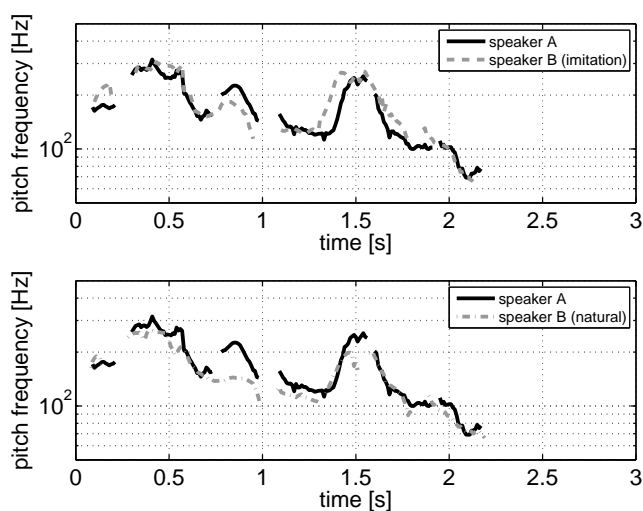


図3 文章1前半部（一度でいいから見てみたい）の基本周波数．上段：話者Aの音声と話者Bの物真似音声，下段：話者Aの音声と話者Bの地声音声．

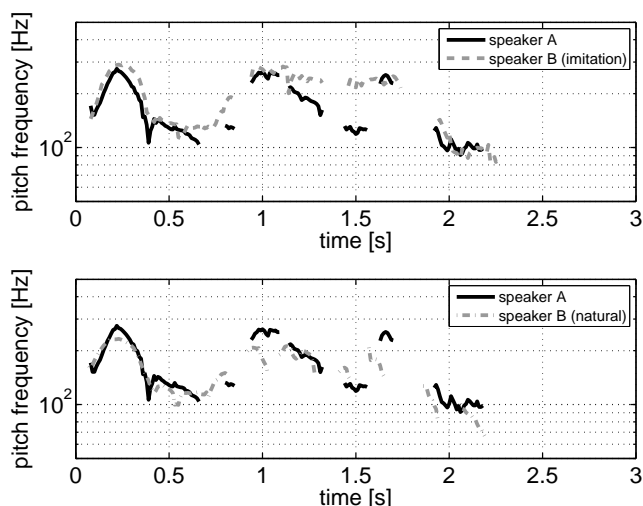


図4 文章1後半部（女房がへソクリ隠すところ）の基本周波数．上段：話者Aの音声と話者Bの物真似音声，下段：話者Aの音声と話者Bの地声音声．

の基本周波数と話者Bの物真似音声の基本周波数を比較すると、変化のタイミングはずれているものの、形状がよく似ている部分があることがわかる．具体的には、文章1前半部全体、文章1後半部の約1 sまでの区間と約1.7 s以降の区間、文章2前半部の約0.5 sまでの区間で類似性が高い．

ここでは、基本周波数の変化パターンの類似性を客観的に評価するため、文章1前半部を対象にして対数軸上での基本周波数の勾配を話者間で比較する．まず、「いちど」における文頭から基本周波数が極大値をとる点までの勾配は、話者Aが0.83、話者Bの物真似音声で0.67、話者Bの地声音声で0.67であった．続く「で」における勾配は、話者Aが-1.15、話者Bの物真似音声で-1.18、話者Bの地声音声で-0.79であった．ここで負値は基本周波数が下降していることを意味している．基本周波数が上昇の後ただちに下降する「いい」における勾配は、話者Aの上昇部が1.06、下降部が-1.51、話者Bの物真似音声の

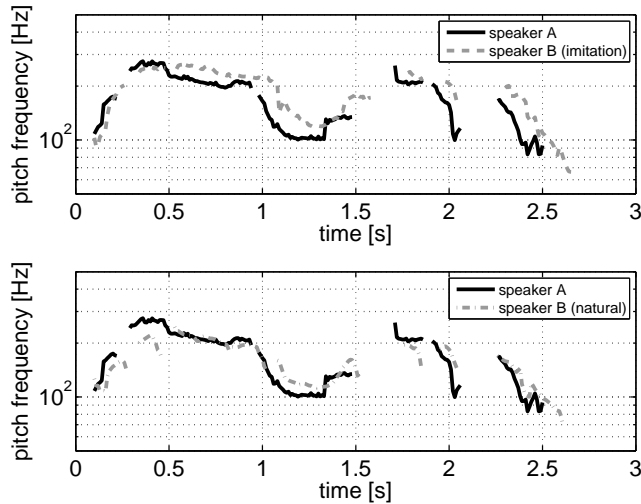


図5 文章2前半部(出かける猫に行き先聞けば)の基本周波数. 上段: 話者Aの音声と話者Bの物真似音声, 下段: 話者Aの音声と話者Bの地声音声.

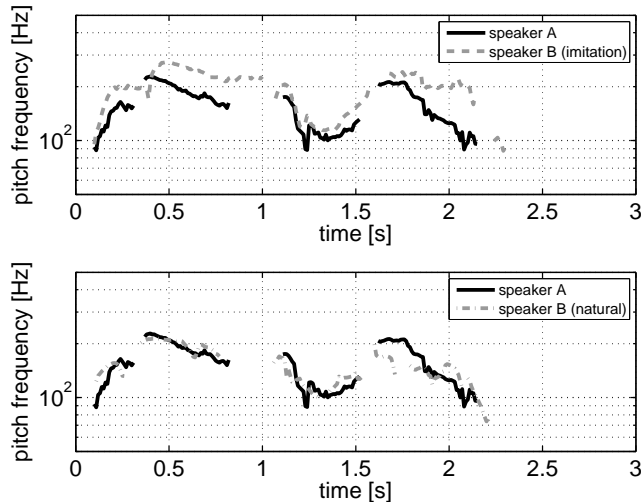


図6 文章2後半部(旅行が好きでまた旅だ)の基本周波数. 上段: 話者Aの音声と話者Bの物真似音声, 下段: 話者Aの音声と話者Bの地声音声.

上昇部が0.75, 下降部が-1.52, 話者Bの地声音声の上昇部が0.21, 下降部が-1.11であった. 基本周波数が緩やかに下降する「から」の有声区間における勾配は, 話者Aが-0.32, 話者Bの物真似音声は-0.02, 話者Bの地声音声は-0.35であった. 基本周波数が急激に上昇する「み」における勾配は, 話者Aが2.67, 話者Bの物真似音声は2.71, 話者Bの地声音声は1.94であった. 勾配を計算した区間は, 基本周波数の2次微分が極大値をとる点(基本周波数が急激に上昇しはじめる点)から, 基本周波数が最大値をとる点までである. 前半部最後の「てみたい」の区間における勾配は, 話者Aが-1.01, 話者Bの物真似音声は-1.03, 話者Bの地声音声は-0.80であった. 以上で比較をした区間のうち, 文頭の「いちど」「いい」の上昇部「から」を除いては, 話者Aの音声と話者Bの物真似音声の基本周波数における勾配が近い値であった. 従って, 話者Aの音声と話者Bは物真似音声の文章1前半部においては, 基本周波数の変

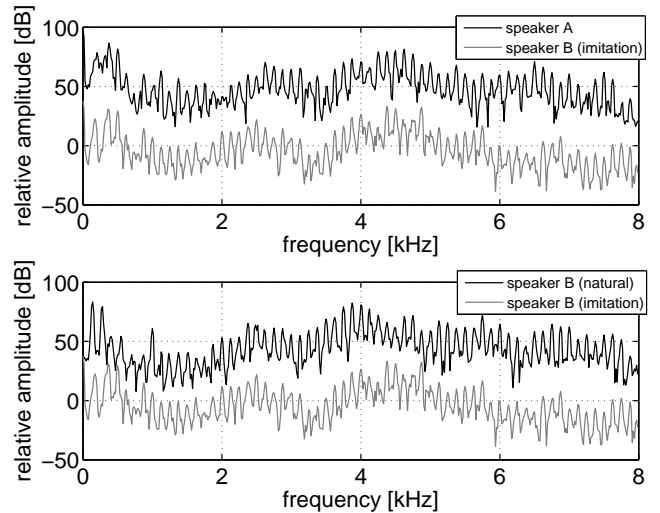


図7 文章1前半部「いい」の平均DFTスペクトル. 上段: 話者Aの音声と話者Bの物真似音声, 下段: 話者Bの地声音声と物真似音声.

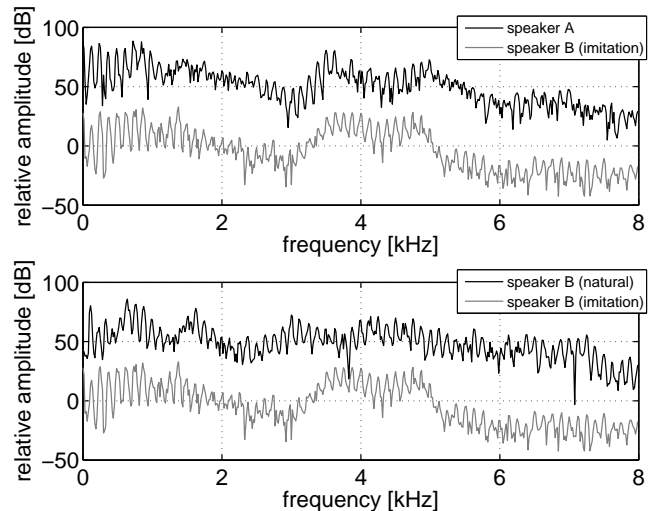


図8 文章1前半部「ら」中の「あ」の平均DFTスペクトル. 上段: 話者Aの音声と話者Bの物真似音声, 下段: 話者Bの地声音声と物真似音声.

化パターンが似ているとみなせる. しかし, 文章1後半部(図4), 文章2前半部(図5), 文章2後半部(図6)には基本周波数の値や勾配が大きく異なる部分も見られる.

4.3 スペクトル

文章1前半部「いい」のDFTスペクトルをフレーム方向に加算平均した結果(平均DFTスペクトル)を図7に示す. 同様に, 文章1前半部「ら」中の「あ」の平均DFTスペクトルを図8に示す. いずれの図でも話者Aの音声と話者Bの物真似音声のスペクトルの概形が似ており(図7上段, 図8上段), 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. 話者Aの音声と話者Bの物真似音声のスペクトルでは, スペクトルのピークだけでなく3kHz前後の谷も一致している. なおかつ, 地声音声ではこの谷が見られない. 従って, 話者Bは物真似の際にこの谷を生み出しているといえる.

表 1 文章 1 前半部「いい」のホルマント周波数 (Hz) .

	F1	F2	F3	F4
speaker A	390.1	2687.5	3171.9	4031.2
speaker B (imitation)	406.2	2468.8	—	4049.9
speaker B (natural)	250.0	2484.4	2906.2	4031.2

表 2 文章 1 前半部「ら」中の「あ」のホルマント周波数 (Hz) .

	F1	F2	F3	F4
speaker A	765.6	1484.4	3578.1	4000.0
speaker B (imitation)	796.9	1343.8	3593.8	3968.8
speaker B (natural)	671.2	1531.2	3078.1	4140.6

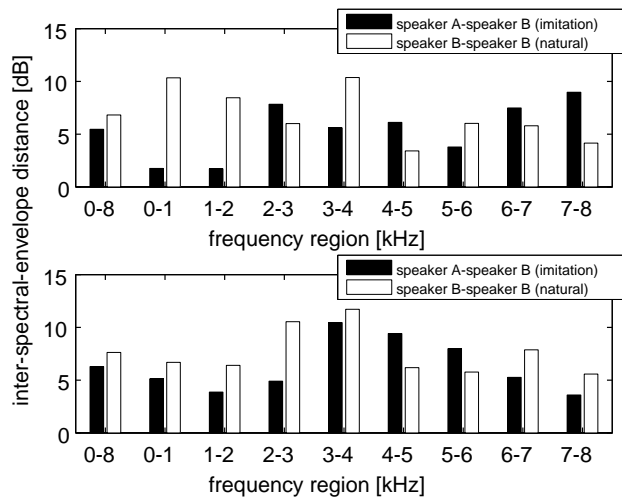


図 9 文章 1 前半部「いい」(上段) および「ら」中の「あ」(下段) に関するスペクトル包絡間距離 .

また、「あ」の物真似音声(図 8)では、約 5 kHz を境に振幅が減衰する様子も話者 A のスペクトルと類似している。このように、話者 B は広い周波数帯域においてスペクトルを話者 A のものに似せていることがわかる。

上記「いい」および「あ」のホルマント周波数をそれぞれ表 1, 2 に示す。話者 A の音声のホルマント周波数と話者 B の物真似音声のホルマント周波数の近さを調べるために、これらの比を求めたところ、「いい」の第 1 ホルマント周波数 (F1) が 1.04, 第 2 ホルマント周波数 (F2) が 0.92, 第 4 ホルマント周波数 (F4) が 1.00 であった。さらに、「あ」の F1 が 1.04, F2 が 0.91, F3 が 1.00, F4 が 0.99 であった。すなわち、F2 以外のホルマント周波数においては話者間の差が 4 % 以内に収まっている。

また、話者 B が物真似の際に地声からどの程度ホルマント周波数を変化させているのかを調べるために、これらの音声のホルマント周波数の比を求めた。その結果、「いい」の F1 が 1.62, F2 が 0.99, F4 が 1.00 であった。さらに、「あ」の F1 が 1.19, F2 が 0.88, F3 が 1.17, F4 が 0.96 であった。すなわち、ホルマント周波数を最大で 62 % 変化させていたことになる。一方で、F4 は地声でも話者 A のものに近いことがわかる。

さらに、上記「いい」および「あ」の 0 kHz ~ 8 kHz (全帯域), 0 kHz ~ 1 kHz, 1 kHz ~ 2 kHz, 2 kHz ~ 3 kHz, 3 kHz ~ 4

kHz, 4 kHz ~ 5 kHz, 5 kHz ~ 6 kHz, 6 kHz ~ 7 kHz, 7 kHz ~ 8 kHz におけるスペクトル包絡間距離 (式 1) を図 9 に示す。全帯域での比較では、話者 B の物真似音声の方が地声音声よりも話者 A の音声とのスペクトル包絡間距離が小さく、その差は「いい」において 1.35 dB、「あ」において 1.34 dB であった。また、多くの周波数帯域で物真似音声の方が話者 A の音声とのスペクトル包絡間距離が小さく、特に 4 kHz 以下では 1 つの例外(「いい」の 2 kHz-3 kHz, F2 を含む帯域)を除いて物真似音声の方がスペクトル包絡間距離が小さい。ただし、現段階では周波数帯域ごとのスペクトル包絡間距離と類似性知覚との関係が明らかではないため、図 9 の結果を知覚上の類似性と結び付けて考えることはできない。

4.4 声帯音源特性

図 7, 8 に関して声帯音源特性の指標となる第 1, 第 2 調波の振幅の差 (H1-H2) を求めた。その結果「いい」(図 7) については、話者 A の音声において -8.03 dB, 話者 B の物真似音声において -8.12 dB, 地声音声において 3.82 dB であった。話者 A の音声と話者 B の物真似音声の H1-H2 の値が極めて近い。一方、「あ」(図 8) については、話者 A の音声において -2.22 dB, 話者 B の物真似音声において -11.55 dB, 地声音声において 18.3 dB であった。「いい」と「あ」のいずれにおいても、話者 B の音声と話者 A の物真似音声の H1-H2 のの符号が一致し、話者 B の地声音声は符号が異っている。この結果は、物真似音声では声帯音源も制御して似せていることを表している。

5. 考 察

話者 B は、物真似音声の多くの区間において、平均基本周波数を高く誇張し、かつ基本周波数の変動パターンを話者 A のものに似せていた。物真似音声において平均基本周波数が高くなる傾向は、Zetterholm [3] も報告している。Akagi と Ienaga [11] 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 “the dynamics of the F0 contour” も個人性知覚に寄与することを示した。そして、彼らは、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、Akagi と Ienaga の結果から見て理にかなっているといえる。

また、話者 B はスペクトルの概形やホルマント周波数も話者 A に似せていた。本研究で比較した「いい」と「あ」においては、話者 A の音声と話者 B の物真似音声の F1, F3, F4 の差は 4 % 以内であった。話者 B の物真似音声と地声音声のホルマント周波数を比較すると、特に「いい」の F1 を大きく変化させていた(表 1)。母音の個人性の知覚要因に関する先行研究ではスペクトルの高周波数成分の寄与が大きいと報告されているが [12] [13] [14], F1 を対象話者のものに近付けることが個人性の類似性知覚に与える影響を明らかにする必要がある。

さらに、話者 B は、物真似発話の際にスペクトルの 3 kHz 前後に谷を生じさせていた。この谷は話者 A のスペクトルにも見られる。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。

この谷の生成要因として、(1) 主声道の分岐管、(2) subglottal resonance、(3) 鼻腔共鳴の3つの可能性が考えられる。以下、それぞれの可能性について考察する。第1の可能性は、物真似発話の際に主声道の分岐管の形状を変えていることである。分岐管としては、左右もしくはそのいずれかの梨状窩が考えられ、喉頭蓋谷の可能性も否定できない。これらの分岐管は変形させることができると報告されており [15] [16]、その形状によっては 3 kHz 付近に谷を生じさせる。

第2の候補に挙げた subglottal resonance は、声門下すなわち気管から肺において生じる共鳴である。声道と声門下が連結した場合には、この共鳴により音声スペクトルに谷を生じる。成人男性の subglottal resonance の第1、第2、第3共鳴周波数は、それぞれ約 600 Hz、1550 Hz、2200 Hz とされている [10]。話者 A はしわがれ声に分類される声質を有している。しわがれ声の発声時は声帯が完全には閉鎖しないため、声道と声門下が連結し、スペクトル上に谷が生じると考えられる。4.4 節で示したように、話者 B はその声帯音源特性も話者 A に似せている。そのため、物真似発話時に声門が完全閉鎖せず、スペクトル上に subglottal resonance による谷が生じた可能性がある。実際、話者 B の物真似音声「いい」のスペクトル (図 7) の 0.8 kHz 付近、1.8 kHz 付近にも零の影響が見られ、subglottal resonance の影響を示唆している。しかし、3 kHz 前後という谷の周波数は上述の第3共鳴周波数 (2200 Hz [10]) とは差があるため、この谷が subglottal resonance に由来すると断定することはできない。

第3の可能性は、物真似の際に主声道と鼻腔とが連結し、それによってスペクトル上に谷が生じたというものである。しかし、物真似音声「いい」のスペクトル (図 7) の谷は約 3.2 kHz、「あ」のスペクトル (図 8) の谷は約 2.9 kHz であり、その間に数百 Hz の差がある。鼻腔の共鳴周波数は基本的に変化しないため、当該の谷が鼻腔共鳴により生じた可能性は低い。ただし、ヒトは鼻孔を動かすことができるため、その可能性を完全に否定することはできない。

話者 B は、基本周波数や声道音響特性の他、声帯音源特性も話者 A のものに似せていた。これは話者 A のしわがれ声に似せる効果があると考えられる。このことは、個人性知覚における声帯音源特性の寄与を示した、Kitamura と Saitou [14]、齋藤と北村 [13] の結果を支持するものである。

話者 A の音声と話者 B の物真似音声の F1、F3、F4 の差が 4% 以内だったのに対して、F2 には 9% 前後の差があった。この結果が音声には意識的に変えられない周波数特性が存在することを意味するのか、もしくは F2 が音声の類似性に対する寄与が小さいことを意味するのかについてはさらなる検討が必要である。

また、音節継続時間長に関しては、話者 A の音声と話者 B の物真似音声の間の相関よりも話者 A の音声と話者 B の地声音声の間の相関の方が高かった。しかし、個人性の類似性知覚に対する音節継続時間長の寄与が否定されたわけではない。話者 B が何らかの誇張をした結果、相関が低くなった可能性もある。今後この点に関する検討も必要である。

6. おわりに

本研究では、物真似においてどのような音響特微量が制御されているのかを明らかにするために、物真似の対象話者 (話者 A) の音声、プロの物真似タレント (話者 B) の物真似音声および地声音声を比較した。2つの文章 (都々逸) の音声データを分析した結果、話者 A の音声と話者 B の物真似音声との間に以下の関係があることが示された。

- (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。

音声の物真似は声帯模写とも呼ばれるが、基本周波数や声帯音源特性だけを真似しているわけではなく、声道音響特性も似せていることが明らかになった。機会があれば、磁気共鳴画像法 (MRI) を用いて物真似音声の生成過程を解明したい。

謝辞

本研究の一部は総務省戦略的情報通信研究開発推進制度 (071705001) の援助を受けた。

文 献

- [1] 鈴木松美, “声紋: 声帯模写から見た声の識別,” 言語生活 207, 37–41 (1968).
- [2] E. Zetterholm, “A comparative survey of phonetic features of two impersonators,” *TMH-QPSR*, 44, 129–132 (2002).
- [3] E. Zetterholm, “Impersonation – reproduction of speech,” *Working Papers Dept. of Linguistics, Lund University*, 49, 176–179 (2001).
- [4] E. Zetterholm, “Same speaker – different voices: A study of one impersonator and some of his different imitations,” *Proc. Int. Conf. Speech Sci. & Tech.*, 70–75 (2006).
- [5] F. Clermont and E. Zetterholm, “F-pattern analysis of professional imitations of “hallå” in three Swedish dialects,” *Working Papers Dept. of Linguistics, Lund University*, 52, 25–28 (2006).
- [6] WaveSurfer, <http://www.speech.kth.se/wavesurfer/>
- [7] Matlab, <http://www.cybernet.co.jp/matlab/>
- [8] D. Talkin, “A robust algorithm for pitch tracking (RAPT),” In W.B. Kleijn and K.K. Paliwal (eds.) *Speech Coding and Synthesis* (495–518). Amsterdam, Lausanne, New York, Oxford, Shannon, Tokyo: Elsevier (1995).
- [9] H.M. Hanson, K.N. Stevens, H.-K. J. Kuo, M.Y. Chen, and J. Slifka, “Towards models of phonation,” *J. Phonet.* 29, 451–480 (2001).
- [10] K.N. Stevens, *Acoustic Phonetics*, Chap. 3, 196–198 (1998).
- [11] M. Akagi and T. Ienaga, “Speaker individuality in fundamental frequency contours and its control,” *Journal of the Acoustical Society of Japan(E)* 18(2), 73–80 (1997).
- [12] 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,” *日本音響学会誌* 54(3), 185–191 (1997).
- [13] 齋藤毅, 北村達也, “3連続母音に含まれる個人性知覚に寄与する音響特微量,” *信学技報 (SP)*, 106(613), 43–48 (2007).
- [14] T. Kitamura and T. Saitou, “Effects of acoustic modification on perception of speaker characteristics for sustained vowels,” *Acoust. Sci. & Tech.* (accepted).
- [15] 竹本浩典, 北村達也, 本多清志, 正木信夫, “声の高さの変化による下咽頭腔の変形とその音響効果,” *音講論 (春)* 253–254 (2006).
- [16] 本多清志, 竹本浩典, 中島淑貴, 足立整治, 平原達也, “ピッチを変えて発声した無声母音の声道形状,” *音講論 (春)* 349–350 (2006).