物真似音声の分析*

○北村達也 (甲南大)

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．物真似音声が似て知覚される要因や，物真似音声の生成機構を明らかにできれば，人間の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．

本研究では，プロの物真似タレントが物真似した音声と物真似の対象者の音声とを比較する．これによって，音声中のどのような成分を似せているのかを分析する．さらに，物真似タレントが物真似した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する．

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 U と物真似タレントの話者 C が「一度でいいから見てみたい，女房 (にょうぼ) がヘソクリ隠すとこ.」という文章を発話したものである．括弧内は発音の仕方を示したものである．なお，本稿では「一度でいいから見てみたい」を前半部，「女房がヘソクリ隠すとこ」を後半部と表す．

話者 C の音声は，物真似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 C は話者 U の発話を聞いた直後に発話した．話者 U と C が 2 回発話した中から音声データを 1 つずつ選択した．一方，話者 C の地声は，物真似音声の収録とは別の日に話者 U 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,「ぼ」の部分に飽和があった．

以上 3 種類の音声データはいずれも標本化周波数 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 (F0)，DFT スペクトルを求め比較した．音声データが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．音節継続時間長はスペクトログラムを参考にしつつ決定した音節区間から求めた．F0 は，音声処理ソフトウェア WaveSurfer[1] の F0 抽出機能を用いて抽出した．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 ESPS method，フレーム長は 7.5 ms，フレーム周期は 10 ms を用いた．明らかな抽出誤りは目視で修正した．DFT スペクトルはフレーム長 64 ms，フレーム周期 8 ms で求めた．窓関数はハニング窓を用いた．

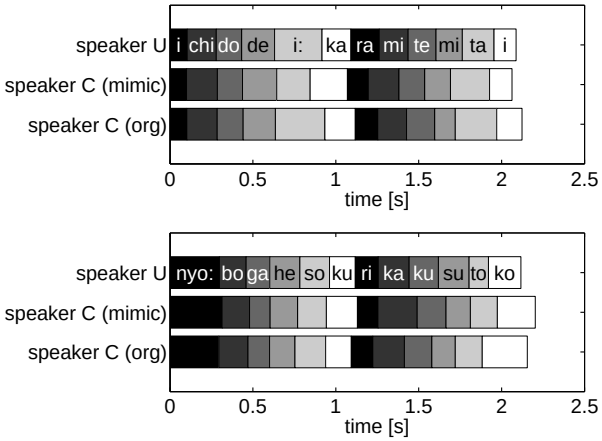

Fig. 1 syllable duration of the first half (top) and second half (bottom) of the sentence..

4 結果

4.1 音節継続時間長

各音節の継続時間長を図 1 に示す．継続時間長に関して音声データ間の相関係数 r を求めた．その結果，前半部では話者 U の音声と話者 C の物真似音声の間に 0.652，話者 U の音声と話者 C の地声音声の間に 0.898 の相関があり，後半部では話者 U の音声と話者 C の物真似音声の間に 0.906，話者 U の音声と話者 C の地声音声の間に 0.797 の相関があった．物真似音声において必ずしも音節継続時間長を似せているわけではない．

4.2 F0

F0 を図 2 に示す．同図下段の約 1.1 s から約 1.6 s の間 (後半部の「ヘソクリ隠すとこ」の「くりかく」に対応) にずれが見られる以外は，話者 U の音声と話者 C の物真似音声の F0 変動パター

*Analysis of impersonated speech. by KITAMURA, Tatsuya (Konan University)

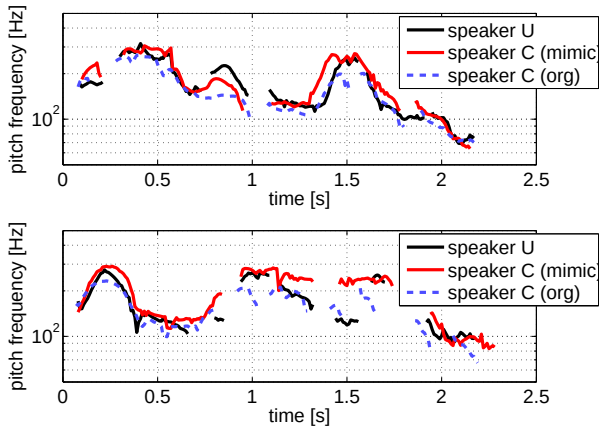

Fig. 2　F0 of the first half (top) and second half (bottom) of the sentence.

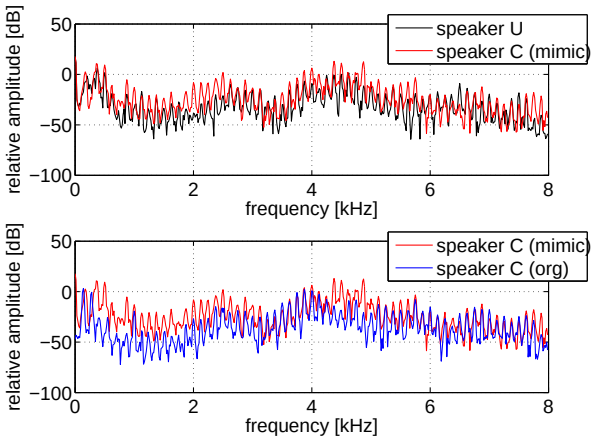

Fig. 3　DFT spectra of /i:/ of speaker U and impersonated /i:/ of speaker C (top) and those of impersonated and natural /i:/ of speaker C (bottom).

ン (各部分の値と勾配) が酷似している．

有声区間の平均 F0 は話者 U が 167.2 Hz，話者 C の物真似音声が 185.1 Hz，地声音声が 152.0 Hz であった．話者 C の物真似音声の平均 F0 が話者 U のものより大幅に高いのは，主に上述の「くりかく」の部分の F0 がずれているためである．

4.3　DFT スペクトル

前半部の「いい」の区間の DFT スペクトルをフレーム方向に平均したものを図 3 に示す．話者 U の音声と話者 C の物真似音声の DFT スペクトル (図 3 上段) は，第 1 ホルマント付近 (約 1 kHz 以下) と約 2.6 kHz 以上の周波数帯域の形状がよく似ている．一方で，話者 C の物真似音声と地声音声の DFT スペクトルはこの周波数帯域の形状が異なっている (図 3 下段)．従って，話者 C は意識的にこの周波数帯域のスペクトル形状を話者 U のものに合わせていると考えられる．

また，図 3 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，話者 C の音声では-8.03 dB，話者 U の物真似音声では- 7.55 dB，地声音声では 3.82 dB であった．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 C は物真似音声において，平均 F0 を高くするだけでなく，F0 変動パターンも似せていた．F0 変動パターンも個人性知覚に寄与すると報告されているので [2]，理にかなっている．また，スペクトルの高周波数帯域には個人性が顕著に現れ [3]，それは主に下咽頭腔の形状の個人差に起因する [4] と報告されている．従って，話者 C は下咽頭腔を変形することでスペクトルの高周波数帯域の形を変え，それと同時に喉頭に力みを加えてしわがれ声を作り出すことで，声質を話者 U に似せていると考えられる．

しかし，話者 U の音声と話者 C の物真似音声の第 2 ホルマント周波数 (F2) は異なる (図 3 上段)．これ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は今後の検討課題である．

6　おわりに

プロの物真似タレントによる物真似音声の分析を行った．その結果，本研究で対象にした音声データにおいては，基本周波数の変動パターン，スペクトルの第 1 ホルマント付近と高周波数帯域，声帯音源を似せていることが明らかになった．

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．

参考文献

[1] `http://www.speech.kth.se/wavesurfer/`
[2] Akagi and Ienaga, J. Acoust. Soc. Jpn(E)., 18(2), 73–80, 1997.
[3] 北村, 赤木, 音響誌, 54(3), 185–191, 1997.
[4] Kitamura *et al.*, Acoust. Sci. Tech., 26(1), 16–26, 2005.