

MRIによる物真似発話時の声道形状観測*

☆川岸容子 北村達也 (甲南大・知能情報)

1 はじめに

物真似タレントは、複数話者の音声によく似た声質を出し分けることができる。この技巧を解明することができれば、発話器官と声質との関係や個人性知覚に関する有益な知見を得られるとともに、音声生成機構に基づく音声合成 [1] において多彩な声質を合成できるようになると考えられる。

このような背景の下、近年、物真似発話の研究が増えてきている。例えば、Kitamura [2] は物真似音声では基本周波数の変動パターン、フォルマント周波数、声道音源特性が目標話者のものと類似していることを報告している。しかし、音響分析の研究が行われている一方で、その生成過程を観測した研究は少ない。Aminら [3] は、物真似発話時の音声と Electroglottograph (EGG) 波形を分析している。しかし、著者らの調べた限り、磁気共鳴画像法 (Magnetic resonance imaging, MRI) 等を用いて物真似発話時の声道形状を直接観測した研究は見られない。

このたび、著者らは物真似タレントの物真似発話時の声道形状をMRIにて観測する機会に恵まれ、さらにそのデータを分析する許可を得たので、結果を報告する。

2 方法

2.1 実験協力者

日本語を母語とする物真似タレント1名(37歳男性)である。

2.2 発話タスク

物真似をしていない状態、および目標話者12名(うち1名は女性)の物真似をした状態にて母音/a/および/e/を持続発話時に撮像した。実験協力者は仰臥位でMRI装置内に入り、その中で発話した。予行演習も兼ねて先に文章を物真似発話し、その後に母音を発話した。撮像直前に発話された音声を実験者ら数名がMRI室外のスピーカーにて聴取し、目標話者の物真似であることと当該の母音であることを確認した。

Table 1 Conditions of MRI measurement.

Echo time (TE)	2.14 ms
Repetition time (TR)	4.34 ms
Flip angle (FA)	12°
Field of view (FOV)	200×200 mm
Pixel size	0.89×0.89 mm
Slice	Sagittal
Slice thickness	4.0 mm
Number of slices	8
Number of average	12

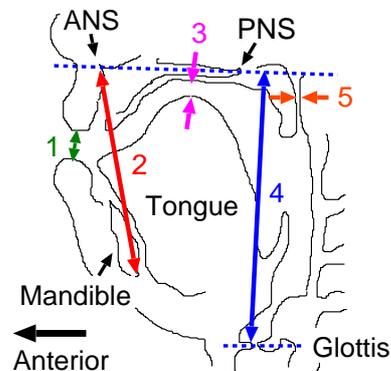


Fig. 1 Definition of distances 1, 2, 3, 4, and 5 on the mid-sagittal plane.

2.3 MRI 撮像条件

神田クリニック (東京) の MRI 装置 Philips Achieva 1.5T にて撮像した。撮像条件は表1の通りである。矢状断面8枚を撮像し、このうち正中矢状断面に相当するMRIデータを以下の分析に用いた。なお、本研究では音声データは入手できていない。

2.4 分析方法

本研究では以下の5つの距離を計測し(図1参照)、その変化の範囲を求めた。

距離1 上下口唇間の最短距離

距離2 Anterior nasal spine (ANS) から下顎骨下端までの距離

距離3 舌の最高点と口蓋の最短距離

* Measurement of vocal tract shape during imitating voice. by KAWAGISHI, Yohko, KITAMURA, Tatsuya (Konan University)

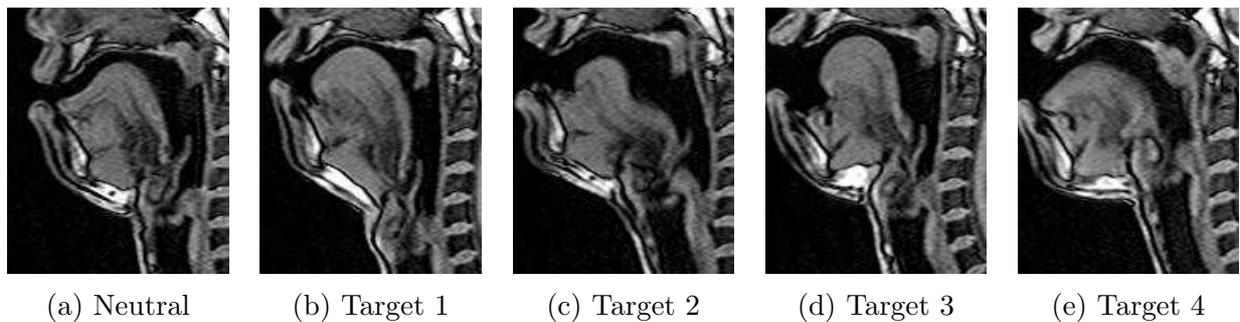


Fig. 2 Mid-sagittal images acquired during speaking the vowel /e/ in neutral (a) and imitated articulations for four target speakers (b, c, d, and e).

距離 4 ANS と Posterior nasal spine (PNS) を通る直線から声門までの最短距離

距離 5 口蓋帆と咽頭後壁の最短距離

距離 1, 2, 3 は開口の程度を評価するため、距離 4 は喉頭の高さを評価するため、距離 5 は鼻咽腔結合の程度を評価するために計測した。距離 1, 3, 5 は各発話器官をトレースしたデータから計測した。また、距離 2, 4 は MRI データ上で ANS, PNS, 下顎骨下端, 声門の位置を目視で定めてから計測した。

3 結果と考察

物真似をしていない状態、および物真似をした状態での母音/e/発話時の正中矢状断面の MRI データを図 2 に示す。後者は目標話者 12 名中 4 名の音声を真似た際のもので、このうち図 2(e) は女性話者を真似たものである。

これらの MRI データから、物真似の際には舌を多彩に変形していることがわかる。図 2(b) では喉頭を約 20 mm 下降させ、声道長を長くし、低い声を作っている。(c) の舌形状は特徴的で、/e/ 発話のためのせばめを作りつつ(ただしせばめはあまり狭くない)口腔後方に大きな空間を作り出している。一方、(d) ではせばめの前方の空間が広がっている。女性話者を真似た(e)では、硬口蓋より前方の歯茎部と舌との間でせばめを作っている。加えて、喉頭を約 20 mm 上昇させることによって、声道長を変化させ身体サイズの違いを表現するとともに、音声の基本周波数を変化させていることがうかがえる。

距離 1~5 の計測結果を表 2 に示す。最も顕著な変化を示したのは距離 4(喉頭の高さに対応)で、両母音とも約 40 mm の範囲で上下している。これは、音声の個人性にとって声道長や基本周波数が重要な情報であることと対応している。

Table 2 Distances 1, 2, 3, 4, and 5 for the vowels /a/ and /e/ during neutral and imitated articulations in mm.

Distance	Neutral		Imitated	
	/a/	/e/	/a/	/e/
1	13	11	8-36	9-32
2	88	91	90-107	92-108
3	18	7	8-23	4-16
4	105	98	77-116	76-118
5	0	0	0-5	0-3

4 おわりに

本研究では MRI を用いて物真似タレントが物真似発話をした際の声道形状を計測した。実験協力者が音声のフィードバックを頼りに複雑な発話器官の制御を習得したことは大変興味深い。今後、3次元声道形状やその動きに関するデータ、EGG 等の喉頭制御に関するデータが計測され、物真似発話のメカニズムが解明されることを期待する。

謝辞 本研究の一部は平成 26 年度科研費(25280026, 25280066)の支援により行われた。研究資料の利用を許諾していただいた日本テレビ放送網株式会社および実験協力者、MRI 撮像にご協力いただいた医療法人社団 裕健会 神田クリニックに感謝する。

参考文献

- [1] 北村, 音響誌, 67(1), 28-33 (2011).
- [2] Kitamura, *Interspeech2008*, 813-816 (2008).
- [3] Amin *et al.*, *IEEE Trans. Multimedia*, 16(3), 668-678 (2014).