

MRI 観測にもとづく音声生成系研究

Studies on the human speech production system based on MRI observations

甲南大学 北村 達也
Tatsuya Kitamura
Konan University

Abstract This paper gives an outline of studies on human speech production system using magnetic resonance imaging (MRI). We have developed techniques on MRI to explore speech production mechanisms. Based on MRI observations, we have investigated the source of speaker characteristics and have proposed a new speech production theory.

1 はじめに

磁気共鳴画像法 (Magnetic Resonance Imaging, MRI) は、この 20 年の間に発話器官の任意断面の形態や動態の観測に盛んに利用されるようになり、今や音声生成系の研究分野の極めて重要なツールになっている。しかし、その過程においては、音声生成系研究にとっての MRI の短所を克服する様々な工夫が加えられてきた。本稿では、それらの工夫と、その観測技術を利用して行われている研究について概説する。

2 発話器官の形態・動態観測技術

音声生成系の研究者にとって、外から見えない発話器官の形や動きを見たい、という思いは今も昔も変わらない。発話器官の形態や動態の観測は MRI が開発される前から試みられてきた。X 線透過静止画は、音声の科学的研究の嚆矢となった Chiba と Kajiyama による *The Vowels*[1] においても重要な役割を果たした。その後、X 線映画による動態観測も実現し、近年には X 線 CT が開発されたが、被曝の問題が避けられなかった。

一方、MRI は被曝の心配がない安全な観測手法といえる。さらに、3 次元的な形態情報を、通常分析には十分な空間分解能 (例えば $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$) で得られるのも利点である。発話機構観測への利用は 1980 年代後半から始まり、以来 20 年以上が経過するが、一般的な MRI を発話機構観測に用いる場合には短所も存在する [2]。

その短所として、(1) 撮像に時間を要し、動画像を撮像できないこと、(2) 喉頭部などの観測においては、空間分解能が不足する場合があること、(3) 歯が空気と同等の信号値しか持たないため、可視化できないこと、(4) 大きな撮像騒音が発生し、発話に影響を与える可能性があること、などが挙げられる。

しかし、これらの問題は、様々な工夫によって部分的にはあるが解決されてきている。以下では、その解決策と、その結果得られた成果について述べる。

3 MRI による動態観測

MRI は時間分解能が低いため、一般には動画を撮像するのが難しい。しかし、心臓の拍動のように同じ運動が周期的に繰り返される場合は、位相をずらしながら撮像することによって一連の運動の動画を得ることができる。発話運動の場合も、被験者に同じ発話を繰り返させ、位相をずらしながら撮像することによって、発話動態を観測できる [3]。ただし、被験者は発話資料を同じタイミング、同じ動作で 100 回程度繰り返す必要があり、その負担は小さくない。

この撮像法は、それ以前に知り得なかった多くの知見をもたらしている。例えば、Takemoto ら [4] は母音連続発話を上記の撮像法で観測し、声道形状並びに声道断面積関数¹の時間変化を抽出した (図 1)。現在、この撮像法は音声生成系の研究のみならず、音声合成、臨床研究、リハビリテーションにも応用されている。

4 高感度小型表面コイルによる高分解能撮像

発話器官の中には喉頭部のように、粘膜に囲まれた微細な軟骨を含む部分がある。また、唇の内部のように複雑に筋繊維が絡み合う構造もある。このような対象を撮像する場合には、通常よりも細かい空間分解能が要求される。一般に撮像対象の空間分解能を上げようとすると、観測対象のボクセルサイズは小さくなり、そこから出力される信号強度は弱まる。これを補うためには、高感度のアンテナ (コイル) が必要となる [2]。

¹声門から口唇までを結ぶ声道中心線に直交する断面の面積をプロットした 2 次元データ。

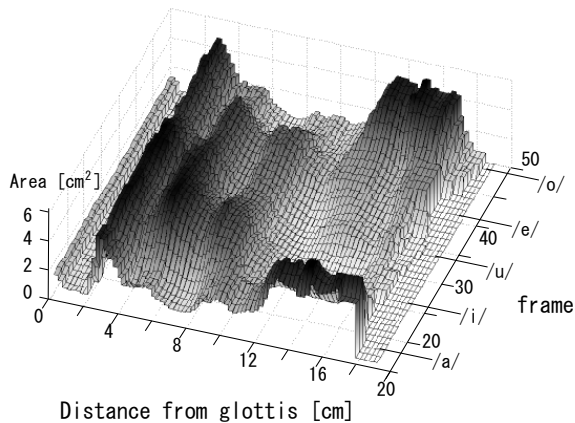


図 1: 3次元 MRI 動画から抽出された声道断面積関数の時間変化 [4].

そこで、高感度小型表面コイル(図 2(左)) が ATR にて開発された。このコイルを用いた撮像により、高分解能 (0.19 mm × 0.19 mm) の MR 画像が取得できた(図 2(右)) [5]。このコイルは、基本周波数の変化に伴う甲状軟骨の動きの観測、口唇の突き出しに伴う筋肉の動きの観測、下咽頭腔形状の個人内変動の観測 [6] などを可能にし、それぞれの検討から多くの知見が得られている。

5 歯列形状の観測

発話器官において歯は重要な要素の一つであるが、MRI では歯が空気と同じく低輝度に撮像され、歯の境界を抽出できない。この問題点を克服する方法が検討されている。一例を挙げると、Takemoto ら [7] はマンガ含有量の多い飲料を口腔内に含んだ状態で MRI 撮像を行い、その飲料と歯の信号値の差を利用して歯列形状を抽出する方法を提案した。そして、別に撮像しておいた発話時の MR 画像に歯列を補填することによって、この問題を解決した。この手法を利用して、摩擦子音の発音原理を探る研究などが行われている。

6 音声の個人性(話者の特徴)の生成要因

音声の個人性は、音声生成系の生得的な特徴に起因する特徴と、話し方や方言などの後天的な特徴に現れる。ここでは、前者に関して筆者らが MRI を用いて行った研究 [6] について紹介する。

音声の個人性は「発話内容が変わっても話者内で変動しない成分」と考えることができる。従って、声道内の発話内容によって変動しない部位は個人性の生成要因の一つと考えられる。この考えにもとづいて日本語 5 母音発



図 2: 高感度小型表面コイルによる喉頭撮像の様子(左) [5] と得られた喉頭の横断面の MR 画像(右) [6].

話時の MRI 観測を行い、声道下部は発話内容による変動が小さいことを発見した。

さらに、この声道下部の個人差が与える音響的影響を調査したところ、音声スペクトルの約 2.5 kHz 以上の周波数帯域に顕著な影響が現れることが明らかになった。この周波数帯域は、音声の個人性知覚において寄与が大きいことがすでに知られていた。従って、上記の研究成果は、個人性知覚に関する知見を音声生成系の研究によって裏付けたと見ることができる。

7 おわりに

MRI は従来の観測手法では得られなかった成果をもたらしてきたが、撮像騒音の低減など未解決の課題もある。これらの課題は、幅広い分野の人材の協力が得られれば解決することができるものと期待している。本稿がそのきっかけとなれば著者にとって望外の喜びである。

謝辞 本研究の一部は 2008 年度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] T. Chiba and M. Kajiyama, *The Vowels – Its Nature and Structure*, Tokyo-Kaiseikan, Tokyo (1942).
- [2] 正木信夫, 音声・聴覚研究の新技术としての MRI, 聴覚研究会資料, 37, 637–664 (2006).
- [3] S. Masaki *et al.*, MRI-based speech production study using a synchronized sampling method, *J. Acoust. Soc. Jpn. (E)*, 20, 375–379 (1999).
- [4] H. Takemoto *et al.*, Measurement of temporal changes in vocal tract area function from 3D cine-MRI data, *J. Acoust. Soc. Am.*, 119, 1037–1049 (2006).
- [5] 高野佐代子ら, 高分解能 MRI による輪状甲状関節の観測, 音声言語医学, 46, 174–178 (2005).
- [6] T. Kitamura *et al.*, Individual variation of the hypopharyngeal cavities and its acoustic effects, *Acoust. Sci. & Tech.*, 26, 16–26 (2005).
- [7] H. Takemoto *et al.*, A method of tooth superimposition of MRI data for accurate measurement of vocal tract shape and dimensions, *Acoust. Sci. & Tech.*, 25, 468–474 (2004).