

音声生成系の様々な可視化手法*

○北村 達也 (甲南大)

1 はじめに

音声の生成原理の理解は音声学における研究領域の1つであり、調音音声学と呼ばれる。その基本的な研究手法は、(1) 音声の分析、(2) 音声生成系の分析、(3) 実体モデルや数理モデルに基づくシミュレーション、の3つに整理することができる。これらの手法は独立して用いられるばかりではなく、複数を組み合わせた研究も行われる。

本稿では、上記2つ目の音声生成系の分析を支える可視化手法について、音響誌に掲載された解説 [1] に基づき概説する。なお、脳活動の観測 [2] は本稿の対象外とする。発声や発話の様子を可視化する意義(魅力)は、研究対象の音声が生じる原因や原理に直接迫ることができる点にある。そのために、従来見ることができなかったものを、従来を超える精度や規模で可視化する努力が続けられている。

一方で、音声生成系の観測には特殊な装置や解剖学的な知識が必要になるなど、参入しにくい印象があることも事実である。しかし、最近では装置の利用法や観測データの解析法に関するセミナーが開催されたり、研究用の観測データが公開されたりするなど状況は変わりつつある。本稿によって音声生成系の観測に興味を持っていただけたら幸いである。

2 発声器官の可視化手法

2.1 間接的観測法(非侵襲的観測法)

医師でない者が発声器官を観測する際には、非侵襲的な手法を用いる必要がある。Electroglottography (EGG) は代表的な非侵襲的手法(間接的観測法)である。EGG では声門開口面積の相対的な変化を計測することができ、装置の使い方も容易である。

EGG による計測においては、2つの直径3 cm~4 cmの円状の電極を対象者の甲状軟骨を左右からはさむように首に固定する。そして、この電極間に高周波数電流を流し、甲状軟骨の内側にある声門の開閉によるアドミタンスの変化を時間波形として計測する。ここで、EGG は声

門開口面積そのものを計測しているわけではないことに注意が必要である。EGG データからは声門開閉のタイミングや声門開放時間率 (Open quotient, OQ) 等の情報を得ることができる。

2.2 直接的観測法

もし医師の協力が得られれば、ファイバースコープや硬性側視鏡を鼻孔もしくは口から咽頭に挿入し、声帯振動を直接観測することができる。これらの手法は、直接的観測法と呼ばれる。

著者らは、最近、耳鼻咽喉科の医師の協力を得て、喉頭ストロボスコープにより声帯振動を観測する機会を得た。この手法は、発声中の声帯に間欠発光するストロボ光を当てることによってスローモーション映像を得るものである。

著者らの実験では、ファイバースコープを経鼻的に挿入し、声帯振動の観測を試みた。その過程で、ファイバースコープ挿入状態への適応には個人差があり、また操作者の手技が得られるデータの質に影響することがわかり、手法ごとのノウハウや経験が欠かせないことを改めて知ることとなった。

3 調音器官の可視化手法

3.1 磁気共鳴画像法

磁気共鳴画像法 (MRI) の利点は、X線CTと異なり被爆の問題がないため健常者を対象にした研究に用いることができ、なおかつ3次元データおよび任意断面のデータが得られる点にある。他の可視化手法では観測しにくい咽頭、喉頭、口蓋帆なども明瞭な画像が得られる。また、近年ではリアルタイムMRI (rtMRI) の技術により動画の撮像が可能になっている。

MRI はコストが高いのが難点の1つであるが、最近、日本語音声発話時のrtMRIデータベース (<https://rtmridb.ninjal.ac.jp/>) [3] が公開され、無料で利用できるようになった。このデータベースでは、14 fpsあるいは27 fpsにて撮像されたrtMRIデータを音素等の属性で検索し、動画を再生することができる。

MRIにて得られるのは画像データであり、調

* Visualization methods of speech production process. by KITAMURA, Tatsuya (Konan Univ.)

音器官の形態や動態について分析する際には、その輪郭を抽出することが多い。かつては手作業により1枚ずつトレースすることが多かったが、作業による抽出結果の差異などの再現性の問題が避けられない。そこで、最近では機械学習を利用して自動的に抽出する方法が研究されている。Takemotoら [4]の手法は、少ない学習データで調音器官の輪郭を高精度に抽出することができる優れた方法である。

3.2 超音波画像診断法

日本国内では調音運動観測における超音波画像診断法の利用が進んでいないが、国外では幅広く利用されており、報告も多い。超音波画像診断法は、MRIと比較してはるかに低コストで、撮像時の騒音も小さく、装置の操作に特別な知識を必要としない。また、得られた画像から舌輪郭を抽出するソフトウェアが複数公開されており、未経験者でも導入しやすい手法である。ただし、可視化できる範囲はほぼ舌に限られる。

装置の形態は様々で、ノートPCタイプのほか、タブレットやスマートフォンに超音波ビームを発するプローブを有線または無線接続する装置もある。プローブにも様々なタイプがあるが、舌運動の観測にはマイクロコンベックス型と呼ばれる、扇形に超音波ビームを発する小型のプローブが用いられることが多い。

観測の際にはプローブを上向きにして下顎に密着させる。その際、実験参加者自身がプローブを手で保持するとその位置や向きが安定しないため、専用のヘッドセットを利用するのが望ましい。著者らは、発話に伴う下顎の動きに対応できるダンパ機構を持たせたヘッドセットを開発中である。

超音波画像診断装置は医療機器のレンタル業者から1日単位でレンタル可能であり、これを活用すればコストを抑えて研究を始めることができる。その意味でも導入しやすい手法である。

3.3 その他の可視化手法

調音運動の可視化手法としては、他にも磁気センサシステム (Electromagnetic articulography, EMA) やエレクトロパラトグラフィ (Electropalatography, EPG) などがある。また、最近では、スマートフォンやタブレットにて高速度かつ高解像度の撮影が可能となり、発話時の口唇の動きや表情の変化の詳細な分析が身近になった。

著者らは、画像処理のクラウド AI を利用して顔のランドマークを自動抽出し、発話時の動きの解析に利用している。この他、音声の放射特性も音声生成における重要な性質であり、マイクロフォンアレイを用いた放射特性の計測 [5] などが行われている。

4 おわりに

本稿では音声生成系の可視化手法について概説した。国立国語研究所では、様々な可視化手法を実際に体験できるハンズオンセミナーを開催している。2023年度の同セミナーでは、EGG, MRI, 超音波画像診断法, EMA, EPGの使用法と観測データの解析法などが解説された。学生、教員、研究者などが参加し、交流の場としても機能している。同セミナーの講義を収録した動画 (YouTubeへのリンク) が公開されているので、興味のある方は視聴していただきたい。また、同研究所が公募している共同利用型共同研究に応募することによって、同研究所が所有する装置を利用した研究が可能である。このようなイベントや制度を起点として、音声生成系に興味を持つ研究者の協力体制を構築することが期待される。

謝辞 本稿の執筆にあたって JSPS 科研費 (Nos. 20H00291, 20H01265, 23K00544), 中山隼雄科学技術文化財団の支援を受けた。可視化手法の研究、実施に関する関係者の皆様、本稿の執筆にあたってご助言いただいた皆様に感謝します。

参考文献

- [1] 北村, 動画で見る音声生成系の観測手法, 音響誌, 76, 700–705 (2020).
- [2] 今泉ら, 聴覚・発話に関する脳活動観測, コロナ社 (2022).
- [3] 前川ら, リアルタイム MRI 動画日本語調音運動データベースの設計, LRW2020, 209–230 (2020).
- [4] Takemoto *et al.*, Speech organ contour extraction using real-time MRI and machine learning method *Proc. INTERSPEECH 2019*, 904–908 (2019).
- [5] Pörschmann & Arend, Investigating phoneme-dependencies of spherical voice directivity patterns *JASA*, 149, 4553–4564 (2021).