

母音発話 MRI データに基づく声道模型の音響特性

北村 達也[†] 竹本 浩典^{††,†††} 本多 清志^{††,††††}

[†] 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

^{††} ATR 認知情報科学研究所

^{†††} NICT ユニバーサルメディア研究センター

^{††††} Laboratories of Phonetics and Phonology, University of Paris III

E-mail: [†]t-kitamu@konan-u.ac.jp

あらまし ATR 母音発話 MRI データから抽出した声道の音響特性を明らかにするために、この MRI データから日本語 5 母音の声道模型を光造形により作製し、その音響特性を測定した。声道模型の口唇部から入力した Optimized Aoshima's time-stretched pulse (OATSP) 信号を声門部に固定したプローブマイクで収録し、声道模型の音響特性を求めた。この測定手法を用いることによって、声門閉鎖時に出現する喉頭腔共鳴を含む声道音響特性を測定することができた。得られた測定結果は、各種の声道音響解析手法に対して、精度を検証するための基準値の 1 つとなる。
キーワード 磁気共鳴画像法 (MRI), ATR 母音発話 MRI データ, 日本語 5 母音, 声道, 声道模型, 音響特性

Acoustic characteristics of solid models based on vowel production MRI data

Tatsuya KITAMURA[†], Hironori TAKEMOTO^{††,†††}, and Kiyoshi HONDA^{††,††††}

[†] Department of Information Science and Systems Engineering, Faculty of Science and Engineering, Konan University

8-9-1 Okamoto, Higashinada, Kobe, Hyogo, 658-8501, Japan

^{††} ATR Cognitive Information Science Laboratories

^{†††} NICT Universal Media Research Center

^{††††} Laboratories of Phonetics and Phonology, University of Paris III

E-mail: [†]t-kitamu@konan-u.ac.jp

Abstract “ATR MRI database of Japanese vowel production” was used to evaluate acoustic characteristics of realistic vocal tracts for five Japanese vowels through the measurements of frequency responses from vocal tract solid models formed by a stereo-lithographic technique. An optimized Aoshima's time-stretched pulse signal generated from a horn driver unit was introduced into the solid model at the lip end. The response signals of the models were recorded at the model's glottis. This method permits accurate measurement of acoustic characteristics of the vocal tract including the laryngeal cavity. The results provide a benchmark for testing numerical analysis methods that have been used to study vocal tract acoustics.

Key words Magnetic Resonance Imaging (MRI), ATR MRI data of Japanese vowel production, five Japanese vowels, vocal tract, vocal tract solid model, acoustic characteristics

1. はじめに

近年、磁気共鳴画像法 (Magnetic Resonance Imaging, MRI) による発話形態・動態観測技術が進歩し、発話中の声道形状を精密に観測できるようになってきた [1]。2005 年には ATR 人

間情報科学研究所から ATR 母音発話 MRI データ [3] が公開され、音声研究の共通基盤として利用できるようになった^(注1)。これには成人男性 1 名の日本語 5 母音発話時の 3 次元 MRI デー

(注1): 現在は (株)ATR-Promotions が管理している。

タおよび音声データが収録されており、研究・教育目的で用いることができる。

そして、MRI の観測技術の進歩を背景にして、有限要素法 (Finite Element Method, FEM) や時間領域差分法 (Finite Difference Time Domain, FDTD) などの手法によって声道音響解析が行われるようになってきている。しかし、これらの研究では、数値計算の精度をいかに評価するのが問題である。声道音響解析の結果を音声のスペクトルと比較した研究 (例えば [2]) も見られるが、実際の音声では声道音響特性以外にも数々の要因が影響しているため、妥当な比較とは言えない。

我々は、声道音響解析の精度検証においては、MRI データから得られる声道形状そのものに着目すべきであり、声道模型の音響特性との比較が最も有効であると考えている。光造形技術を用いれば現在の MRI の分解能を越える精度で声道模型が作製でき、その音響特性は信頼に足るものと考えられるからである。

以上のような背景から、本研究では、ATR 母音発話 MRI データから抽出した声道の音響特性を示すことを目的とする。MRI データに基づいて日本語 5 母音の声道模型を光造形により作成し、これらの音響特性を測定する。得られた測定結果は、声道音響解析手法の基準値の 1 つとして利用することができる。

著者らは従来より、声道音響解析の精度検証を目的として、声道模型の音響特性の測定を試みてきた [4] [5]。この方法は、音響管の入力端の粒子速度と出力端の音圧から音響特性を求めるものである。音響管の入力端の粒子速度は、音響管とホーンドライバユニットを接続する直径 5 mm の管の 2 点の音圧から求めた。この方法によって一様音響管や二段音響管の音響特性は正確に測定できたものの、声道模型の音響特性において予想されるより共鳴の数が 1 つ少ないという問題があり、その原因もわからないでいた。

ところが、ごく最近になって、声門の閉鎖と開放によって喉頭腔に由来する共鳴 (喉頭腔共鳴) が出現と消失を繰り返すことが実験的、理論的に明らかになった [6] [7] [8]。喉頭腔とは、声道の最下部に位置する漏斗状の細い管である。この喉頭腔が声門閉鎖時には閉管となって喉頭腔共鳴が出現させ、声門開放時には開管となるため喉頭腔共鳴が消失することが示されたのである。上述の音響測定法においては、声道模型の声門部が直径 5 mm の管と接続されていたため、声門開放時と同じ状態になり喉頭腔共鳴が出現しなかったと推測される。

このことから、声道模型の音響特性を喉頭腔共鳴も含めて測定するためには、声門閉鎖時のように声門インピーダンスを無限大に近くする必要があることがわかる。そこで、本研究では声道模型の声門部をプラスチック板でふさぎ、そこに開けた穴にブローマイクを挿入し、声道模型の口唇部から入力した Optimized Aoshima's time-stretched pulse (OATSP) 信号 [9] の応答を測定することにした。

本稿では、まず声道模型のもとになった ATR 母音発話 MRI データについて説明し、次に声道模型の作製方法および音響測定方法について述べる。最後に音響測定結果を示し、得られた結果を検討する。

2. ATR 母音発話 MRI データ

この MRI データは、骨伝導と体伝導を用いて MRI 同期撮像法のガイド音を呈示することによって、従来法よりも普通の調音に近い状態を高 S/N で撮像できている [10]。

2.1 MRI 同期撮像法

MRI 同期撮像法は、被験者に周期的なガイド音にタイミングを合わせて発声させ、発声区間のみで撮像する手法である [11]。この手法を用いることにより、吸気や唾液の嚥下に起因するアーチファクトを抑制し高 S/N の MR 画像を得ることができる [12] [13]。この撮像では、120 Hz の倍音から成る調波複合音をガイド音として用いている。

2.2 骨伝導と体伝導によるガイド音の呈示

MRI 撮像時には非常に大きな騒音が発生する。この撮像で用いられた Fast Spin Echo 法による撮像時の音圧レベルは 110 dB を越える [14]。このため、被験者は自分自身の音声をフィードバックできず調音が歪んでしまう。

そこで、この撮像では、被験者に耳栓をさせた上にイヤマフ (Bilsom 715) を装着させて気導音をできる限り遮断し、ガイド音を呈示されている。耳栓をすることによって被験者は自分の声を骨伝導でフィードバックできるため、調音の歪みの軽減が期待できる。これに加えて、この手法には大きな撮像ノイズによる心的ストレスを軽減する効果もある。

ガイド音の呈示には、非磁性体で構成され、かつ十分な音量が得られるピエゾスピーカ (栄進電機 PZT02) が用いられている。撮像の際には、被験者の鎖骨付近にこのピエゾスピーカを両面テープで接着し、アンプ (SONY TA-F37R) で駆動した。

2.3 被験者およびタスク

被験者は成人男性 1 名である。タスクは日本語 5 母音 /a/、/e/、/i/、/o/、/u/ の持続発声である。被験者はガイド音の基本周波数 (120 Hz) に声の高さを合わせ、かつガイド音にタイミングを合わせて繰り返し発声を行った。

2.4 撮像条件

撮像に用いられた MR 装置は、ATR-Promotions 脳活動イメージングセンタ (BAIC) に設置された島津 Marconi 社製 MAGNEX ECLIPSE 1.5T Power Drive 250 である。撮像にはアトラスアレイコイル (頭頸部・胸部用コイル) を用いた。シーケンスは Fast Spin Echo 法、TE = 9 ms、TR = 4900 ms、FA = 90 度。撮像領域は 256 × 256 mm、分解能は 512 × 512 pixels である。スライスは矢状方向で、厚み 2 mm、間隔 2 mm、41 スライスで撮像した。加算回数は 1 回で、撮像時間は 1 母音あたり約 6 分である。

3. 声道模型の作製方法

まず、MR 画像を補間して 0.5 × 0.5 × 0.5 mm/voxel のボリュームデータに変換した。このボリュームデータに、事前に作成していた歯列のボリュームデータを Takemoto ら [15] による方法で補填した^(注2)。歯列のボリュームデータは、被験者に

(注2): ATR 母音発話 MRI データには、この歯列補填済み画像データが含ま



図 1 母音/a/の声道模型の写真。

造影剤を口にふくませて撮像した MRI データから得たものである。

次に、閾値処理に目視による修正を加えて声道領域を抽出した。この被験者は母音/a/発声時に主声道と鼻咽腔が連結するが、この鼻咽腔開口部もふさいだ。

抽出した声道領域を Materialise 社の 3 次元画像編集ソフトウェア MIMICS で STL 形式に変換し、次に同社の STL 編集ソフトウェア MAGICS を用いてスムージングを施し、3 mm 厚の壁をつけた。さらに、上下約 30 mm の口唇部分を残して顔部分を除去した。そして、光造形装置によって図 1 に示すような声道模型を作製した^(注3)。この声道模型はエポキシ系樹脂の一種でできており、造形の精度は 0.2 mm 程度である。

4. 音響測定方法

音響測定系の構成を図 2 に示す。声道模型はその口唇部がホーンドライバーユニットの出力部に正対するように設置した。声道模型の声門部を直径 1.2 mm の穴の開いたプラスチック板 (B&K UA0929) でふさぎ、その穴にプローブマイク (B&K 4182) のプローブを挿入した。これは、声門部の音響インピーダンスを高くすることにより、声門開放が声道音響特性に与える影響 [6] [7] [8] を抑えるためである [17]。

声道音響特性の計測にあたっては、アンプ (Accuphase E-305) を介してホーンドライバーユニット (エール音響研究所 7550DE) から音響測定系の音響特性を補正した Optimized Aoshima's time-stretched pulse (OATSP) 信号 [9] を出力した。そして、声門部における音圧をアンプ (B&K 5935) を介してソリッドステートレコーダ (Marantz PMD-670) に標準化周波数 48 kHz、量子化 16 bit にて保存した。つまり、声道模型の口唇部から入った信号が声道模型内を通り声門部で収録されることになる。補正済み OATSP 信号に対する応答を 10 回加算平均し、声道模型の音響特性を求めた。なお、測定は気温 25 °C の無響室にて行った。

声道模型の音響特性の測定に先立ち、音響測定系の音響特性

れている

(注3): 母音/a/の声道模型は、(株) ATR-Promotions にて販売されている。

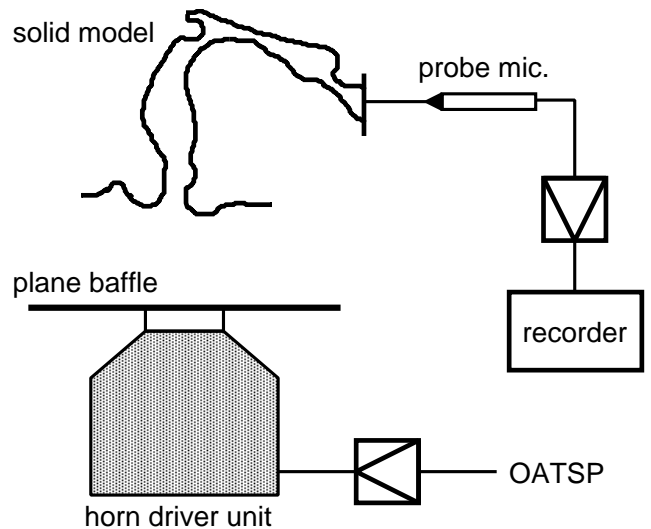


図 2 音響測定系の構成。

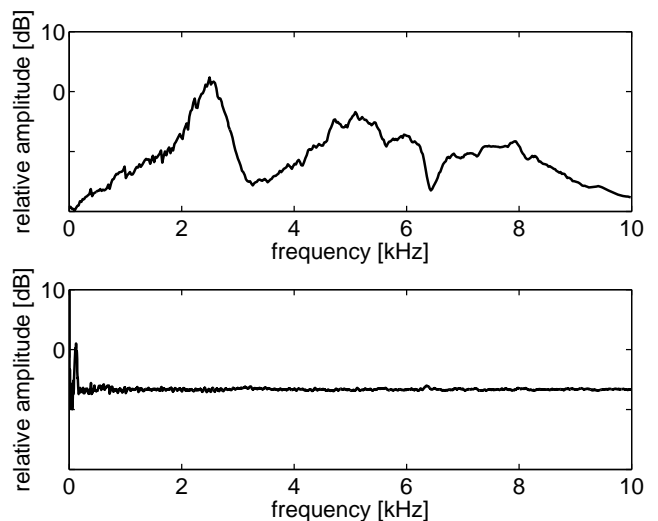


図 3 (上) 音響測定系の音響特性。(下) 補正後の音響測定系の音響特性。

を補正した。まず、声道模型がない状態でホーンドライバーユニットから OATSP 信号を出力して収録した。音響測定系の音響特性を図 3(上) に示す。音響測定系の音響特性の逆フィルタを OATSP 信号にかけ、音響特性を補正した OATSP 信号を得た。この補正済みの OATSP 信号を用いることによって、図 3(下) の音響特性が得られた。この図に見られるように、補正済みの OATSP 信号を用いても 160 Hz 以下にピークが見られるため、本研究では音響測定系の音響特性がほぼ平坦な 160 Hz 以上の測定結果を示す。

5. 結果と考察

音響測定により得られた声道模型の音響特性を図 4 に示す。共鳴によるピークのみならず、梨状窩などに由来する反共鳴による谷も現れている。この図の中には音響特性の主要なピークと谷にその周波数を示している。これらのうち第 1、第 2、第 3、第 4 共鳴の周波数を表 1 にまとめた。母音/i/の第 1 共鳴周波数は、音響測定系の音響特性が平坦でない周波数帯域 (160

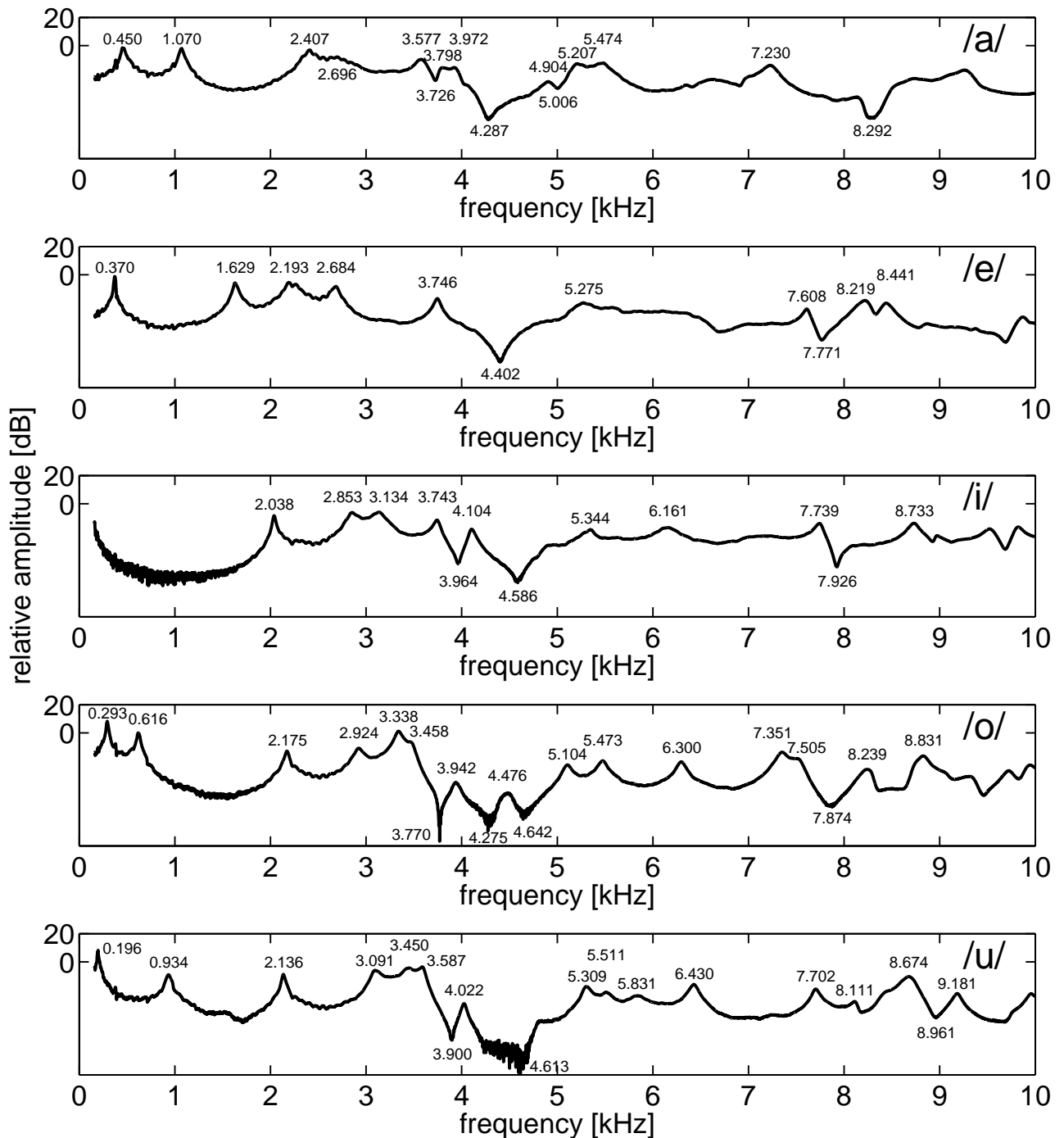


図4 日本語5母音の声道模型の音響特性．上から母音/a/、/e/、/i/、/o/、/u/．主要なピークと谷にその周波数 [kHz] を示している．

Hz 以下) に含まれるため、測定できなかった。

表1の共鳴周波数は、一般的な母音のホルマント周波数よりも低い。この傾向は低周波数領域ほど顕著である。これは声道模型の壁が剛壁であることに起因する。音響管の壁の硬度が上がると、その共鳴周波数は低周波数方向へシフトし、その影響は低周波数領域ほど大きいからである[18]。また、図4の音響特性が一般的な母音のスペクトルと傾斜が異なるのは、後者には声帯音源の周波数特性による -6 dB/octave の傾斜が含まれるためである。

以下では、同じ声道形状データに基づく有限要素法による声

道音響解析の結果[2][19][17]との比較により、図4に示した音響特性のピークと谷の由来を説明する。この声道音響解析においては、FEM解析プログラムの制約から声道内を $2 \times 2 \times 2$ mmの要素で分割し、放射領域を $2 \times 2 \times 4$ mmの要素で分割した。また、声道壁は剛壁とした。

まず、各母音の第4共鳴は喉頭腔に由来する。これは、有限要素法による解析において、声道から喉頭腔を除去することにより第4共鳴が消失することから明らかである[19]。

次に、声道下部の左右に存在する分岐管である梨状窩によって生じる谷について説明する。これは、声道の有限要素モデル

表 1 声道模型の音響特性における第 1, 第 2, 第 3, 第 4 共鳴周波数 (R1, R2, R3, R4) [Hz]. 音響特性からピークピッキングにより求めた.

vowel	R1	R2	R3	R4
/a/	450	1070	2407	2696
/e/	370	1629	2193	2684
/i/	—	2038	2853	3134
/o/	293	616	2175	2924
/u/	196	934	2136	3091

から梨状窩を除去した実験によって示されたものである。まず、母音/a/の声道模型の音響特性においては、3726 Hz と 4287 Hz の谷が梨状窩によるものである。母音/e/においては、4402 Hz の谷が梨状窩によるものである。母音/e/の音響解析では、梨状窩によって 2 つの谷が現れていたが、3.8 kHz 付近に現れる谷のバンド幅が小さかった。今回の音響測定ではこの谷が明確に現れなかったものと考えている。母音/i/では、3964 Hz と 4586 Hz の谷が梨状窩によるものである。母音/o/では、3770 Hz と 4275 Hz の谷が梨状窩によるもので、4642 Hz の谷は声道内の横モードによるものと考えている。最後に、母音/u/においては、3900 Hz と 4613 Hz の谷が梨状窩によるものである。

梨状窩は声道における分岐管であるにもかかわらず、ピークも生じさせる。梨状窩の下端から口唇までを結ぶ経路に沿って定在波が生じるためである [17]。まず、母音/a/の声道模型の音響特性においては、5207 Hz と 5474 Hz のピークが梨状窩によるものである。母音/e/では 5275 Hz のピーク、母音/i/では 5344 Hz のピークが梨状窩によるものである。これらの母音において、梨状窩によるピークが 1 つしか生じない理由は明らかになっていない。母音/o/においては、梨状窩によって 5104 Hz と 5473 Hz のピークが出現している。最後に、母音/u/においては、5309 Hz と 5511 Hz のピークが梨状窩によるものである。

6. おわりに

本研究では、ATR 母音発話 MRI データに収録された MRI データから光造形によって日本語 5 母音の声道模型を作成し、その音響特性を測定した。声道模型の声門部を高インピーダンスにして音響特性を測定する手法を用いることによって、声道模型の喉頭腔共鳴を含めた音響特性を得ることができた。ATR 母音発話 MRI データは研究・教育目的に公開されているので、本研究の測定結果は各種の声道音響解析手法の精度検証に用いることができる。シミュレータのプログラムのデバッグにも有効である。声道断面積関数の抽出方法の検証にも利用することができる。

ただし、実際の声道音響特性と異なる部分があることに注意が必要である。まず、今回対象にした声道模型では声道壁が剛壁である。また、声道模型の口唇部は上下約 30 mm しかなく人間の放射特性とは異なる。この他、本研究の被験者は母音/a/発話時に鼻咽腔結合が生じているが、声道模型では鼻咽腔開口部をふさいでいる。さらに、MRI 観測、声道領域抽出、音響

測定の段階の誤差も含まれる。本研究の測定結果の利用にあたっては、これらの点の認識が不可欠である。

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「音声の中の非言語情報の生成・知覚の特性解析と多言語間コミュニケーションへの応用 (071705001)」の援助を受けた。

本研究で使用した MRI データは、ATR 人間情報科学研究所が独立行政法人情報通信研究機構からの研究委託「人間情報コミュニケーションの研究開発」に基づいて収録し、公表した『ATR 母音発話 MRI データ』の一部である。本データの使用および成果の発表は、株式会社 ATR-Promotions との使用許諾契約に基づいている。

また、実験にご協力いただいた西本博則氏 (現 ATR-Promotions) と藤田覚氏 (現 北陸先端科学技術大学院大学)、ご助言いただいた足立整治氏 (現 Fraunhofer institute for building physics) と一ノ瀬裕先生 (九州産業大学) に感謝します。あわせて、無響室の利用を許可していただいた ATR 認知情報科学研究所 (井ノ上直己所長)、一部の音響機器を貸与していただいた尾崎正和氏 (ブリュエル・ケアール・ジャパン) に感謝します。

文 献

- [1] 北村, 正木, “MRI 観測を基礎にした音声生成系研究の進展,” 音響誌, 62(5), 385–390, 2006.
- [2] 北村, 竹本, 本多, “ボクセルメッシュを用いた有限要素法による声道の音響解析,” 信学技報 (音声), 104(470), 41–46 (2004).
- [3] ATR-Promotions, “ATR 母音発話 MRI データ CD-ROM 説明書 第 4 版,” 2007.
- [4] 北村, 西本, 藤田, 本多, “音響計測, 電気回路理論および有限要素法により求めた声道模型の伝達関数の比較,” 信学技報 (音声), 103(26), 37–42, 2003.
- [5] Kitamura, Fujita, Honda, Nishimoto, “An experimental method for measuring transfer functions of acoustic tubes,” *Proc. ICSLP2004*, TuB602, 2004.
- [6] 本多, 北村, 竹本, 能田, 平田, 正木, 藤田, “母音発話 MRI データに基づく声道模型の製作と音響分析,” 音講論 (秋), 313–314, 2005.
- [7] Takemoto, Adachi, Kitamura, Mokhtari, Honda, “Acoustic roles of the laryngeal cavity in vocal tract resonance,” *J. Acoust. Soc. Am.*, 120(4), 2228–2238, 2006.
- [8] Kitamura, Takemoto, Adachi, Mokhtari, Honda, “Cyclicity of laryngeal cavity resonance due to vocal fold vibration,” *J. Acoust. Soc. Am.*, 120(4), 2239–2249, 2006.
- [9] Suzuki, Asano, Kim, Sone, “An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses,” *J. Acoust. Soc. Am.*, 97(2), 1119–1123, 1995.
- [10] Nota, Kitamura, Takemoto, Hirata, Honda, Shimada, Fujimoto, Syakudo, Masaki, “A bone-conduction system for auditory stimulation in MRI,” *Acoust. Sci. & Tech.*, 28(1), 33–38, 2007.
- [11] Masaki, Tiede, Honda, “MRI-based speech production study using a synchronized sampling method,” *J. Acoust. Soc. Jpn. (E)*, 20(5), 375–379, 1999.
- [12] 高野, 本多, 正木, 島田, 藤本, “MRI コイルと外部トリガ同期撮像法による発声動作の高分解能撮像,” 音講論 (春), 291–292, 2003.
- [13] 藤本, 赤土, 島田, “外部トリガを用いた MRI 同期撮像法,” 第 20 回放射線技師総合学会 CONGRESS BOOK, 252, 2004.
- [14] 北村, 正木, 島田, 藤本, 赤土, 本多, “光マイクロフォンを用いた MRI 撮像時の騒音測定,” 音響誌, 62(5), 306–315 (2006).
- [15] Takemoto, Kitamura, Nishimoto, Honda, “A method of

- tooth superimposition on MRI data for accurate measurement of vocal tract shape and dimensions,” *Acoust. Sci. & Tech.*, 25(6), 468–474, 2004.
- [16] Fujita, Honda, “An experimental study of acoustic characteristics of hypopharyngeal cavities using vocal tract solid models,” *Acoust. Sci. & Tech.*, 26(4), 353–357, 2005.
- [17] 北村, 竹本, 本多, “梨状窩によって生じる声道伝達特性の極零パターン,” 音講論 (春), 357–358, 2006.
- [18] Rabiner, Schafer, *Digital processing of speech signals*, Chap. 3, Prentice Hall, 1978.
- [19] 北村, 竹本, 本多, “声道伝達特性における喉頭腔の影響,” 音講論 (春), 275–276, 2005.