MRIにより計測した声道長と音声データとの関係*

☆波多野 博顕 (神戸大), 北村 達也 (甲南大), 竹本 浩典, パーハム モクタリ (NICT),
本多 清志 (パリ第 3 大学), 正木 信夫 (ATR/ATR-Promotions)

1 はじめに

声道は音声の生成にとって根本的で重要な性質を持つ．声道の持つ形態的特徴は共鳴特性に影響し，音響的特徴としてフォルマントとなって現れる [1]．言語音の生成には，形態と音響が密接な関係で結びついている．

近年, 調音器官の形態観察に MRI を利用した多くの研究によって，各器官の発話中の動態が次第に明らかになってきている．また，撮像技術や分析法の進展に伴い，これまでより詳細な情報が得られるようになった [2]．とりわけ，Fitch & Giedd [3] では 2〜25 才の 129 名から採取された大規模な MRI データに基づいて声道長（Vocal-Tract Length, VTL）の詳細な計測を行い，身長・体重との相関や成長に伴う形状変化，性差による違いを明らかにしている．しかし，彼らの MRI データは発話中に撮像されたものではなく，従って声道長と音声との関連は示されていない．

最近，ATR-Promotions 脳活動イメージングセンタより, 日本人成人男性 15 名の母音発話時の MRI データベースが公開された. このデータベースには，MRI データから計測された声道長や，撮像と同時に録音された音声データも収録されている．このデータベースは音声生成の研究のみならず，音声認識技術における話者正規化や，音声合成技術における声質変換の研究などにも活用できる貴重なデータである．本研究では，この MRI データベースを対象にして，声道長と音声の関連に着目した分析を行った。

2 材料と方法

2.1 被験者および分析対象発話

被験者は日本人成人男性 15 名（年齢範囲 24 歳〜55 歳．平均 37.2 歳．標準偏差 9.1 歳）であった (Table 1)．データには各被験者の身長も含まれる（身長範囲 162 cm〜184 cm．平均 172.9 cm．標準偏差 5.6 cm)．

Table 1 Subjects' age and body height [cm].

ID	Age	Body height
M01	29	169
M02	29	175
M03	30	171
M04	34	178
M05	36	176
M06	38	177
M07	38	175
M08	47	175
M09	52	165
M10	55	168
M11	24	175
M12	40	162
M13	27	184
M14	44	169
M15	35	175

分析対象発話は，日本語 5 母音（/a/, /e/, /i/, /o/, /u/）の単独発話である．収録は各母音につき 3 回行っている．

2.2 MRI 撮像および声道長抽出

MRI の撮像は ATR-Promotions 脳活動イメージングセンタに設置されている MRI 装置（ Shimadzu-Marconi ECLIPSE 1.5T Power Drive 250) を用いた. 撮像パラメータは, FOV 256 x 256 mm，image size 512 x 512 pixels（1 ピクセルは 0.5 x 0.5 mm)，TE=3.9 ms，TR=15 ms, スライス厚 5 mm, 加算回数 2 回である．撮像したのは正中矢状断面 (1 断面) のみで, 1 回の撮像に要した時間は約 5 秒である．被験者はこの間，当該の母音を一息で発話し，

*Relationship between speech data and vocal tract length measured by magnetic resonance imaging, by HATANO, Hiroaki (Kobe Univ.), KITAMURA, Tatsuya (Konan Univ.), TAKEMOTO, Hironori, MOKHTARI, Parham (NICT), HONDA, Kiyoshi (University of Paris III) and MASAKI, Shinobu (ATR/ATR-Promotions).

もし撮像中に息が切れた場合には，口の構えを維持するよう指示された．なお，MRI 装置の制約から被験者は仰臥位にて発話している．

撮像された各 MRI データから，Takemoto *et al.* [4] によるアルゴリズムで声道幅が抽出された（撮像されたのが正中矢状断面のみのため声道断面積は抽出できない）．このアルゴリズムでは，まず目視にて MRI データ上の声門の位置を決定する．次に声道領域内に声門からの距離に関するコンターマップを作り，それに基づいて声道中心線を引く．声道長は 2.5 mm 単位で計測された．

発話中の口蓋帆挙上の有無も目視により確認した．

2.3 音声およびフォルマント周波数計測

一部の被験者については，MRI 撮像と同時に音声も収録された．収録にはレコーダ（Marants PMD-670）と，MRI 装置内でも使用可能な光マイクロフォン（Phone-Or SOM）が用いられた．このデータの標本化周波数は 16 kHz で，量子化ビット数は 16 bit である．

この MRI データベースには，静かなオフィス環境で録音された日本語 5 母音も収録されている．この録音にはコンデンサマイクロフォン（Audio-technica AT9820X）とノート PC が用いられた．標本化周波数は 16 kHz で，量子化ビット数は 16 bit である．

MRI 装置内部では撮像中に大きなバックグラウンドノイズが発生するため，撮像が始まる前の僅かな時間（約 1 秒間）に発声された音声を切り出してフォルマント周波数の計測に用いた．フォルマント周波数は 3 回の撮像時の音声全てでもとめ，その平均値を結果とした．

フォルマント周波数の抽出には対数スペクトルの不偏推定[5]を利用した。分析窓はハニング窓，フレーム長 64 ms，フレーム周期 16 ms，繰り返し 3 回，ケプストラム次数 60 次でスペクトル包絡を求め，フレーム間で加算平均した．得られた平均スペクトル包絡から目視にて第 1 から第 4 フォルマント周波数を計測した。

Table 2 Average of vocal-tract length [cm]. Each column shows average of each vowel utterances.

ID	/a/	/e/	/i/	/o/	/u/	mean
M01	16.3	15.5	16.0	17.2	16.7	16.3
M02	16.2	15.3	15.8	17.3	17.2	16.4
M03	16.5	15.8	16.5	17.3	17.0	16.6
M04	17.3	16.2	17.3	18.5	17.9	17.4
M05	17.7	16.9	17.6	18.4	18.8	17.9
M06	16.9	16.3	16.8	17.3	17.0	16.9
M07	16.8	16.1	16.3	18.4	17.2	17.0
M08	17.1	16.4	16.8	17.3	17.0	16.9
M09	16.7	16.0	16.2	17.5	17.0	16.7
M10	18.3	17.5	18.0	19.3	18.7	18.4
M11	15.8	15.6	16.1	16.8	16.8	16.2
M12	16.8	16.3	16.6	17.8	16.8	16.8
M13	17.0	16.6	16.9	18.0	18.0	17.3
M14	16.8	16.4	16.3	17.4	17.9	17.0
M15	16.1	15.5	15.8	17.7	17.4	16.5

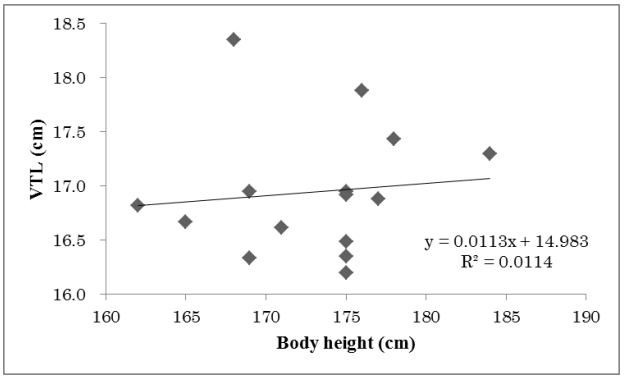

Fig.1 Correlations between VTL and body height.

3 結果

3.1 声道長

各母音発話時の声道長を示す（Table 2, 単位は cm）．各 3 回の母音発話時の声道長を平均した数値である．また，被験者の身長と声道長の相関を求めたところ，相関はみられなかった（r^2=0.011）（Fig. 1）．

3.2 フォルマント周波数

撮像中に収録した母音/a/のフォルマント周波数（F1, F2, F3, F4）を口蓋帆の咽頭壁への接触の有無と共に示す（Table 3）．被験者 15 名のうち，撮像中の音声が収録されたものは 12 名であった．口蓋帆の接触－非接触は，接触を「○」で，非接触を「×」で示した．各典型例を Fig. 2 に示す．

Table 3 Existence or non-existence of velum contact with pharyngeal wall and formant frequencies (F1, F2, F3, F4) [Hz] of Japanese vowel /a/ uttered during scanning.

ID	Velum contact	F1	F2	F3	F4
M01	×	599.0	869.8	2776.0	3083.3
M02	○	609.4	1041.6	2400.7	2766.0
M03	○	604.2	1119.7	2677.0	3385.3
M04	×	713.5	1062.7	2359.3	2750.0
M05	○	578.1	984.5	2828.3	3333.3
M06	○	645.8	1182.7	2479.3	3062.7
M07	○	604.2	960.0	2359.3	2885.7
M08	×	578.1	1010.4	2296.7	2620.0
M09	○	656.3	1135.7	2879.7	3182.3
M10	×	609.4	1021.0	2526.0	2896.0
M11	×	666.7	1125.0	2791.7	3302.3
M12	×	687.5	1000.2	2724.0	3130.3

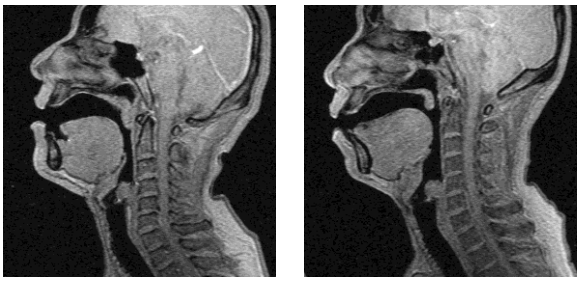

Fig. 2 Typical examples of velum contact with pharyngeal wall (left) and noncontact (right)

Table 4 Formant frequencies (F1, F2, F3, F4) [Hz] of /a/ uttered at quiet office environment.

	F1	F2	F3	F4
M01	807.3	1145.7	3250.0	4038.0
M02	677.1	1125.0	2449.7	3177.3
M03	630.2	1161.3	2916.7	3760.7
M04	812.2	1099.0	2255.0	3583.7
M05	619.8	1068.0	2875.0	3474.0
M06	692.7	1260.7	2677.0	3526.0
M07	541.7	994.8	3088.7	3781.3
M08	697.9	1135.3	2677.0	3573.0
M09	645.8	1146.0	2916.7	3432.0
M10	619.8	1005.3	2463.3	3338.7

口蓋帆の接触－非接触によって全体を 2 群に分け, 各フォルマント値について *t* 検定（等分散を仮定した 2 標本による検定）を行った. 結果は，いずれも有意差はみられなかった（ F1: $t(10)=1.03$, $p=0.33$; F2: $t(10)=1.11$, $p=0.29$; F3: $t(10)=0.2$, $p=0.85$; F4: $t(10)=0.96$, $p=0.36$）.

静かなオフィス環境で収録した母音/a/のフォルマント周波数を Table 4 に示す（被験者 M11, 12 は未収録）.

収録環境（撮像時とオフィス時）の相違による被験者 M01～M10 の母音/a/のフォルマント周波数について，*t* 検定（一対の標本による平均の検定）を行った．結果は全てのフォルマント周波数で，オフィス時のものが撮像時のものより有意に高かった.（F1: $t(9)=2.3$, $p<0.05$; F2: $t(9)=2.93$, $p<0.05$; F3: $t(9)=2.39$, $p<0.05$; F4: $t(9)=5.89$, $p<0.05$）.

3.3 声道長とフォルマントの相関

声道長と撮像中に収録した母音/a/の各フォルマント周波数の値を Figs. 3～6 に示す．声道長とフォルマント周波数の相関をもとめたところ，いずれも相関はみられなかった（F1: $r^2=0.03$; F2:$r^2=0.02$; F3: $r^2=0.04$; F4: $r^2=0.05$）.

4　考察

Fitch & Giedd [3] では身長と声道長に強い相関があることが示されているが（$r^2=0.86$），本研究ではこのような相関関係が得られなかった（$r^2=0.011$）．両者は被験者の年齢範囲に大きな相違がある．前者では成長に伴った声道長の形態的発達に注目しているため，年齢範囲が大きい（2 歳～25 歳）．一方，本研究では 24 歳～55 歳であるため，声道を含む諸調音器官の発達が終了していると思われる．そのため，少なくとも身体発達が成熟した成人以上の段階では，身長と声道長に相関はみられないといえる.

また，少なくとも母音/a/に関しては，声道長とフォルマント周波数にほとんど相関がない（F1: $r^2=0.03$; F2: $r^2=0.02$; F3: $r^2=0.04$; F4: $r^2=0.05$）．このことから，フォルマント周波数の個人差は声道長ではなく，声道の形状の個人差に主に起因することが示唆される．北村ら [6]は男性 8 名の母音/i/と/e/の声道断面積関数から声道伝達関数を求めた. その結果, 声道長を正規化しても共鳴周波数には大きな個人差が見られた．この結果は上の指摘を支持している.

撮像時とオフィス環境で収録された音声では，フォルマント周波数に大きな差がみられた. 母音/a/の F1 から F4 までの全てにおいて，

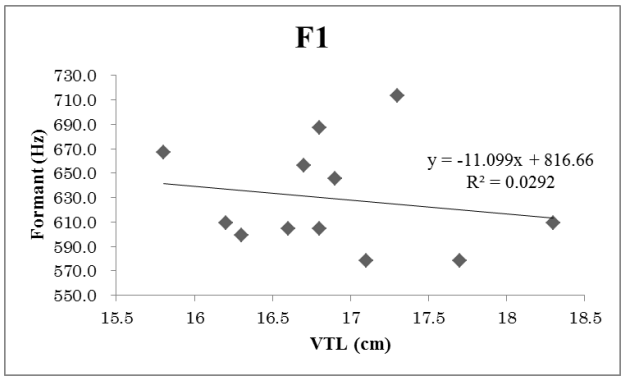

Fig. 3 Correlations between VTL and F1 of /a/.

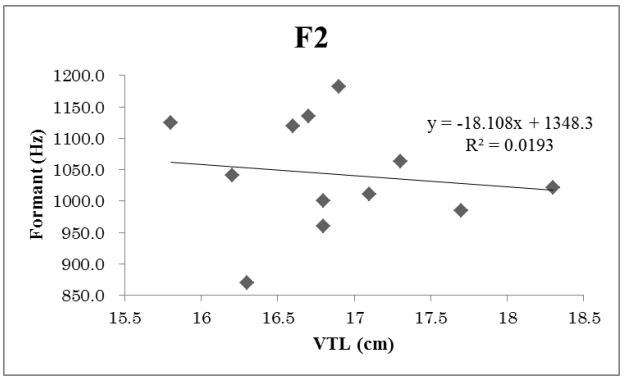

Fig. 4 Correlations between VTL and F2 of /a/.

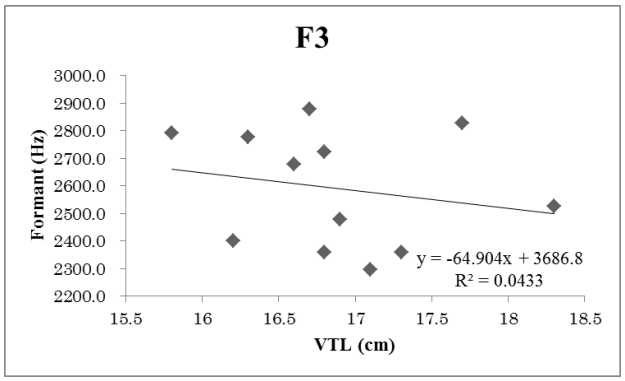

Fig. 5 Correlations between VTL and F3 of /a/.

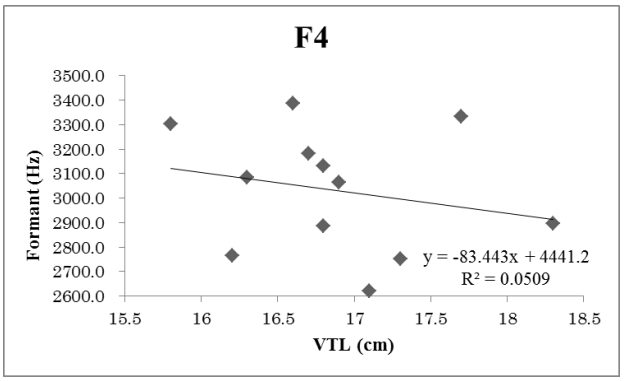

Fig. 6 Correlations between VTL and F4 of /a/.

オフィス環境で収録されたもののフォルマント周波数が有意に高かった（撮像時平均; F1: 619.8 Hz; F2: 1038.8 Hz; F3: 2558.2 Hz; F4: 2996.5 Hz, オフィス時平均; F1: 674.4 Hz; F2: 1114.1 Hz; F3: 2756.9 Hz; F4: 3568.5 Hz）．この差は，撮像時は仰臥位でオフィス時は座位という, 発話時の姿勢に起因する. Kitamura *et al*. [7]では，開放型 MRI 装置を用いて母音発話時の姿勢に起因する調音器官の相違を調べている．その結果，仰臥位では各器官が重力からの影響を受けることが明らかにされている．このような形態変化がフォルマント周波数の相違を引き起こしたと考えられる．本データベースを使用するには，このような点を十分認識する必要がある．

5　おわりに

本研究では，母音/a/を対象にして声道長と音声の関連について分析した．今後，他の母音にも同様の分析を行う予定である．また，MRI データから喉頭腔，咽頭腔，口腔の長さをもとめ，各フォルマント周波数との相関を明らかにする．

謝辞

本研究は，平成 23 年度科研費（21300071, 21500184）によりおこなわれた．ここに記して深謝する．

参考文献

[1] Chiba and Kajiyama, *The Vowel: Its Nature and Structure*, Tokyo-Kaiseikan, 1941.
[2] 鏑木ら，*音声生成の計算モデルと可視化*, コロナ社, 2010.
[3] Fitch and Giedd, JASA, 106 (3), 1511-1522, 1999.
[4] Takemoto. *et al*., JA SA, 119 (2), 1037-1049, 2006.
[5] 今井，古市，信学論 A, 70 (3), 471-480, 1987.
[6] 北村ら，音講論（春）, 285-286, 2004.
[7] Kitamura. *et al*., Acoustic. Sci. Tech., 26 (5), 465-468, 2005.