

# Wave を用いた歌唱中の舌と口唇の動きの分析

## Analysis of Tongue and Lip Movements During Singing Using the Wave Speech Research System

高橋 純<sup>†</sup> 志々目 樹<sup>‡</sup> 竹本 浩典<sup>‡</sup> 能田 由紀子<sup>††</sup>

Jun TAKAHASHI<sup>†</sup> Itsuki SHISHIME<sup>‡</sup> Hironori TAKEMOTO<sup>‡</sup> Yukiko NOTA<sup>††</sup>

<sup>†</sup> 大阪芸術大学短期大学部 〒546-0023 大阪府大阪市東住吉区矢田 2-14-19

<sup>‡</sup> 千葉工業大学 〒275-0016 千葉県習志野市津田沼 2-17-1

<sup>††</sup> 国立国語研究所 〒190-8561 東京都立川市緑町 10-2

<sup>†</sup> Osaka University of Arts Junior College 2-14-19 Yada, Higashisumiyoshi-ku, Osaka-shi, Osaka 546-0023 Japan

<sup>‡</sup> Chiba Institute of Technology 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan

<sup>††</sup> National Institute for Japanese Language and Linguistics 10-2 Midori-cho, Tachikawa, Tokyo 190-8561, Japan

E-mail: <sup>†</sup> jun\_takahashi@osaka-geidai.ac.jp, <sup>‡</sup> hironori.takemoto@p.chibakoudai.jp,

<sup>††</sup> ynota@ninjal.ac.jp

**あらまし** 本研究では1名のプロの歌手と歌唱訓練1年程度の一般大学生1名の歌唱技術の違いを検討した。プロと学生が歌唱中の音声からSF占有率（全帯域に対する2-4 kHz帯域のエネルギーの割合）を計算し、同時に舌・口唇の動きを計測した。その結果、全てのタスクにおいてプロは学生よりSF占有率が高く安定していた。そして、プロは母音の変化は主に舌で制御し、音高の変化は下顎を徐々に変位させて高音に応じた口の構えを作ることで制御していた。一方、学生は母音の変化を舌、口角および下顎の動きによって制御し、音高の変化ではプロに見られるような制御が見られなかった。これらの舌や口唇の制御の差がSF占有率の差となったと考えられる。

**キーワード** オペラ, 歌唱, 歌声, 歌い手のフォルマント, EMA, Wave

### 1. はじめに

男性のオペラ歌手では歌声のスペクトルの約3 kHz付近に「歌い手のフォルマント Singer' formant (以下SF)」と呼ばれる顕著な高まり[1]が存在することが知られている。これまで筆者は、音声スペクトルの全帯域に対してこのSFを中心とする2-4 kHzの帯域のエネルギーが占める割合をSF占有率と定義してSF占有率と歌声の音楽的な評価の関係を検討してきた。プロの歌手と音楽大学で声楽を学ぶ学生の計18名が

音高や母音を変えて歌唱した際のSF占有率と評価を分析した。その結果、一般的にプロは学生に比べてSF占有率が高く、音高や母音が変わってもSF占有率の変化が少なかった。そして、そのような特徴を持つほど歌声の音楽的な評価が高かった[2][3]。

この結果を踏まえて、まずSFの生成機序を検討するために、プロと学生が歌唱中の声道形状をリアルタイム磁気共鳴画像法 (rtMRI: real-time Magnetic Resonance Imaging) を用いて観測した。その結果、プ

口は学生に比べて喉頭の位置が低く，咽頭が拡大していた[4].これは，SFの生成には喉頭に比べて咽頭が十分に広い必要がある[1]という知見を裏付けていた.

次に，音高や母音が変化してもプロは学生に比べて SF 占有率が安定している要因を rtMRI データで検討しようと試みた.しかし，rtMRI は空間・時間的な分解能が低く，舌や口唇の大まかな形状や動きは観察できるが，特定の部位の変位を正確に検討することはできなかった.

そこで本研究では，特定の部位の変位をより高い時間・空間分解能で計測できる磁気センサシステムの 1 種である Northern Digital Inc. の Wave speech research system(以下 Wave)を導入した.そしてこの装置を用いて，プロと学生が母音と音高を変化させて歌唱している際の舌と口唇の動きを計測し，プロで SF 占有率が高く安定している要因について基礎的な検討を行ったので報告する.

## 2. 実験方法

### 2.1. 実験参加者

歌声の収録と Wave を用いた舌と口唇の計測を行う被験者として，国内外でプロの歌手として活躍するテノール(pTen) 1 名と，歌唱訓練 1 年程度の一般大学生バリトン(aBar) 1 名の合計 2 名を選定した.

### 2.2. 歌唱課題

歌唱中の母音と音高の変化に伴う調音運動を検討するため，以下の歌唱課題を作成した. Fig. 1 に pTen の歌唱課題を示す. Task1-1 は 3 音全てが同一の/a/母音と音高であり，Task1-2 は 3 音の音高は一定で母音のみ a→i→a と変化する. Task2-1 は，母音は一定で，音高が 1 音目から 2 音目に完全 5 度

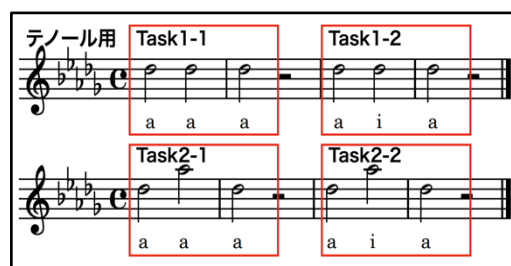


Fig. 1 音高と母音の変化を含む歌唱課題

上行跳躍し，さらに 2 音目から 3 音目に完全 5 度下行跳躍する. Task2-2 は母音が a→i→a と変出し，音高も先述した完全 5 度超躍進する. このうち，音高が一定で母音が変化する Task1-2, 音高が変化するが母音が一定の Task2-1, および音高も母音も変化する Task2-2 の音声と調音運動を分析した. なお，また，歌唱技術の差が顕著に現れるように Task2 の最高音は声区のブレイク (換声点: 地声と裏声が切り替わる音域) よりも高い音高に設定した[5]. また，声域が異なる aBar は 1 音目を pTen より短 3 度下の B $\flat$ 3 とした.

### 2.3. 音声分析

収録された歌声データから，フレーム間隔 10 ms で離散フーリエ変換によってパワースペクトルを求めた. その後，各フレームにおいて 2-4 kHz の周波数帯域のパワーの総和と全帯域のパワーの総和の比を計算して百分率としたものを SF 占有率とし，その時系列データに移動平均を適用して平滑化した. 理論上，この値が 100% の場合，全ての音響エネルギーが 2-4 kHz の帯域に集中していることを意味する.

### 2.4. 調音運動の計測と分析

Wave は，磁気を利用して調音器官に貼り付けた小型のセンサーの時時刻刻の位置を計測する装置である. 得られるデータは時系列に沿ったセンサーの 3 次元の座標値で

あり，取り扱いが容易である．さらに近年，能田ら[6]によってセンサーの改良やバイトプレートの開発が行われ，口蓋形状と合わせたセンサー位置の検討が可能となった．被験者は，舌尖（T1），右口角（RL），下顎（LI）に Wave のセンサーを装着し，国立国語研究所の防音室で歌唱課題を実施した．センサーの座標は 100 Hz のフレームレートで記録し，音声は Sony 製のコンデンサーマイクロホン C-357（サンプリング周波数 22 kHz，量子化ビット数 16 bit）で録音した．本研究では，計測開始時における各センサーの座標を基準位置とし，基準位置から各時刻におけるセンサーの距離を計算してその変化を分析した．また，Wave では口蓋形状の計測ができないため，上顎の歯列と口蓋の形状を歯科用印象材で採取し，光学スキャナで計測した．その後，この形状を Wave の座標系に写像した．

### 3. 結果と考察

表 1 は，各 Task の SF 占有率である．全ての Task において，pTen の SF 占有率が高かった．pTen の平均値が 23.9%，aBar が 22.3%であり，t 検定の結果有意な差が見られた( $t(4) = 3.07$ ,  $p = .042$ )．この結果は先行研究[2]と一致した．

表 1 pTen と aBar の SF 占有率

pTen	Task1-2	aia	24.6%
	Task2-1	aaa	23.5%
	Task2-2	aia	23.6%
aBar	Task1-2	aia	22.9%
	Task2-1	aaa	21.5%
	Task2-2	aia	22.5%

Fig. 2 と Fig. 3 は，Task1-2 における pTen と aBar の SF 占有率および基本周波数(fo)，

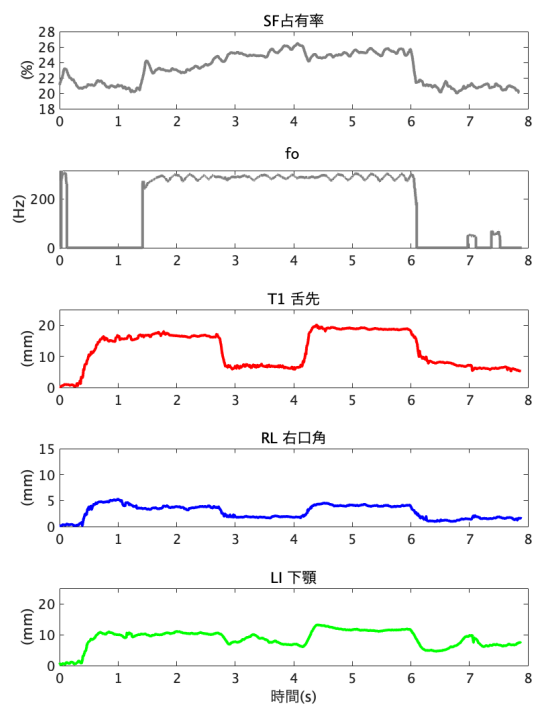


Fig. 2 Task1-2 の a→i→a pTen

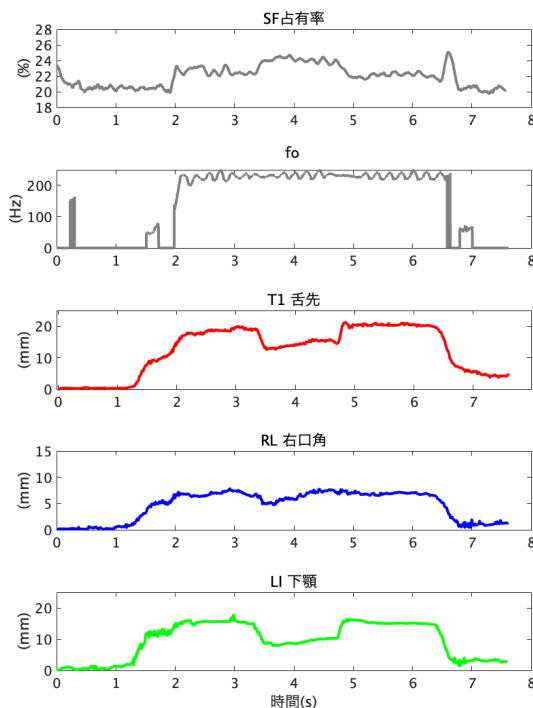


Fig. 3 Task1-2 a→i→a ABar

そして各センサー（T1: 赤, RL: 青, LI: 緑）の時間変化を示している．まず，両者の SF 占有率に着目すると，pTen の値は平均 24.6%で aBar の 22.9%に比べて高く，また pTen は母音間での SF 占有率の変化が少な

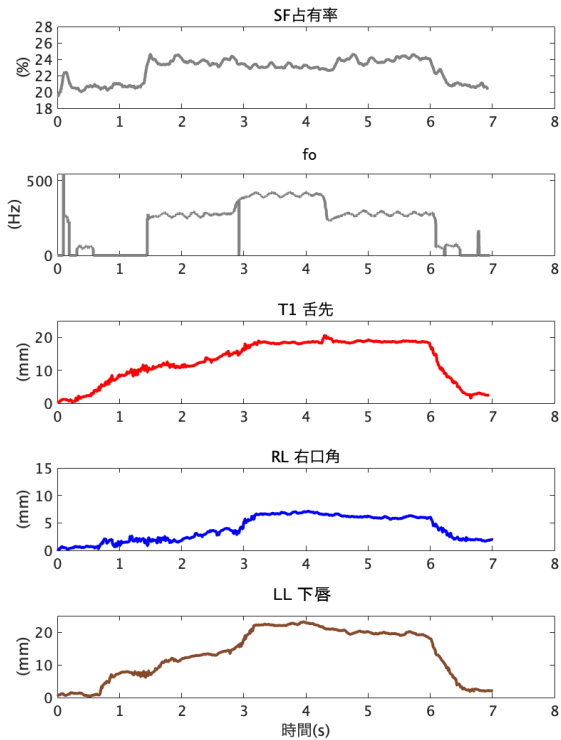


Fig. 4 Task2-1 a→a→a の pTen

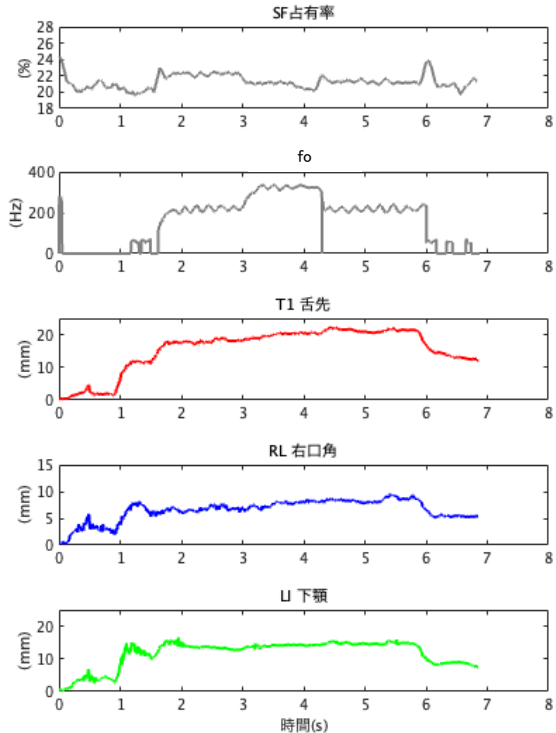


Fig. 5 Task2-1 a→a→a の aBar

いが、aBar では母音によって SF 占有率の変動が大きい。これは先行研究[2]と一致した。また、fo および各センサーの時間的変化に注目すると、aBar は各センサーが動き

始め、動作の変遷部から定常部に移行したタイミングで発声しているのに対し、pTen は各センサーが定常状態になってから発声していた。この結果から、pTen は aBar に比べて口の構えを十分に整えてから発声を開始していることが明らかとなった。

さらに、各センサーの動きに注目すると、pTen の T1 は RL および LI に比べて相対的に動きが大きく、特に母音の変化時に大きな動きを示している。一方で、aBar の T1 は歌い出しでは口の構えをとるために pTen と同程度の動きが見られたが、母音の変化時では RL や LI に比べて相対的に小さかった。また、LI のセンサーの動きは pTen よりも大きく、特に母音の変化にともなって大きく変化した。これは、pTen では下顎の動きは最小限に抑えて主に舌の動きによって母音を変化させているが、aBar は下顎と舌の双方の動きで母音を変化させていることを示唆する。/i/母音を調音するためには、声道の形状において口腔内に狭窄部を形成する必要がある。学生は顎を閉じることでこの狭窄部を作るのに対し、プロは顎をあまり閉じず、舌の隆起や前方への動きによって/i/母音を調音していると考えられる。

Fig. 4 と Fig. 5 は、Task2-1 の a→a→a における pTen と aBar の SF 占有率、fo、および各センサーの時間的変化を示している。なお、Fig. 5 の pTen の下顎データは、センサーエラーにより座標が欠損したため、予備的に下唇に取り付けたセンサー (LL:茶色) のデータを使用した。この Task2-1 では/a/母音のみの発声であり、下顎と下唇は同様の動きを示すことを確認した。

両者の SF 占有率に注目すると、Task1-2 の a→i→a と同様に、pTen の SF 占有率は

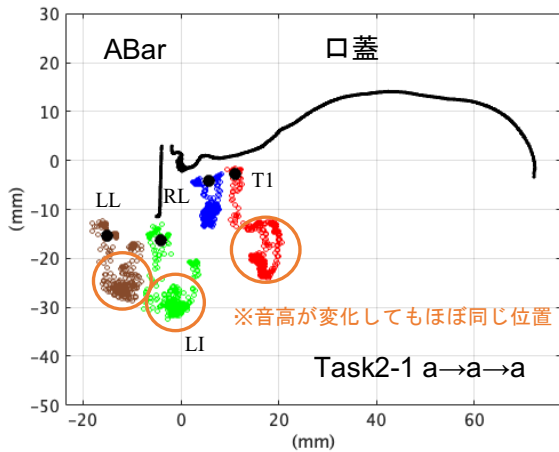


Fig. 6 ABar の口蓋の正中矢状断面

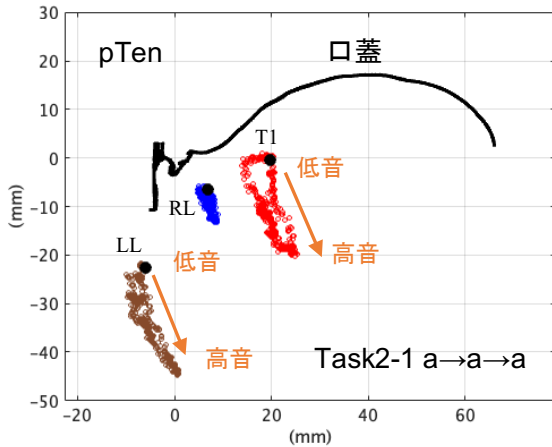


Fig. 7 pTen の口蓋の正中矢状断面

高く、母音間で安定しているのに対し、aBar は母音によって SF 占有率が変動した。また、fo および各センサーの時間的変化に注目すると、aBar の各センサーは発声と同時に定常部に達し、音高の変化にかかわらずほぼ動きがない。一方、pTen は 2 音目の最高音に向かってセンサーが徐々に動いていることが観察された。この両者の動きの方向を確認するために、各センサーの座標を正中矢状断面に投影し、口蓋の正中矢状断面 [7] とともに可視化した (Fig. 6, 7)。黒点は各センサー基準位置を示す。また、aBar にも下唇のセンサー (LL) を示す。aBar の舌と下唇、そして下顎のセンサーは、開口してからの動きがほぼ変化しないのに対して、

pTen の舌と下唇のセンサーは口を開く方向に変化し、さらに LL が、aBar よりも大きく下降していることが確認された。この結果は、学生は、発声のために開口してからタスクを歌い終わるまで口腔の形状は変えていないのに対して、pTen は高音に向けた音高の変化に対して徐々に口の構えを開ける方向に調整し、下顎を下方に開いていることを示唆している。先行研究 [8] では、プロの歌手は高音発声時に口を開き、下顎を下げ、喉頭が下降して咽頭が拡大することにより、高い SF 占有率を維持していることが報告されており、本研究の結果もそれと整合していると考えられる。

Fig. 8 と Fig. 9 は、Task2-2 の a→i→a における pTen と aBar の SF 占有率、fo、および各センサーの時間的変化を示している。両者の SF 占有率に注目すると、他の Task と同様に、pTen の値は高く安定しているが、aBar は母音が変わり音高が最高音に移行する際に一時的な増大が見られるが全体的に pTen より低く不安定である。また、fo および各センサーの時間的変化を観察すると、pTen の T1 は、Task1-2 の a→i→a における母音変化時と同様に急激に変化し、RL の動きは小さく、LI は、Task2-1 の a→a→a における音高変化時と同様に、2 音目の最高音に向けて徐々に動いている。一方、aBar は全てのセンサーが母音と音高の変化に応じて同様の動きを示しており、T1 の母音変化に対する動きが pTen よりも小さい。また、LI は pTen のような最高音に向けて徐々に変化する動きが見られず、母音の変化に対する動きが確認できる。これらの結果は、pTen は主に舌の動きで母音を制御し、下顎を徐々に変位させて高音に適切な口の構えを

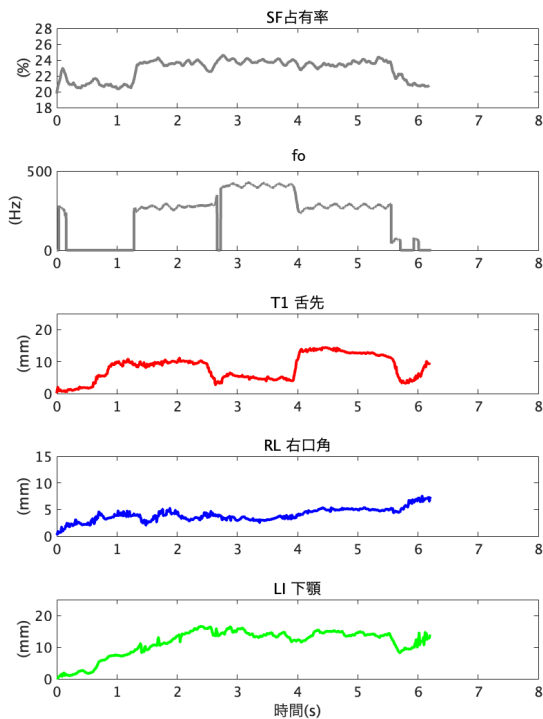


Fig. 8 Task2-2 a→i→a の pTen

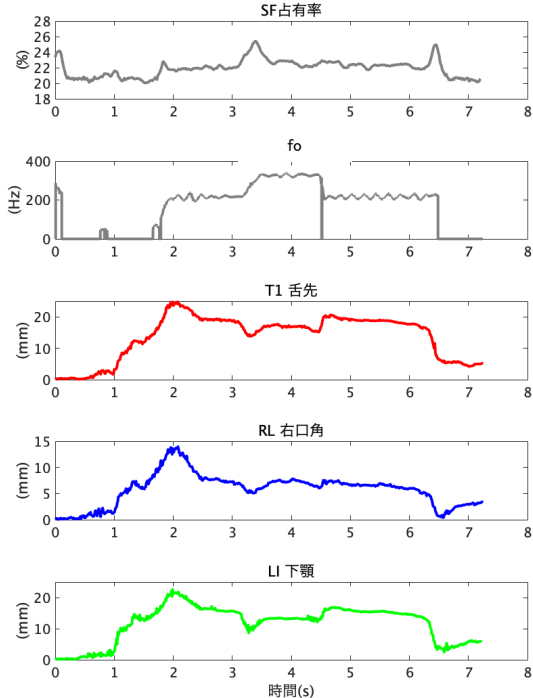


Fig. 9 Task2-2 a→i→a の ABar

つくって音高を制御していることが示唆される。一方、aBar は舌だけでなく、口唇や下顎によっても母音を制御し、pTenに見られるような音高に適切な口の構えをとっていないと考えられる。

## まとめ

本研究では、プロの歌手と声楽を学ぶ学生各1名を対象に、歌唱中の母音と音高の変化における舌と口唇の動きを計測し、SFとの関係を検討した。その結果、プロの歌手のSF占有率は学生よりも高く安定していた。また、プロの歌手は発声前に口の構えを整えており、母音の変化は主に舌で制御し、音高の変化は下顎を徐々に変位させて高音に応じた口の構えを作ることによって制御していた。一方、学生は母音の変化を舌、口唇および下顎の動きによって制御し、音高の変化ではプロに見られるような制御が見られなかった。これらのことから、プロの歌手は、発声前の口の構えや、母音変化時の舌の動き、そして音高変化時の下顎の開き具合によってSF占有率を含む歌声の音質の安定に必要な喉頭や咽頭の共鳴腔を最適化していると考えられる。しかし、本研究の知見は各1名のデータに基づくため、一般化には限界があり、今後はさらに被験者を増やして検討する必要がある。

## 謝辞

本研究は、JSPS 科研費 22K13773 と国立国語研究所 共同利用型共同研究(B)の支援を受けて行われた。

## 参考文献

- [1] Sundberg, J., *THE SCIENCE OF THE SINGING VOICE*, 1987.
- [2] 高橋他, 音講論(秋), 795-798, 2019.
- [3] 高橋他, 日本音楽知覚認知学会, 36-41, 2016
- [4] 高橋他, 音講論(春), 773-774, 2020.
- [5] コーネリウス・L・リード, 渡部東吾訳, ベル・カント唱法 その原理と実践, 音楽之友社, 1987.
- [6] 能田他, 音講論(春), 747-748, 2021.
- [7] Nota *et al.*, *JASA Express Lett.* 4, 015201, 2024.
- [8] 高橋他, 音講論(春), 737-738, 2021.