Vol.2019-MUS-123 No.24
Vol.2019-SLP-127 No.24
2019/6/22

歌声生成過程の観測に向けた歌声の声帯音源波形と声道共鳴特性の同時推定法

高橋 響子[†1,a)]　赤木 正人[†1,b)]

概要：歌声には，喉頭および口腔内で生じる乱流雑音が多い．そのため，音声生成過程の数理モデルによる歌声の声道共鳴特性および声帯音源特性の推定は困難であった．本研究では，歌声生成過程の解明に向け，乱流雑音を含む歌声の声帯音源波形と声道共鳴特性の同時推定法を提案する．推定される声道共鳴特性の安定性を保つために，声道共鳴特性は声帯音源特性に比較してゆっくりと変化するという仮定のもと声道共鳴特性を推定した．歌声のシミュレーション波形をデータとして用いた分析実験により，乱流雑音が声帯音源に含まれる歌声において，声帯音源波形と声道共鳴特性が同時推定可能であることが確認された．

キーワード：歌声分析，ARX-LF モデル，乱流雑音，声道共鳴特性

1. はじめに

幅広い音域で自在に歌唱するには，目的の音高に応じた声質（声区）への転換が重要である [1]．しかしながら，歌唱時の発声器官の瞬時的な動きを観察するのは未だに難しいため，ヒトがどのように喉頭や声道を変化させて声区転換を実現しているのかは明らかになっていない．近年 MRI やハイスピードカメラを用いた声帯や声道の直接観察は可能となってきたが，歌唱時の声帯と声道の動きを同時に見ることはできない．一方，音声生成過程の数理モデルを用いた声帯音源波形と声道共鳴特性の同時推定法であれば，計測条件の問題を克服して，声帯と声道の動きを推定できる．

これまでに，ARX-LF モデルを用いた幅広い音域を持つ歌声の声帯音源波形と声道共鳴特性の同時推定法を提案してきた [2]．しかし，この方法では裏声のような喉頭や口内で発生する乱流雑音が多く含まれる歌声において，声道共鳴特性を表すフィルタの安定性に問題があることが確認された．声区転換における声帯音源波形と声道共鳴特性の時間変化を調べるためには，乱流雑音を含む歌声においても声帯音源波形および声道共鳴特性が同時推定できることが重要である．

そこで，本稿では乱流雑音を含む歌声の声帯音源波形と声道共鳴特性の同時推定法を提案することを目的とする．これまでの方法 [2] が抱える声道共鳴特性を表すフィルタの推定方法に関する問題を解決するために，声帯音源特性と比較して，声道共鳴特性の時間変化はゆっくりしたものであると仮定し，フィルタ係数の推定を行った．これにより，声帯音源波形および声道共鳴特性の同時推定において，声帯音源特性の時間変化に追随しながら声道フィルタの安定な推定を行える手法を構築した．

†1 現在，北陸先端科学技術大学院大学
Presently with Japan Advanced Institute of Science and Technology
a) kyoko.takahashi@jaist.ac.jp
b) akagi@jaist.ac.jp

2. ARX-LF モデル

LF モデルは声帯音源波形の微分波形（dGSW）の数理モデルであり，6 つのパラメータ $T_p, T_e, T_c, T_a, T_0, E_e$ を持つ [3,4]．図 1 に LF モデルから生成される典型的な dGSW を示す．声帯音源波形の時間変化の観察を容易にするため，3 つのパラメータ O_q, α_m, Q_a を式 (1)〜(3) のように定義する．

$$O_q = \frac{T_e}{T_0}, \tag{1}$$

$$\alpha_m = \frac{T_p}{T_e}, \tag{2}$$

$$Q_a = \frac{T_a}{(1-O_q)T_0}, \tag{3}$$

O_q は声門開口時間率，α_m は開口区間の左右対称性，Q_a は声門完全閉鎖までに要する戻り区間の時間率を表す．

ARX モデルは声道フィルタの同定モデルである [5]．

Vol.2019-MUS-123 No.24
Vol.2019-SLP-127 No.24
2019/6/22

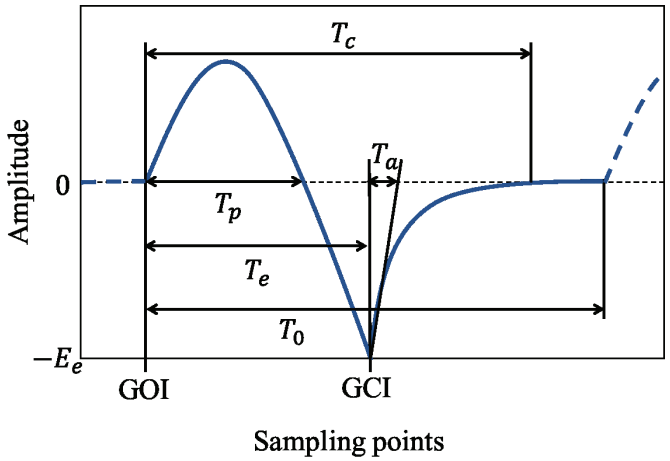

図 1 LF モデルによって生成される典型的な声帯音源波形の微分波形

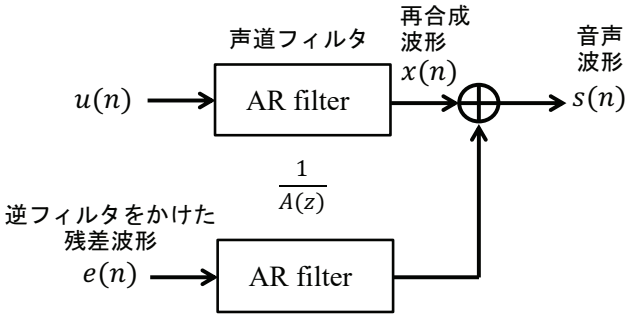

図 2 ARX モデルにおいて表現される音声生成過程

ARX モデルを用いたフィルタ同定において，音声生成過程は図 2 のように表現され，音声波形 $s(n)$ は式 (4) のように表される．

$$s(n) + \sum_{k=1}^{p} a_k(n)s(n-k) = u(n) + e(n), \qquad (4)$$

$a_k(n)$ は声道フィルタである p 次の AR フィルタの時変係数，$u(n)$ は音源，$e(n)$ は ARX モデルの逆フィルタをかけた残差波形（IFR）である．推定結果から再合成される波形 $x(n)$ は式 (5) のように表される．

$$x(n) = -\sum_{k=1}^{p} a_k(n)x(n-k) + u(n). \qquad (5)$$

LF モデルを ARX モデルにおける音源 $u(n)$ を生成するモデルとして組み込むことで，声帯音源波形の同定も可能となる．このモデルを ARX-LF モデルという．

3. 音高が高く乱流雑音を含む歌声における声道共鳴特性推定の問題

裏声のような歌声では，乱流雑音が多く含まれることがわかっている．これまでの方法 [2] では，音高の高い歌声であっても，乱流雑音のない条件下においては精度よく声帯音源波形および声道共鳴特性を推定できることが確認された．しかし，裏声のような乱流雑音を多く含む歌声において，これまでの方法は声道共鳴特性の推定結果の安定性に問題があることがわかった．

ARX モデルでは，外部入力付きの自己回帰フィルタを仮定している．そのため，自己分散法は適用できず，共分散法を用いた推定となる．共分散法ではフィルタの安定性は保証されない [6]．また，声帯音源波形および声道共鳴特性の同時推定法では，1 周期ごとに声帯音源波形と声道共鳴特性を推定する．音高が高い歌声では 1 周期の時間長が短くなり，フィルタへの入力信号の長さが短くなる．つまり，フィルタの分析窓幅が短くなりすぎてしまい，フィルタの安定性に問題が生じる．ARX モデルを用いた推定においてフィルタの安定性を保証するには，1 周期ごとの推定ではなく，分析窓長を長くすればよい．しかし，声区転換では声帯音源波形が時間的に変化することが推測されるため，声帯音源波形の 1 周期ごとの推定は必須である．

したがって，1 周期ごとの声帯音源波形・声道共鳴特性の同時推定ができ，音高が高くかつ乱流雑音がある条件下で安定なフィルタが推定できることが必要である．

4. 解決方法

本稿で提案する推定方法では，音高が高く乱流雑音を含む歌声における声道共鳴特性推定の問題を解決するために，短い時間の中では声道共鳴特性は変化しないという仮定をおいた．たとえば，短い時間を 20 ms とすれば，これまでに音声符号化における分析窓長として用いられてきた値である [7]．声帯音源波形については，これまでの方法 [2] と同様に 1 周期ごとに時間変化するものと仮定する．

ARX モデルによるフィルタ同定法は次のようになる．ARX モデルにおいて，1 周期分の音声波形の Z 変換を $S(Z)$，フィルタを $B(Z)/A(Z)$，dGSW を $U(Z)$，IFR を $E(Z)$ とすると，

$$S(Z) = \frac{B(Z)U(Z) + E(Z)}{A(Z)}, \qquad (6)$$

と書ける．式（6）を展開して，

$$A(Z)S(Z) = B(Z)U(Z) + E(Z) \qquad (7)$$

$$A(Z) = 1 - \sum_{k=1}^{p} a_k Z^{-k} \qquad (8)$$

$$B(Z) = b_0 = 1 \qquad (9)$$

となる．式（7）を時間領域で記述すると，

$$\boldsymbol{e} = \boldsymbol{x_0} - \boldsymbol{X}\boldsymbol{a} - \boldsymbol{U} \qquad (10)$$

と書ける．$\boldsymbol{e}^{\mathsf{T}}\boldsymbol{e}$ が最小となるとき，フィルタ係数 $\boldsymbol{a}$ が決定される．

そこで，声帯音源波形は 1 周期ごとに時間変化し，q 秒間声道共鳴特性は変化しないという仮定をおくと，$\boldsymbol{U}$ は式（11）のように表すことができる．

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{u_{-Q}} & \boldsymbol{u_{-Q+1}} & \cdots & \boldsymbol{u_0} \end{bmatrix} \tag{11}$$

$$Q = \frac{q}{T_0} \tag{12}$$

$\boldsymbol{u_{-Q}}, \cdots, \boldsymbol{u_{-1}}$ はそれぞれ前の周期で推定された dGSW であり，$\boldsymbol{u_0}$ は分析対象周期の dGSW である．T_0 は分析対象周期の長さである．仮定を置かない場合と比べて，$\boldsymbol{U}$ はフィルタを推定するのに十分な長さをもつ．また，仮定より q 秒間，式（4）は時不変フィルタとなり，式（13）となる．

$$s(n) + \sum_{k=1}^{p} a_k s(n-k) = u(n) + e(n) \tag{13}$$

このように，声道共鳴特性は声帯音源特性に比較してゆっくりと変化するという仮定をおくことにより，声道共鳴特性の分析窓長を広げることができるため安定なフィルタが推定できる．

5. 評価実験

これまでの方法 [2] では，音高が高く乱流雑音を含む歌声において声道共鳴特性の安定性が保証されないという問題を抱えていた．提案法の解決方法により，この問題が解決されたか確認するために分析実験を行った．また，音高が高く乱流雑音を含む歌声における声帯音源波形と声道共鳴特性の推定の正確性についても評価した．

実験データは乱流雑音を含む歌声を模擬した合成音を用いた．それぞれの評価実験で用いた実験データの作成条件を表 1，2 に示す．合成音は河原らの方法 [8] を用いて作成し，乱流雑音を含む歌声を模擬するために，雑音（ホワイトノイズ，ピンクノイズ）を次のように付加した．乱流雑音が声帯付近で発生する場合（Glottis），合成した dGSW に雑音を付加し，歌声生成フィルタに通す．表 1，2 の SNR は dGSW と雑音の比である．乱流雑音が口内で発生する場合（Other），表 1，2 に従って合成音を生成し，その合成音に雑音を付加する．表 1，2 の SNR は合成音と雑音の比である．SNR= ∞ は乱流雑音がない歌声であり，雑音を付加しないで合成した実験データである．歌声生成フィルタは典型的な母音/a/を想定して設計した．

5.1 提案法の解決方法に関する評価実験とその結果

これまでの方法 [2] と提案法について，フィルタの安定性の評価実験を行った．表 1 の条件で作成したデータを作成した．地声を想定して基本周波数（Fo）が 221 Hz，裏声として Fo が 441 Hz の 2 条件を設定した．これまでの方法 [2] と提案法を用いて 10 周期分の長さを分析した．

表 3 と表 4 に 10 周期中何周期が発散したか示す．表 3 より，地声程度の高さであれば乱流雑音が含まれていてもこれまでの方法 [2] と提案法ではフィルタの推定に問題はないことが確認できる．表 4 の裏声程度の高さのデータについて，これまでの方法 [2] による結果において，安定なフィルタが推定されない問題が発生している．一方，表 4 の提案法では，これまでの方法 [2] と比較して推定できていない周期が少なくなっている．したがって，本稿の提案した解決法により，これまでの方法 [2] が抱えていたフィルタ推定に関する問題が解決されたことが確認できる．

5.2 提案法の乱流雑音への耐性に関する評価実験とその結果

本稿の提案法が乱流雑音の量や種類によらず分析できるか確認するため，ホワイトノイズを付加雑音とするデータだけでなくピンクノイズを付加雑音とするデータを作成し，提案法によって声帯音源波形と声道共鳴特性を推定した．表 2 の条件で作成した実験データの 0.16 s を分析範囲とした．

推定結果を表 5 と表 6 に示す．表 5 と表 6 では LF モデルパラメータの平均誤差率 [%]，声道フィルタの周波数特性（ピーク周波数）の平均誤差率 [%] を示している．これらの結果より，Glottis のデータも Other のデータも雑音の量や種類にかかわらず，SNR= ∞ のデータと同等程度に各パラメータの誤差率が小さいことがわかる．このことから，提案法の雑音への十分な耐性が確認できた．LF モデルパラメータの Q_a の誤差率が他の LF モデルパラメータの誤差率より大きいが，T_a のオーダが他のパラメータに比べ一桁小さいため，式（3）より少しの誤差でも大きな誤差率となるためである．声帯音源波形の推定では，提案法に限らず他の手法でも Q_a の推定は特に難しい．

6. まとめ

本稿では，声区転換の歌声生成メカニズム解明に向け，乱流雑音を含む歌声の声帯音源波形と声道共鳴特性の同時推定法を提案した．乱流雑音を含み，音高の高い歌声におけるフィルタ推定での問題について，声道共鳴特性の時間変化に関する仮定を置くことによって解決した．評価実験より，雑音の種類やパワーにかかわらず，提案法であれば安定なフィルタを推定できることが確認された．提案法により，裏声の歌声の声帯音源波形と声道共鳴特性が同時推定できるため，声区転換における声帯音源波形および声道共鳴特性の時間変化の観察が可能となる．

地声声区と裏声声区では乱流雑音源特性が異なることがわかっている．声区転換の歌声生成メカニズムの解明のた

表 1 提案法の解決方法に関する評価実験のために作成した実験データの条件

Source				Filter	Aspiration noise		
Fo[Hz]	Oq	α_m	T_a/T_0	Vowel /a/	kinds of noise	SNR [dB]	Position
221, 441	0.3, 0.4, 0.5	0.83	0.04	F1 923 Hz F2 1179 Hz	White noise	20, 40, 60, ∞	Glottis, Other

表 2 提案法の乱流雑音への耐性に関する評価実験のために作成した実験データの条件

Source				Filter	Aspiration noise		
Fo[Hz]	Oq	α_m	T_a/T_0	Vowel /a/	kinds of noise	SNR [dB]	Position
441	0.3, 0.4, 0.5	0.83	0.04	F1 923 Hz F2 1179 Hz	White noise, Pink noise	20, 40, 60, ∞	Glottis, Other

表 3 ホワイトノイズを乱流雑音として付加した実験データ ($Fo = 221$ Hz) の推定結果: 不安定なフィルタの平均個数

	20 dB		40 dB		60 dB		∞ dB
	Glottis	Other	Glottis	Other	Glottis	Other	
これまでの方法 [2]	0	0	0	0	0	0	0
本稿の提案法	0	0	0	0	0.33	0	0

表 4 ホワイトノイズを乱流雑音として付加した実験データ ($Fo = 441$ Hz) の推定結果: 不安定なフィルタの平均個数

	20 dB		40 dB		60 dB		∞ dB
	Glottis	Other	Glottis	Other	Glottis	Other	
これまでの方法 [2]	9.7	10	10	10	10	10	1.3
本稿の提案法	2.3	2	1	1.3	1	1	2

表 5 ホワイトノイズを乱流雑音として付加した実験データの各パラメータの誤差率 [%]

	20 dB		40 dB		60 dB		∞ dB
	Glottis	Other	Glottis	Other	Glottis	Other	
O_q	8.48	14.3	12.7	12.4	13.4	6.37	13.5
α_m	6.55	6.91	8.52	9.21	3.96	7.67	3.64
Q_a	77.7	88.0	94.1	97.6	67.9	88.3	67.6
F1	1.67	3.33	2.22	2.09	0.772	2.21	0.627
F2	2.75	7.88	4.96	5.67	1.72	4.30	1.86

表 6 ピンクノイズを乱流雑音として付加した実験データの各パラメータの誤差率 [%]

	20 dB		40 dB		60 dB		∞ dB
	Glottis	Other	Glottis	Other	Glottis	Other	
O_q	5.76	29.8	13.1	17.9	11.2	6.30	13.5
α_m	6.14	5.45	4.04	8.61	5.09	7.83	3.64
Q_a	59.8	19.0	73.3	93.6	71.8	90.2	67.6
F1	1.60	5.11	0.741	2.39	0.483	2.16	0.627
F2	2.80	7.98	2.10	6.11	2.58	4.22	1.86

めに，声区転換において乱流雑音源特性がどのように変化するか調べることは重要である．したがって，歌声の乱流雑音の推定については今後の検討項目である．

参考文献

[1] 榊原： 世界の歌唱法 : 様々な歌唱様式における supranormal な声 (歌声の科学), 日本音響学会誌, Vol. 70, No. 9, pp. 499—505, 2014.

[2] K. Takahashi and M. Akagi: Estimation of glottal source waveforms and vocal tract shape for singing voices with wide frequency range, 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Honolulu, HI, USA, pp. 1879–1887, 2018.

[3] G. Fant, J. Liljencrants and Q. Lin: A four-parameter model of glottal flow, *STL-QPSR,* Vol. 26, No. 4, pp. 1–13, 1985.

[4] Q. Fu and P. Murphy: Robust Glottal Source Estimation Based on Joint Source-Filter Model Optimization, *IEEE Transactions on Audio, Speech, and Language Processing,* Vol. 14, No. 2, pp. 492–501, 2006.

[5] W. Ding and H. Kasuya: Simultaneous estimation of vocal tract and voice source parameters based on an ARX model, *IEICE TRANS. INF. & SYST.* Vol. E78–D, No. 6, 1995.

[6] 板倉, 東倉： 音声と情報処理：音声の特徴抽出と情報圧縮, 情報処理, Vol. 19, No. 7, pp. 644–656, 1978.

[7] 守谷, 鎌本, 原田, 杉浦： 音声音響符号化技術の進展, 電子情報通信学会 基礎・境界ソサイエティ Fundamentals Review, Vol. 10, No. 4, pp. 246–256, 2017.

[8] H. Kawahara, K. Sakakibara, H. Banno, M. Morise, T. Toda and T. Irino: Aliasing-free implementation of discrete-time glottal source models and their applications to speech synthesis and F0 extractor evaluation, *Signal and Information Processing Association Annual Summit and Conference (APSIPA),* 2015 Asia-Pacific, Hong Kong, pp. 520–529, 2015.