修 士 論 文

音域が広い歌声の声帯音源波形と声道形状の推定に関する研究

1610111　　高橋 響子

主指導教官　赤木正人
審査委員主査　赤木正人
審査委員　赤木正人
鵜木祐史
党建武
吉高淳夫

北陸先端科学技術大学院大学
先端科学研究科 [情報科学]

平成 30 年 2 月

概 要

ヒトは，自由な音高変化や声質の使い分けによって，表現豊かな歌声を実現している．声質は，声帯振動様式の違いによって特徴づけられることがわかっている．同一の声質で発声される声域で区分された音域を声区という．歌声における声質表現の中で，声区は重要な要素である．また，地声声区と裏声声区の間には，声区の変換点があり，地声声区と裏声声区へ声区を遷移すると，急激な基本周波数変化（ピッチジャンプ）が起こる．オペラやポップスなどの歌唱では，聴取者に声区変換での急激な途切れを感じさせないように，連続的に変換することが重要とされている．

これまでに，ヒトの歌唱の計算機による模擬を目指し，様々な手法が提案されてきた．ヒトの声質や声区は声帯振動様式に特徴づけられることから，声帯と声道を独立に制御できる手法が必要となる．声帯振動による喉頭音源波形（声帯音源波形）と声道フィルタをそれぞれ独立にモデル化し，音声生成過程を表現したモデルをソースフィルタモデルという．ソースフィルタモデルの中でも，声帯音源波形の表現と声道フィルタの同定に優れているモデルとして，Liljencrants-Fant (LF) モデルと auto-regressive with exogenous input (ARX) モデルがある．

ARX-LF モデルを用いた歌声の声質と声区の模擬は，Lu らと元田らによって実現されている．しかし，歌声の声区変換部について考慮された模擬は，未だ実現されていない．歌声の声区変換の計算機による模擬のためには，声帯音源波形の時間変化を精度よく分析する必要がある．しかし，先行研究の声帯音源波形と声道形状の推定方法は，時間的変動する声帯音源波形と声道形状の推定が困難であり，基本周波数が高い歌声における声帯音源波形と声道形状の推定精度の低さ，声帯音源波形中の周期成分と非周期成分の不完全な分離という 2 つの問題を抱えていた．

そこで，本研究では，幅広い音域に対応可能な，歌声の声帯音源波形と声道形状の推定方法を提案し，声区変換を含む歌声の分析を行うことを目的とする．

本研究の推定方法では，1 つ目の問題に対して，最小二乗法を用いた声道フィルタフィッティング，前の周期の応答の加算分補正した歌声の再合成による声帯音源波形の時間変化への対応，2 つ目の問題に対して，サンプリング周波数 44.1 kHz での周期波形の推定，EGG 信号と全探索法と Simulated annealing 法を用いた最適化によるパラメータ値探索によって解決できることを示した．歌声をシミュレーションしたデータと実際の歌声データを用いた評価実験によって，先行研究が抱えていた 2 つの問題が解決されたことが確認された．

地声と裏声の歌声の分析結果から，声区の声帯音源特性に関する知見との一致が確認された．地声から裏声へ声区変換する歌声の分析から，声区変換での声帯音源特性の滑らかな変化が見られた．また，声区変換する際，裏声声区のような特性を地声声区の時点で持つ場合があることがわかった．

目 次

図 目 次

表 目 次

第1章　序論

1.1　研究の背景

1.1.1　ヒトの歌唱

ヒトの音声コミュニケーションにおいて,「感情や意図を正確に伝えること」は永遠のテーマである．感情をはじめとする，書き言葉では表現しきれない非言語的な情報の表現に，最も適したコミュニケーション方式として歌声がある．ヒトは，自由な音高変化や声質の使い分けによって，表現豊かな歌声を実現している．

ヒトの話声および歌声は，主に声帯の振動と声道形状の変化によって特徴づけられる [1]．ヒトの発声器官は，呼吸器官，声帯，声道の 3 つで成り立つ．図 1.1 に正中面でのヒトの発声器官の形状を示す．呼吸器官は，肺にある空気を圧縮して声門や声道を通る空気流を生成する．声帯は，圧力変化や呼吸器官からの空気流により発生したベルヌーイ力によって振動し，音（喉頭音源）を生成する．声道は，喉頭音源を音響的に調節する．

声質は，声帯振動様式の違いによって特徴づけられることがわかっている [2,3]．音声の声質の分類には，喉頭音源に関連した分類と，個人的特徴に関連した分類がある [2]．喉頭音源に関連した分類は，声帯振動様式や声道形状によって区別される分類である．個人的特徴に関連した分類は，性差や年齢などに起因する差異によって区分される分類である．喉頭音源に関連した分類である Laver の 5 名義尺度によると，modal voice, falsetto, whisper, creak, harshness, breathiness がある [4]．breathiness については，さらに grade, roughness，breathiness，asthenia，strain に細かく分類される [5]．breathiness では，声帯での乱流や声帯ノイズ（aspiration noise）が重要であることが報告されている [5–7]．

声区は，歌声における声質表現の中で重要な要素である [8]．同一の声質で発声される声域で区分された音域を声区 [9] という．基本周波数の低い方から，フライ（vocal fry），地声（modal），裏声（falsetto），ホイッスル（whistle）と分類される [8]．図 1.2 に声区の分類と音高の関係を示す．

声区ごとの音響的特徴や声帯振動様式の違いについて，さまざまな手法による解明が進められてきた．modal の喉頭音源のスペクトル傾斜は-12 dB/oct であるのに対し，falsetto では modal より傾斜が急峻となる [1]．森下らは，STRAIGHT を用いた音響分析によって，falsetto の基本周波数とケプストラム 1 次項は，modal のものに比べ高くなることを確認した [10]．Henrich らの Electroglottogram (EGG) 信号を用いた計測結果より，1 周期中で声門が開いている割合（声門開口時間率，open quotient）は，modal は 0.3-0.8，falsetto

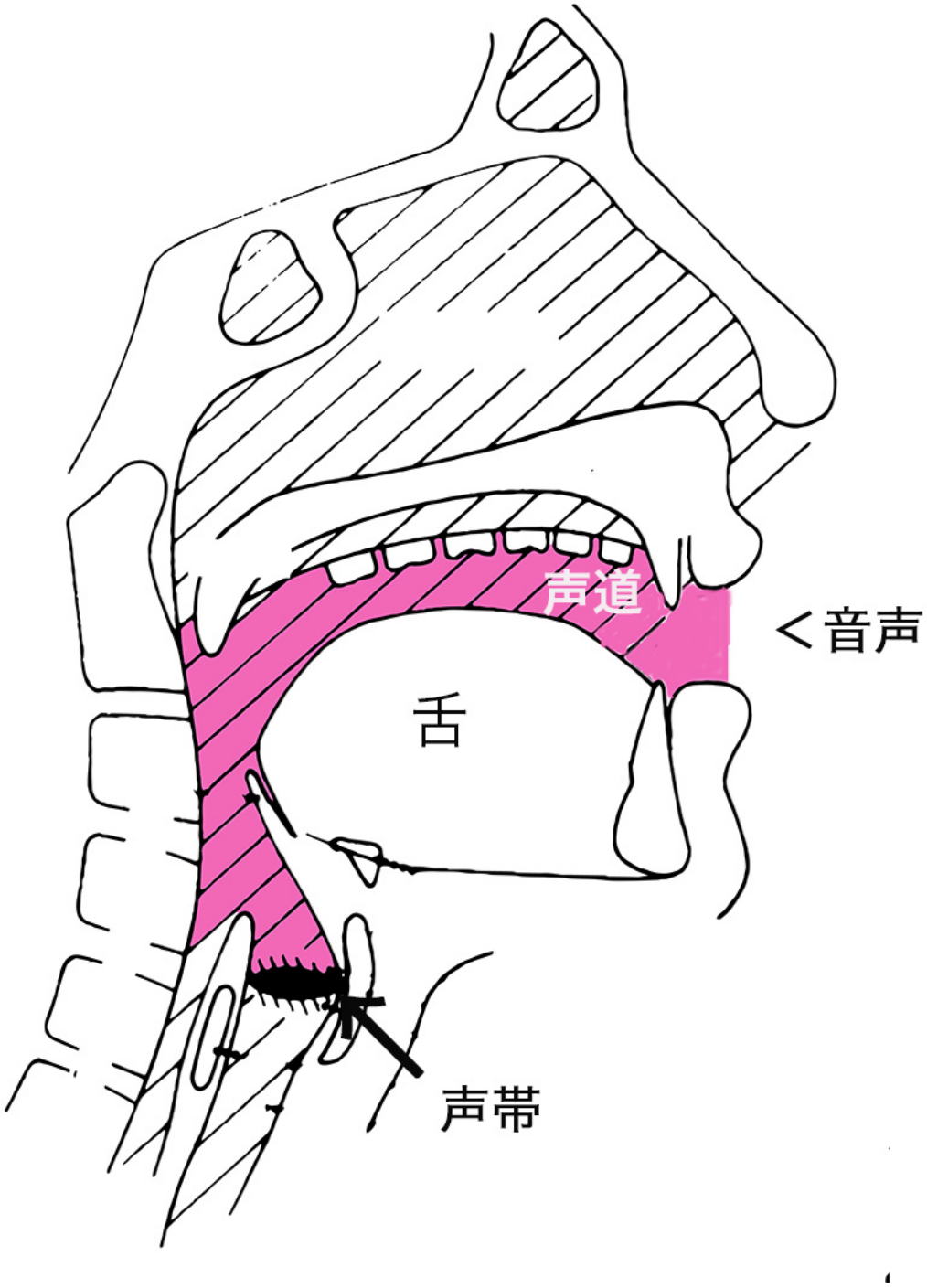


図 1.1: 正中面でのヒトの発声器官の形状

は 0.5-0.95 である [11]．今川らのハイスピードカメラを用いた計測結果より，声門面積の最大値は，modal が falsetto と比較して約 2 倍となる [12]．声門の閉鎖について，modal は完全閉鎖であるが，falsetto はしばしば定常的な間隙が存在する [8]．また，falsetto では声門辺縁部の限局的な振動が見られることがわかっている [13]．

modal と falsetto の間には，声区の変換点がある．modal から falsetto へ声区を変換すると，急激な基本周波数変化（ピッチジャンプ）が起こる．modal と falsetto の声区変換点は，男女ともに C4 − E4（261.6 − 329.6 Hz）の音高に存在する [8]．オペラやポップスなどの歌手では，聴取者に声区変換での急激な途切れを感じさせないように，連続的に変換することが重要とされている [8,14]．

1.1.2 計算機によるヒトの歌唱の模擬

これまでに，計算機上でのヒトの歌声の模擬が試みられてきた．歌声を話声の一種であるという考えを基にして，齋藤らはヒトの話声の歌声への変換を実現した [15]．剣持らは，音素片を接続することによる歌声合成方法を提案した [16]．徳田らは，音声の動的特徴量を含む隠れマルコフモデルを用いた音声合成方法を提案し，それを応用した歌声合成システムを構築した [17,18]．これらの研究 [15–18] で提案された手法は，音声の物理的特徴量を制御する方法であり，声帯と声道の独立な制御は考慮していない．ヒトの声質や声区は声帯振動様式に特徴づけられることから，声帯と声道を独立に制御できる手法が必要となる．

ソースフィルタ理論より，ヒトの音声や歌声は，声帯振動による喉頭音源を声道フィルタに入力した出力と定義される [19]．図 1.3 のように声帯振動による喉頭音源波形（声帯音源波形）と声道フィルタをそれぞれ独立にモデル化し，音声生成過程を表現したモデルをソースフィルタモデルという．声帯音源波形に関するモデルは，Rosenberg-Klatt (RK) モデル [6] や Liljencrants-Fant (LF) モデル [20] が提案されている．声道フィルタの同定モデルは，auto-regressive with exogenous input (ARX) モデル [21] が提案されている．

ソースフィルタモデルを用いた歌声の声質と声区の模擬は，Lu らと元田らによって実現されている．Lu らは ARX-LF モデルを用いた声帯ノイズの推定・合成方法を提案した [7,22]．そして，ARX-LF モデルを用いた breathy voice の声質の歌声合成を実現している [23]．元田らは，ARX-LF モデルを用いて声区ごとの声帯音源波形の特徴を分析し，声区ごとに独立した歌声合成を実現した [24,25]．

1.1.3 先行研究の問題点

歌唱者は，一つの声区に囚われず，複数の声区を遷移することで，自由に音高変化して歌うことが可能である．ヒトの歌唱の模擬には，声区変換の合成の実現も必須であると考えられる．元田らは，声区ごとに独立な制御規則を構築したが，声区から異なる声区への遷移する場合における制御規則の構築は達成できていない．声区変換のための制御規則構

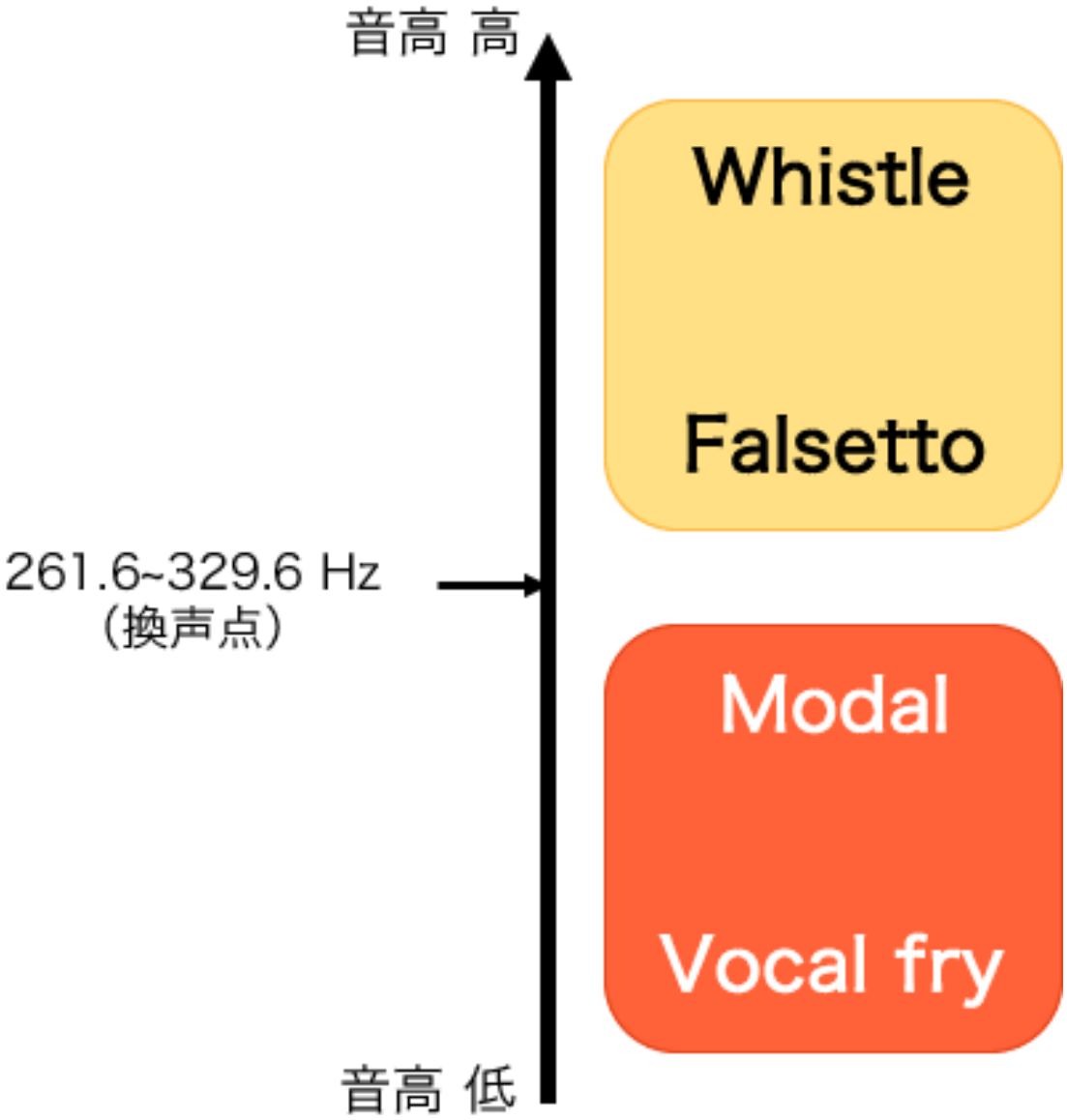


図 1.2: 声区の分類

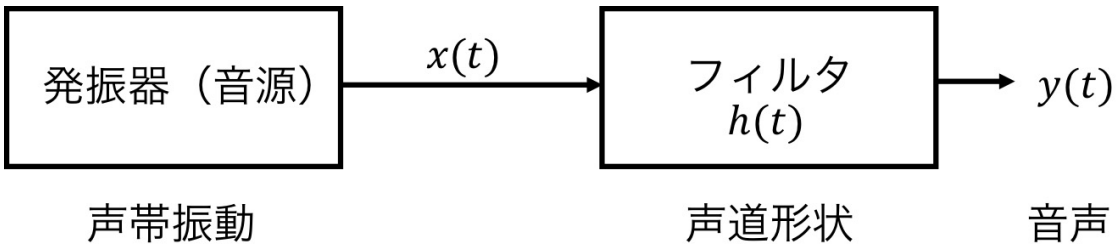


図 1.3: ソースフィルタモデル

築のためには，声区から異なる声区への遷移する際の声帯音源波形の時間変化を分析する必要がある．しかし，Lu らや元田らの歌声分析方法では，声帯音源特性の傾向を得ることができても，その時間変化を観察するのは困難であった．

1.2 研究の目的

本研究では，幅広い音域に対応可能な，ARX-LF モデルを用いた歌声の声帯音源波形と声道形状の推定方法を提案し，声区変換する歌声の声帯音源波形の時間的特徴を分析することを目的とする．

従来の声帯音源波形と声道形状の推定方法において，声帯音源波形と声道形状の時間的変動の観察が困難である．これには，基本周波数が高い歌声における声帯音源波形と声道形状の推定精度の低さ，声帯音源波形中の周期成分と非周期成分の不完全な分離という 2 つの問題点が関連している．これらの問題を解決し，歌声の声帯音源波形と声道形状の推定方法について提案する．

広い音域の歌声の声帯音源波形と声道形状を推定できれば，声区変換だけでなく，他の歌唱表現の分析にも利用出来る．声帯音源波形の各周期が高精度に推定できれば，綿密な制御規則を構築できるため，高品質な歌声合成が可能となる．声帯音源波形と声帯ノイズの十分な分離ができれば，声帯ノイズの合成が容易になる．また，声帯と声道の時間変化がわかれば，声楽などの教育にも寄与できる．

1.3 本論文の構成

第 2 章では，有声音源モデルとして RK モデルと LF モデル，声道フィルタ同定モデルとして線形予測分析法と ARX モデルについて説明する．ARX-RK モデルと ARX-LF モデルを用いた，音声分析合成の先行研究，歌声分析合成の先行研究について述べ，歌声分析に関して先行研究の抱える問題点と原因を述べる．

第 3 章では，本研究で提案する声帯音源波形と声道形状の推定方法について述べる．ARX-LF モデルパラメータ初期値の決定方法，ARX-LF モデルパラメータ値の探索方法，前の周期からの影響を考慮した歌声合成方法について詳細に述べる．

第 4 章では，本研究の推定方法についてシミュレーションデータと実際の歌声を用いた分析実験で評価する．

第 5 章では，声区ごとの声帯音源波形の特性と，声区変換する歌声の声帯音源波形と声道形状の時間変化の分析結果を報告する．

第 6 章では，本研究で得られた成果をまとめる．

第2章　音声生成過程に着目した歌声分析・合成手法

2.1　はじめに

ソースフィルタ理論より，ヒトの音声や歌声は，喉頭音源を入力とする声道フィルタの出力である [19]．ソースフィルタ理論に基づいた音声生成モデルはいくつか提案されている．有声音源モデルとして，RK モデル，LF モデルがある．声道フィルタの分析モデルとして，線形予測分析法と ARX モデルがある．これらのモデルについて説明し，これらを利用した音声・歌声の分析合成に関する先行研究の概要と問題点について述べる．

2.2　音声生成モデル

2.2.1　有声音源モデル

声門体積流（声帯音源波形）に口唇での放射特性（微分特性）を含んだ波形を微分声帯音源波形とよび，その形状を多項式で記述した数式モデルに RK モデル [6] と LF モデル [20] がある．RK モデルは LF モデルと比較して，モデルパラメータが少ない分制御が容易であるが，声帯音源波形のスペクトル傾斜を表現出来るパラメータを持たない．そこで，本研究では LF モデルを用いる．

RK モデル

RK モデルは式 2.1 で定義される [6, 21]．

$$
\begin{aligned}
g(t) &= \begin{cases} 2mt - 3nt^2 & 0 \leq t \leq T_0 \cdot OQ \\ 0 & T_0 \cdot OQ \leq t \leq T_0 \end{cases} \\
m &= \frac{27 \cdot AV}{4 \cdot (OQ^2 \cdot T_0)}, n = \frac{27 \cdot AV}{4 \cdot (OQ^3 \cdot T_0^2)}
\end{aligned}
\tag{2.1}
$$

T_0 は周期の長さ，AV は最大振幅，OQ は声門開放時間率である．RK モデルでスペクトル傾斜を表すには，IIR フィルタによって $g(t)$ をフィルタすることで調節する [26]．

LF モデル

LF モデルは，図 2.1 に示すような，$T_p, T_e, T_a, T_c, T_0, E_e$6 つのパラメータを持つモデルである [27,28]．T_p は声帯音源波形の最大値となる時間を表し，T_e は声門開放区間，T_a は声門閉鎖までの戻り区間，T_c は声門完全閉鎖時間，T_0 は周期の長さ，E_e は最大振幅を表す．Glottal Opening Instant(GOI) は波形の始点であり，Glottal Closure Instant(GCI) は声門閉鎖開始点である．LF モデルは式 2.2 で表される．

$$u(t) = \begin{cases} E_1 e^{at} \sin(\omega t) & 0 \leq t \leq T_e \\ -E_2[e^{-b(t-T_e)} - e^{-b(T_0-T_e)}] & T_e \leq t \leq T_c \\ 0 & T_c \leq t \leq T_0 \end{cases} \tag{2.2}$$

E_1, E_2, a, b, ω は $T_p, T_e, T_a, T_c, T_0, E_e$ に関係している．

2.2.2 声道フィルタ同定モデル

声道フィルタの分析方法には，線形予測分析法（LPC）[29] と ARX 分析法 [30] がある．図 2.2 と図 2.3 に，それぞれが仮定する音声生成モデルを示す．LPC 分析法はフィルタ係数が簡単に推定できるが，ARX モデルでは声帯音源波形を入力とするため近似度の高いフィルタ係数を推定できる [31]．本研究では，ARX モデルを用いる．

線形予測分析法

LPC 分析法では，白色雑音あるいは単一インパルスを入力した全極型声道フィルタの応答を音声信号として考える．LPC 分析法は，図 2.2 のように単純な音声生成モデルを仮定するため，フィルタ係数が簡単に推定できる．LPC における音声生成モデルは式 2.3 で表される．

$$s(n) = -\sum_{k=1}^{p} a_k(n) s(n-k) \tag{2.3}$$

$s(n)$ は音声信号，$a_k(n)$ は p 次の AR フィルタの時変定数である．ホルマント声帯音源のスペクトル特性と声道の周波数伝達特性を区別できず，音源と声道フィルタ特性は全極型 AR フィルタにまとめて表される．

ARX モデル

ARX モデルでの音声生成過程のモデルは図 2.3 となる．ARX モデルにおける音声信号は，単一インパルスでない声帯音源パルス列を入力した極・零型声道フィルタの応答と，

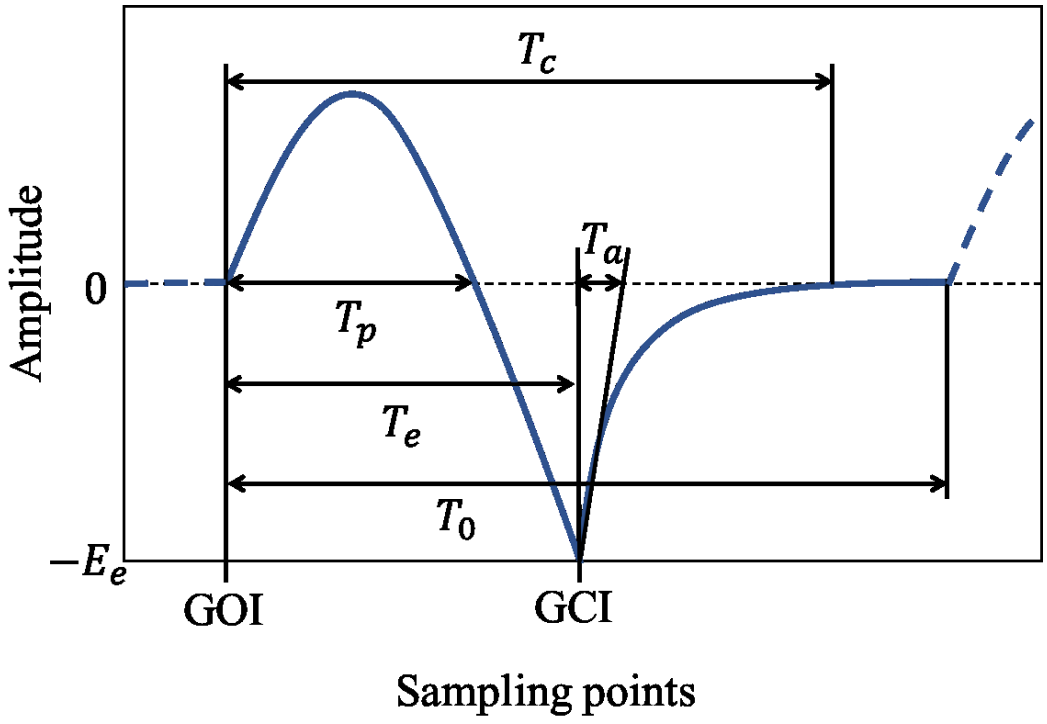


図 2.1: LF モデルのパラメータ

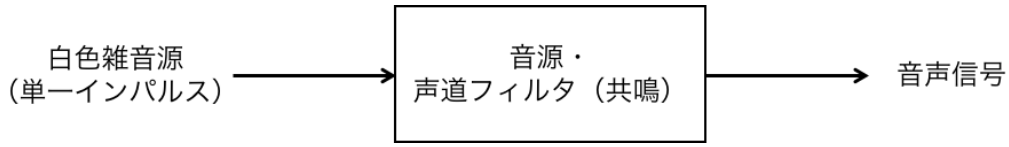


図 2.2: 線形予測分析の音声生成モデル

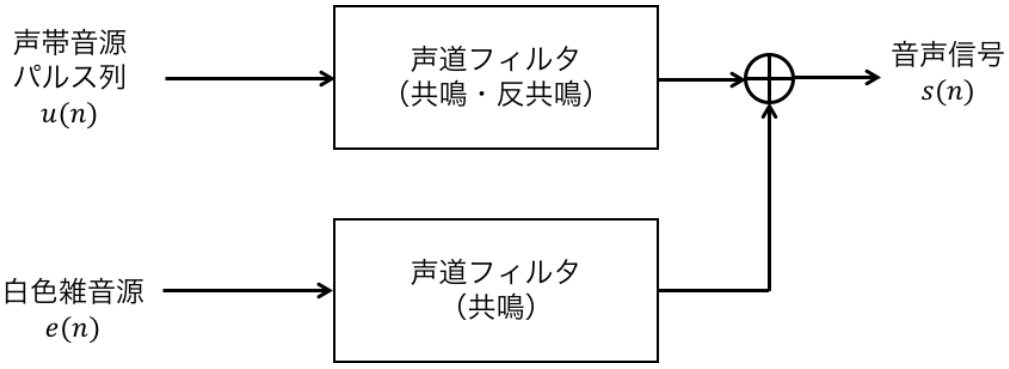


図 2.3: ARX 分析法の音声生成モデル

白色雑音を入力した声道フィルタの応答の足しあわせで表現される．音声信号 $s(n)$ は式 2.4 のように表される．

$$s(n) + \sum_{k=1}^{p} a_k(n)s(n-k) = u(n) + e(n) \tag{2.4}$$

$a_k(n)$ は p 次の AR フィルタの時変定数，$u(n)$ は声帯音源波形の微分形（周期波形），$e(n)$ は ARX モデルの式誤差と声帯ノイズ（非周期波形）を表す．$u(n)$ は，LF モデルの出力である．式誤差がない理想的な推定ならば，$e(n)$ は白色雑音のような非周期波形のみとなる．再合成される音声信号 $x(n)$ は，式 2.5 で表される．

$$x(n) = \sum_{k=1}^{p} a_k(n)s(n-k) + u(n) \tag{2.5}$$

2.3 ソースフィルタモデルを用いた先行研究

ソースフィルタ理論に基づいた音声分析・合成方法は，いくつか提案されている．Ding と粕谷は，ARX-RK モデルを用いた音声分析・合成方法を提案した [21]．ARX モデルによる声道フィルタ分析法を提案したことで，LPC 分析法より高精度な推定を実現した．大塚と粕谷は，音源パルス列を用いることで，ARX-RK モデルを用いた音声分析方法を向上させた [32]．話声の分析結果より，女性や子供の話声のような基本周波数が高い音声についても高精度に分析可能であることが示された．Vincent らは，ARX-LF モデルを用いた音声分析・合成方法を提案した [33]．LF モデルの低周波数帯域と高周波数帯域部分を分けた推定法 [34]，音声から推定した基本周波数を利用した *GCI* 特定法 [35]，Harmonic plus noise モデル [36] を用いた声帯ノイズの推定法によって，高精度な分析・合成を実現した．

ARX-LF モデルを用いた歌声の分析・合成も提案されている．Lu と Smith Ⅲ は，歌声の声帯ノイズに注目し，歌声に含まれる声帯ノイズの抽出・合成方法を提案した [7,23,37]．元田と赤木は，ARX-LF モデルを用いて声区ごとの声帯音源特性を分析し，声区ごとの独立した歌声合成を実現した [24,25,38]．結果より，声区ごとに異なる声帯音源波形の傾向が確認された．

2.4 先行研究が抱えている問題点

しかし，これらの先行研究 [7,23–25,37,38] は，歌声の声帯音源波形と声道形状の推定に関して，次の 2 つの問題を抱えている．

1 つ目は，高い音高の歌声，つまり基本周波数が高い歌声における声帯音源波形と声道形状の推定精度の低さである．基本周波数が高い歌声では，各周期の長さは短い．その場合，先行研究で用いているカルマンフィルタアルゴリズム [39] では，各周期の声道フィ

ルタのフィッティングが困難であり，解が収束しない．また，実際の歌声では，声帯音源波形は時間的に変動する．先行研究では，声帯音源波形は変動しないと仮定して各周期の声道フィルタを推定している．周期ごとに声道フィルタを推定する場合，声道フィルタの整定時間が周期の長さを超過し，前の周期の声道フィルタの応答が対象周期にずれこむ．したがって，歌声の再合成では，前の周期からの影響を考慮する必要がある．

2つ目は，声帯音源波形中の周期成分と非周期成分の不完全な分離である．この問題では，LF モデルパラメータの推定誤差が原因となっている．特に，GCI の誤差が大きいため，声帯音源波形中の周期成分と非周期成分が十分に分離できない．音声の基本周波数の推定精度の影響も受けるため，音声波形から正確な GCI を特定することは非常に困難である．Li らは，GCI の特定に Electroglottogram (EGG) 信号を用いて，感情音声の ARX-LF モデルパラメータの推定を行った [40]．結果より，EGG 信号が有用であることは確認されたが，基本周波数が高い歌声の推定は未だ困難である．

2.5 まとめ

この章では，従来の歌声の声帯音源波形と声道形状の推定方法，先行研究の推定方法が抱える問題点について説明した．声帯音源波形の数理モデルとして，スペクトル傾斜を表現出来るパラメータを持つ LF モデル，声道フィルタの同定モデルとして ARX モデルが，声帯音源波形と声道フィルタの推定に最適であると考えられる．このような ARX-LF モデルを用いた先行研究が抱えている問題点として，次の 2 点がある．

- 基本周波数が高い歌声における声帯音源波形と声道形状の推定精度の低さ
- 声帯音源波形中の周期成分と非周期成分の不完全な分離

次の章において，この 2 つの問題点の解決策と本研究で提案する声帯音源波形と声道形状の推定方法について述べる．

第3章　声帯音源波形と声道形状の推定方法

3.1　はじめに

この章では，本研究で提案する，歌声の声帯音源波形と声道形状の推定方法について述べる．

先行研究 [7,23–25,37,38,40] では，声帯音源波形と声道フィルタの時間的変動の観察が困難である．これには，基本周波数が高い歌声における声帯音源波形と声道形状の推定精度の低さ，声帯音源波形中の周期成分と非周期成分の不完全な分離という 2 つの問題点が関連している．

まず，本研究の推定方法の概要を述べる．続いて，先行研究の抱える問題への解決方法の詳細を述べていく．

3.2　推定方法の概要

本研究で提案する推定方法の全体の流れを図 3.1 に示す．Step1 と 2 は，ARX-LF モデルのパラメータ値の推定を行う段階である．Step2 と 4 では，前の周期からの影響を考慮した歌声の再合成を行う．

Step1 LF モデルパラメータの最適値の存在範囲を特定するために，EGG 信号から LF モデルパラメータ初期値を計算する．

Step2 ARX-LF モデルパラメータの最適値を求めるために，Step1 で求めたパラメータ初期値を基に，全探索法と Simulated Annealing 法 [41, 42] で ARX-LF モデルパラメータ値を推定する．

Step3 周期波形 $u(n)$ を合成する．

Step4 前の周期からの影響を考慮した合成波形 $x(n)$ を再合成する．

Step5 非周期波形 $e(n)$ を計算する．

Step6 対象周期の推定結果を保存する．

そして，同様に次の周期の推定を行う．

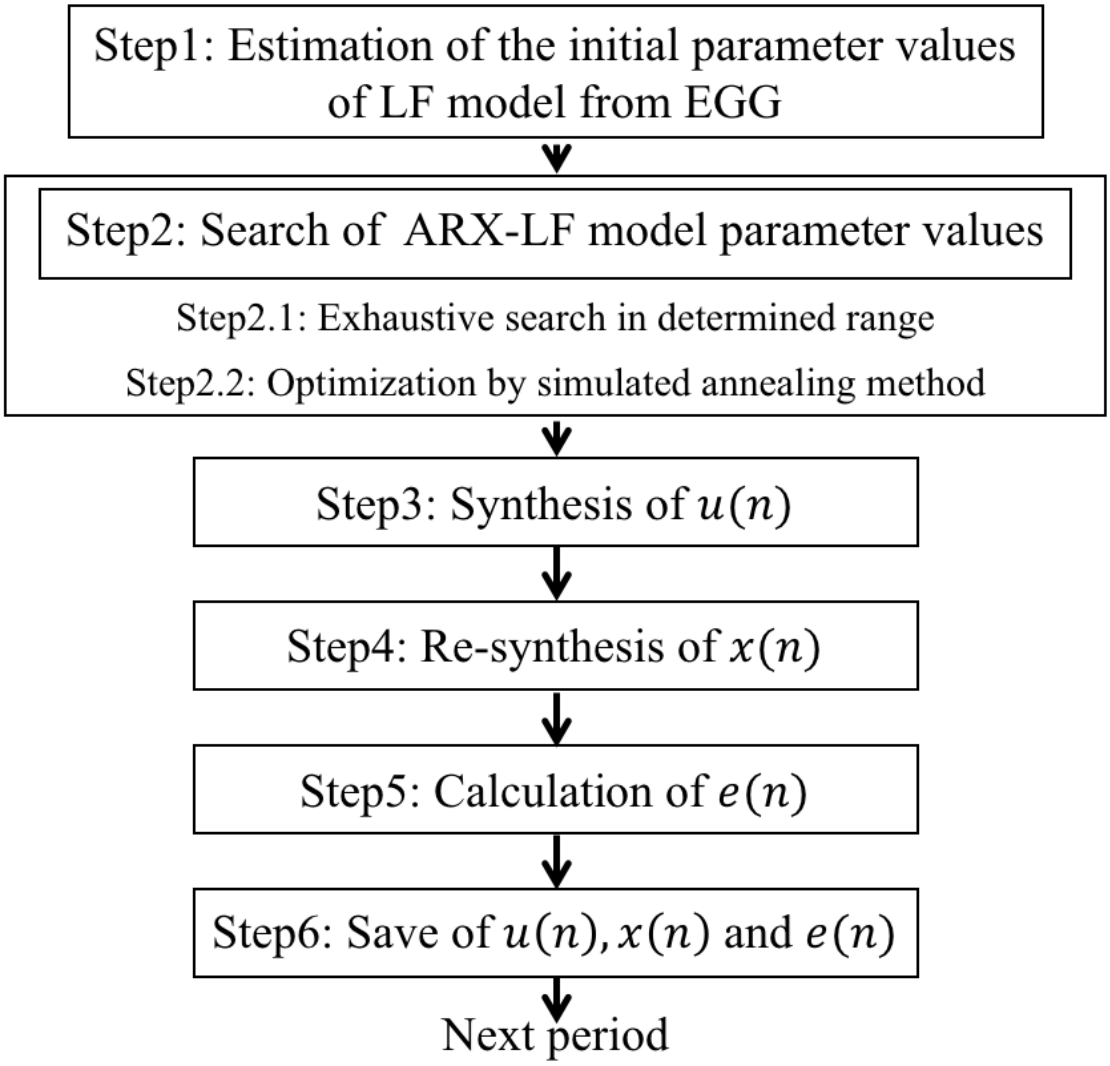


図 3.1: 本研究の声帯音源波形と声道形状の推定手順

3.3 基本周波数が高い歌声における声帯音源波形と声道形状の推定精度の低さへの解決方策

基本周波数が高い歌声における声帯音源波形と声道形状の推定精度の低さを改善するための，各周期の短さへ対応した声道フィルタの推定方法と声帯音源波形の時間変化の考慮方法について述べる．

3.3.1 各周期の短さへ対応した声道フィルタのフィッティング

声帯音源波形と声道フィルタの時間変化を観察するために，それぞれ 1 周期ごとに推定している．そのため，各周期の声道フィルタを推定する際，入力信号である周期波形 $u(n)$ は 1 周期分の長さをもつ波形となる．そして，基本周波数が高い歌声では，$u(n)$ の長さは短いものとなる．先行研究で用いているカルマンフィルタアルゴリズム [39] では，入力信号がある程度の長さを持っていないと，解が収束せず，フィッティングが困難である．それにより，基本周波数が高い歌声における声帯音源波形と声道形状の推定精度の低さ，あるいは声帯音源波形と声道形状が推定できないという状況が生じていた．そこで，本研究では入力信号の長さに依らずフィッティング可能な最小二乗法を用いた．

3.3.2 声帯音源波形の時間変化の考慮方法

Step2，4 では，推定された声道フィルタと周期波形 $u(n)$ から合成波形 $x(n)$ を再合成する．実際の歌声では，声帯音源波形は時間的に変動するため，歌声の再合成において前の周期からの影響を考慮する必要がある．特に，基本周波数が高い歌声の推定では，前の周期からの影響が大きい．基本周波数が高い歌声では，声道フィルタの整定時間が周期の長さを超過し，前の周期の声道フィルタの応答が対象周期にずれこむ．そのため，前の周期からの影響を考慮することは重要である．

前の周期からの影響を含めた合成波形 $x(n)$ を再合成するために,「数周期の間，声道フィルタは時不変である」という仮定する．ここで，分析対象周期を N 周期目とする．まず，推定された声道フィルタの整定時間 L（許容範囲 2%）を計算する．得られた整定時間から，何周期前からの影響を考慮する必要があるか計算する（式 3.1）．

$$M = \frac{L}{T_0} \tag{3.1}$$

続いて，$N - M$ 周期目から $N - 1$ 周期目の周期波形と，Step3 で合成された N 周期目の周期波形 $u(n)$ から，整定時間以上の長さの周期波形 $u_l(n)$ を作成する．$u_l(n)$ を声道フィルタに入力し，$M + 1$ の長さを持つ合成波形 $x_l(n)$ を得る．$x_l(n)$ の後半 1 周期分が $x(n)$ となる．これによって，前の周期からの影響を考慮した合成波形の再合成が可能となる．

3.4 声帯音源波形中の周期成分と非周期成分の不完全な分離への解決方策

声帯音源波形中の周期成分と非周期成分の不完全な分離を改善するための，ARX-LF モデルパラメータ値の探索方法について述べる．

3.4.1 周期波形 $u(n)$ の表現

LF モデルパラメータにおいて，最も重要なパラメータは GCI である．GCI は，分析対象の $s(n)$ の中で推定した波形 $x(n)$ の対応する点を決定する役割を担っている．また，GCI の誤差はサンプリング周波数によって左右される．GCI の微小な誤差により，非周期波形 $e(n)$ に周期成分として現れてしまう．先行研究 [24,25,38] では，最大周波数 6 kHz の声道フィルタ推定のために，周期波形 $u(n)$ を 12 kHz サンプリングで推定している．しかし，12 kHz サンプリングでは，GCI の表現が難しい．そのため，周期波形 $u(n)$ のサンプリング周波数は，$s(n)$ と同様の 44.1 kHz とした．そして，声道フィルタを推定する直前に，$u(n)$ を 12 kHz にダウンサンプリングして，声道フィルタの入力信号として用いた．この手順は，Step2.1 の全探索法と Step2.2 の Simulated annealing 法による最適化にふくまれている．

3.4.2 EGG 信号を用いた LF モデルパラメータ初期値の計算

Step1 では，Li らと同様に EGG 信号から，GCI，GOI の初期値を求める [40]．そして，GCI，GOI 初期値から，LF モデルパラメータ T_e，T_0 を計算する．EGG 信号からは，GCI はかなり明瞭に計測できるが，GOI については GCI ほど明確ではない．そのため，これらを LF モデルパラメータ初期値とし，Step2 でさらに詳細に探索する．

3.4.3 ARX-LF モデルパラメータの探索

Step2 で求められた LF モデルパラメータ初期値を基に，GCI，GOI，T_p, T_e, T_a, T_c, E_e の探索範囲を決定する．まず，この探索範囲内で ARX-LF モデルパラメータ値の全探索を行う．続いて，全探索の結果から探索範囲を狭め，焼きなまし法（Simulated annealing 法）[41,42] で ARX-LF モデルパラメータ値を最適化する．全探索と Simulated annealing 法の探索条件は式 3.2 となる．

$$\begin{aligned} &minimize \quad f = \sum \{s(n) - x(n)\}^2 \\ &limitation \quad 0 < T_p < T_e < T_0 \\ &\qquad\qquad 0.8 < T_c/T_0 < 1 \\ &\qquad\qquad 0.01 < T_a/T_0 < 1 \end{aligned} \tag{3.2}$$

3.5 まとめ

この章では，本研究で提案する歌声の声帯音源波形と声道形状の推定方法について述べた．基本周波数が高い歌声における声帯音源波形と声道形状の推定精度の低さへの解決方策として，

- 最小二乗法を用いた声道フィルタフィッティング
- 前の周期の応答の加算分補正した歌声の再合成による声帯音源波形の時間変化への対応

を提案した．声帯音源波形中の周期成分と非周期成分の不完全な分離への解決方策として，

- サンプリング周波数 44.1 kHz の周期波形 $u(n)$
- EGG 信号を用いた LF モデルパラメータ初期値の計算
- 全探索法と Simulated annealing 法の最適化による ARX-LF モデルパラメータ値探索

を提案した．

第4章　推定方法の評価

4.1　はじめに

本研究で提案する，歌声の声帯音源波形と声道形状の推定方法の評価を行う．歌声をシミュレーションしたデータと実際の歌声データを用いた分析実験によって評価する．シミュレーションデータの分析実験によって，基本周波数が高い歌声への対応を検証した．実際の歌声データの分析実験によって，声帯音源波形中の周期成分と非周期成分の分離を確認した．

まず，歌声をシミュレーションしたデータの分析実験について述べる．シミュレーションデータの作成方法と作成条件，分析結果と先行研究 [24] との比較結果について述べる．つづいて，実際の歌声データの分析実験について述べる．分析した歌声データの条件，先行研究 [24] との比較結果について述べる．

4.2　シミュレーションデータの分析

4.2.1　シミュレーションデータの作成

シミュレーションデータの作成には，河原らの「SparkNG: Matlab realtime speech tools and voice production tools」を用いた [43]．河原らの合成システムでは，LF モデルパラメータ値と声道フィルタ係数（声道形状），声帯ノイズの量を設定できる．分析実験で用いたシミュレーションデータ合計 9 個は，次のように設定して作成した．

- LF モデルパラメータ値
  - $T_e/T_0 = 0.3, 0.4, 0.5$
  - $1/T_0 = f\mathrm{o} = 147, 221, 441$ Hz
- 声道フィルタ係数（声道形状）
  - 典型的な/a/の声道形状
  - 第 1 ホルマント周波数: 969 Hz
  - 第 2 ホルマント周波数: 1184 Hz

- 声帯ノイズはないものと仮定
- 44.1 kHz サンプリング

4.2.2 分析結果

表 4.1 に，シミュレーションデータの分析結果の各パラメータの平均誤差率を基本周波数（fo）ごとに示す．Fr1，Fr2 は声道フィルタの第 1 ホルマント，第 2 ホルマントである．すべてのデータにおいて，LF モデルにおいて重要な意味を持つパラメータ T_p, T_e の誤差率が十分小さいことが示された．また，Fr1，Fr2 とも誤差率が十分小さい．

先行研究 [24] の推定方法と比較する．図 4.1 に，分析結果の残差 $e(n)$ の平均二乗誤差を示す．残差 $e(n)$ の最小二乗誤差 $\varepsilon(n)$ は，式 4.1 のように計算した．

$$\varepsilon(n) = \frac{1}{M} \sum e(n)^2 \tag{4.1}$$

残差 $e(n)$ には，ARX モデルの式誤差と非周期成分が含まれる．シミュレーションデータでは，声帯ノイズはないと設定したので，非周期成分がない．そのため，シミュレーションデータの分析実験での残差 $e(n)$ は，推定誤差を表す．図 4.1(a) に先行研究 [24] の推定方法による $\varepsilon(n)$，図 4.1(b) に本研究の推定方法による $\varepsilon(n)$ を示す．これらの結果より，先行研究 [24] と比較して，本研究の推定結果の誤差の減少が見られる．fo 147 Hz の 3 データで平均 91.8%，fo 221 Hz の 3 データで平均 84.2%，fo 441 Hz の 3 データで平均 71.9% 減少した．

シミュレーションデータの分析実験により，本研究の推定方法によって，基本周波数が高い歌声への対応が確認された．

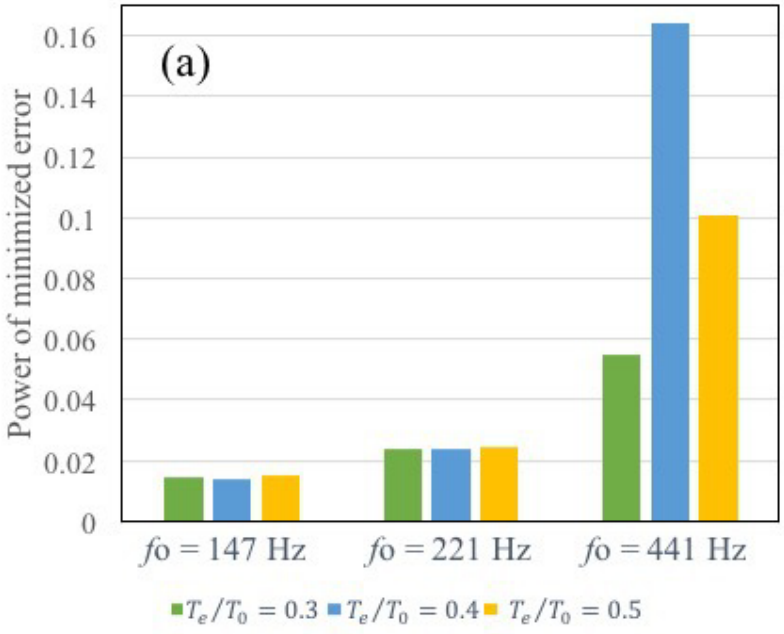
4.3 歌声の分析

4.3.1 分析した歌声データ

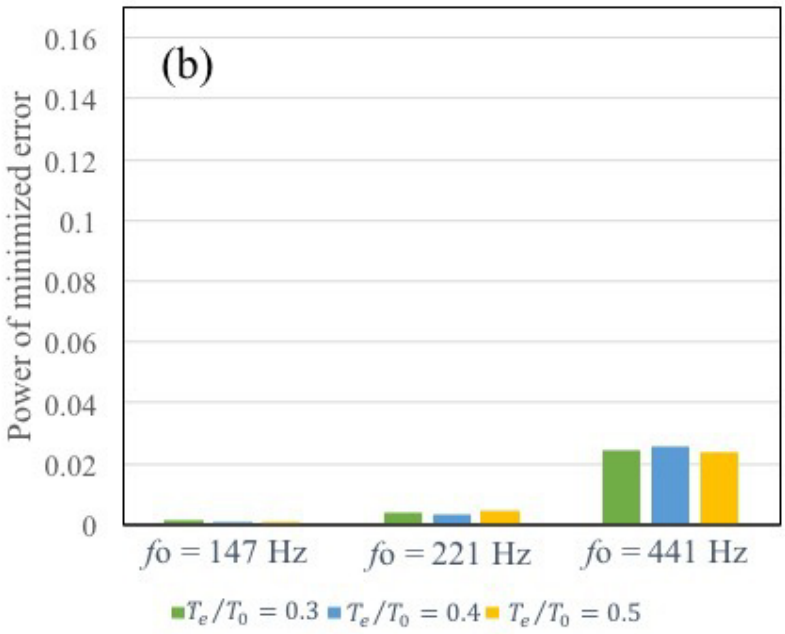
分析した歌声データは，京都市立芸術大学の津崎研究室提供のデータである．歌声とともに，同時収録した EGG 信号がふくまれている．この分析実験では，バリトンの声種の/a/の歌声を用いた（図 4.2(a)）．音程は一定であり，STRAIGHT [44] を用いた分析より fo は 233 Hz である．サンプリング周波数は 44.1 kHz である．

表 4.1: ARX-LF モデルパラメータの平均誤差率 [%]

fo	T_p	T_e	T_a	T_c	E_e	Fr1	Fr2
147 Hz	4.28	3.13	30.6	46.0	70.7	5.63	1.11
221 Hz	6.07	4.38	29.3	46.0	48.8	5.80	1.43
441 Hz	5.11	3.91	33.3	46.0	15.3	8.84	1.78



(a) 先行研究の推定方法による結果



(b) 本研究の推定方法による結果

図 4.1: シミュレーションデータの残差 $e(n)$ の最小二乗誤差 $\varepsilon(n)$

4.3.2 分析結果

図 4.2 に分析した歌声の音声波形と，その非周期波形 $e(n)$ の推定結果を示す．図 4.2(b) は先行研究 [24] の推定方法による結果，図 4.2(c) は本研究の推定結果による結果である．図 4.2(b) と図 4.2(c) を比較すると，図 4.2(c) には図 4.2(b) に見られるような周期成分は見られない．つまり，本研究の推定結果は声帯音源波形中の周期成分と非周期成分が分離されたことを示している．

歌声の分析実験の結果より，本研究の推定方法による声帯音源波形中の周期成分と非周期成分の分離が確認された．

4.4 まとめ

本研究で提案する，歌声の声帯音源波形と声道形状の推定方法の評価実験を行った．シミュレーションデータの分析実験より，基本周波数が高い歌声において推定誤差の減少が確認された．歌声の分析実験より，本研究の推定方法は，声帯音源波形中の周期成分と非周期成分の分離が確認された．

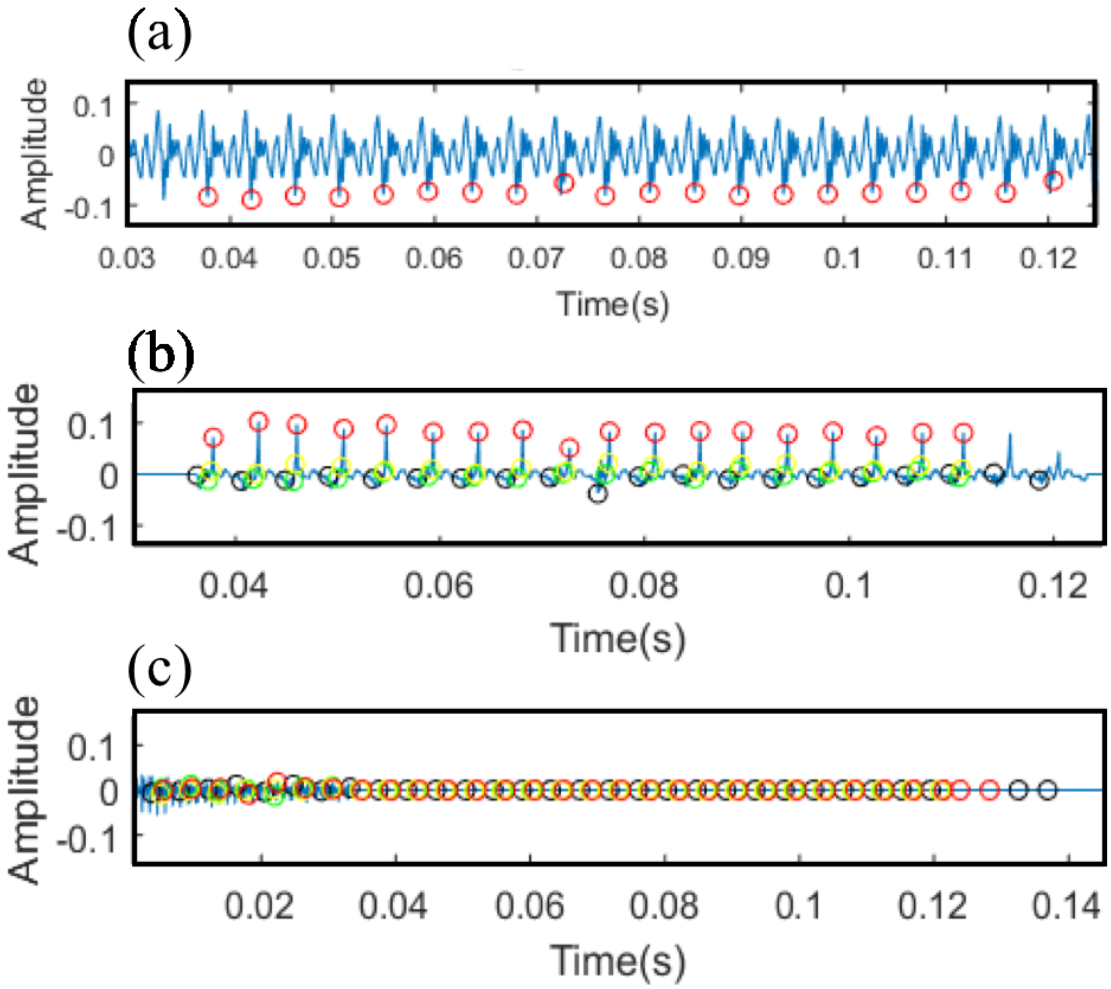


図 4.2: バリトンの歌声/a/の非周期波形 $e(n)$ の推定結果，(a) 歌声の音声波形，(b) 先行研究の推定方法による結果，(c) 本研究の推定方法による結果

第5章　歌声の声区と声区変換部分の分析

5.1　はじめに

本研究の推定方法を用いて，声区ごとの声帯音源波形の特性と，声区変換での声帯音源波形と声道形状の時間変化を観察する．

まず，声区ごとの声帯音源特性に関する知見と，特性を表現するパラメータについて述べる．そして，modal と falsetto の声帯音源波形の推定結果について述べる．最後に，声区変換を含む声帯音源波形の時間変化の推定結果について述べる．

5.2　声区ごとの声帯音源特性

声区ごとの声帯振動様式の違いについて，次のような知見がある．音響的には，modal の喉頭音源のスペクトル傾斜は-12 dB/oct であるのに対し，falsetto では modal より傾斜が急峻となることがわかっている [1]．これは，声帯の緊張と弛緩に関連し，falsetto と比較してスペクトル傾斜が緩やかな modal の声帯が緊張しているといえる [8]．また，声門開口時間率は，声帯が緊張すると値が小さくなることがわかっている [11]．これらの知見より，modal は falsetto と比較して声帯が緊張し，声門開口時間率が小さいと考えられる．また，声門開口時間率は，modal は 0.3-0.8，falsetto は 0.5-0.95 であることがわかっている [11]．声門の閉鎖について，modal は完全閉鎖であるが，falsetto はしばしば定常的な間隙が存在する [8]．また，falsetto では声門辺縁部の限局的な振動が見られる [13]．

声帯音源波形の特性を表すパラメータとして，声門開口時間率 O_q，微分声帯音源波形の声門開口区間の左右対称性 α_m，声門完全閉鎖までに要する戻り区間の時間率 Q_a がある [24]．O_q, α_m, Q_a は式 5.15.25.3 で定義される．

$$O_q = \frac{T_e}{T_0} \tag{5.1}$$

$$\alpha_m = \frac{T_p}{T_e} \tag{5.2}$$

$$Q_a = \frac{T_a}{(1 - O_q)T_0} \tag{5.3}$$

T_e, T_0, T_p, T_a は LF モデルのパラメータである．O_q は，1 周期中で声門が開いている割合を表す．α_m は，声門の開き閉じの速さの比率を表す．声門抵抗・声帯緊張度が小さいと α_m は小さくなる．Q_a は，声門閉鎖の強さを表す．閉鎖が弱い（部分閉鎖）であれば，Q_a は大きくなる．

知見とこれらのパラメータを照らしわせると，式 5.4 の関係が成り立つ．

$$\begin{aligned} O_q: &\quad modal < falsetto \\ \alpha_m: &\quad modal > falsetto \\ Q_a: &\quad modal < falsetto \end{aligned} \tag{5.4}$$

5.3 各声区ごとの歌声の分析および結果

5.3.1 分析対象

分析した歌声データは，京都市立芸術大学の津崎研究室提供のデータである．歌声とともに，同時収録した EGG 信号がふくまれている．この分析実験では，プロの歌唱者 1 名のバリトンの声種の/a/の歌声とテノールの声種の/a/の歌声を用いた．図 5.1 に音声波形と fo の時間変化を示す．音程は一定であり，STRAIGHT [44] を用いた分析より平均 fo はそれぞれ 233 Hz と 430 Hz であった．サンプリング周波数は 44.1 kHz である．

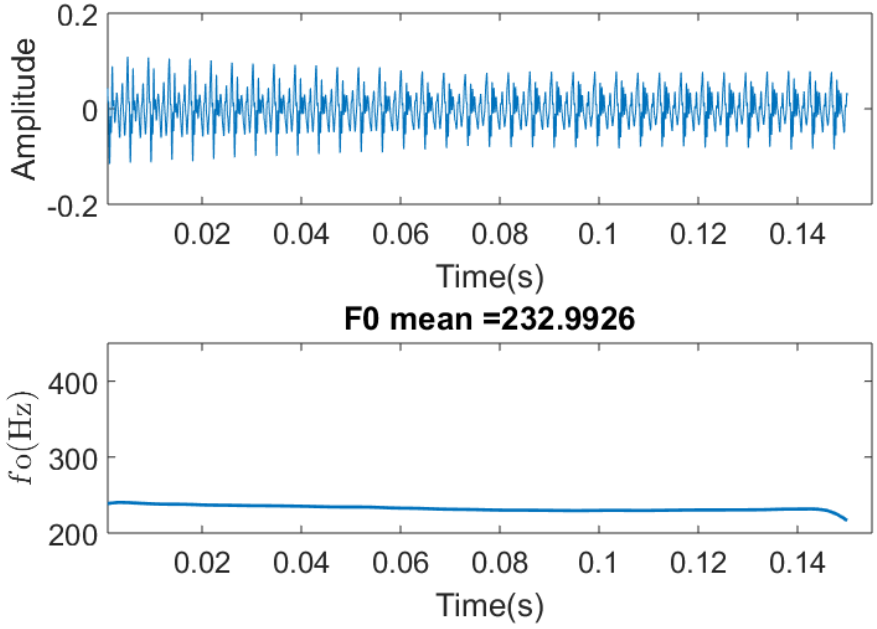
男性の声区変換点が 261.6 − 329.6 Hz の音高に存在する [8] ことから，バリトンの歌声データを modal，テノールの歌声データを falsetto とする．

5.3.2 分析結果

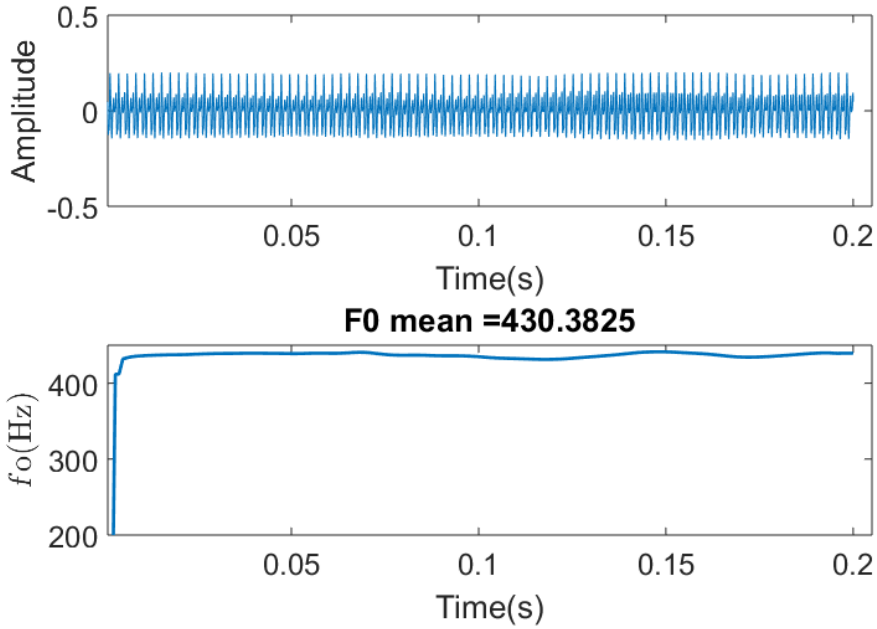
各データの分析結果を表 5.1 に示す．この結果より，falsetto の O_q が modal の O_q より大きく，falsetto の α_m が modal の α_m より小さく，falsetto の Q_a が modal の Q_a より大きいことがわかる．これらの結果は，知見から得られた声帯音源特性の式 5.4 と一致する．したがって，本研究の声帯音源波形と声道形状の推定方法によって，声区に関連したの声帯音源特性の分析が十分可能であるといえる．

表 5.1: ARX-LF モデルパラメータを声区ごとに分析した平均値

声区	O_q	α_m	Q_a
Modal	0.301	0.927	0.0826
Falsetto	0.585	0.872	0.0888



(a) バリトン（Modal）



(b) テノール（Falsetto）

図 5.1: 各声区ごとの歌声の分析で用いた歌声の音声波形と fo

5.4 声区変換を含む歌声の分析および結果

5.4.1 分析対象

分析した歌声データは，京都市立芸術大学の津崎研究室提供のデータである．歌声とともに，同時収録したEGG信号がふくまれている．この分析実験では，プロの歌唱者1名のテノールの声種の/a/の歌声を用いた．図5.2に音声波形とfoの時間変化を示す．低い音程から高い音程へ変化する歌声であり，STRAIGHT [44] を用いた分析よりfoはそれぞれ289 Hzから433 Hzへ変化，280 Hzから418 Hzへ変化が見られた．foが289 Hzから433 Hzへ変化するデータをテノールA，foが280 Hzから418 Hzへ変化するデータをテノールBと呼称する．各データの前半をmodal，後半をfalsettoととする．サンプリング周波数は44.1 kHzである．

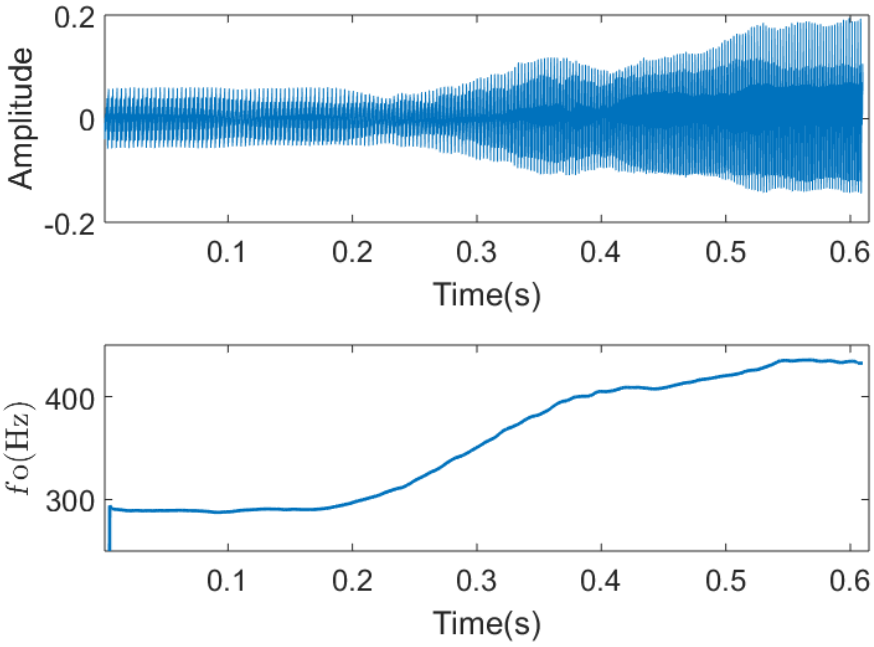
5.4.2 分析結果および考察

図5.3にテノールA，図5.5にテノールBのO_q, α_m, Q_aの推定結果を示す．横軸は周期番号，縦軸は各パラメータの値である．O_q, α_m, Q_aは，推定結果を黒い記号で表し，2次の多項式曲線近似をそれぞれ赤線，青線，緑線で表した．そして図5.4にテノールA，図5.6にテノールBの第1ホルマントと第2ホルマントの推定結果を示す．横軸は周期番号，縦軸は周波数である．

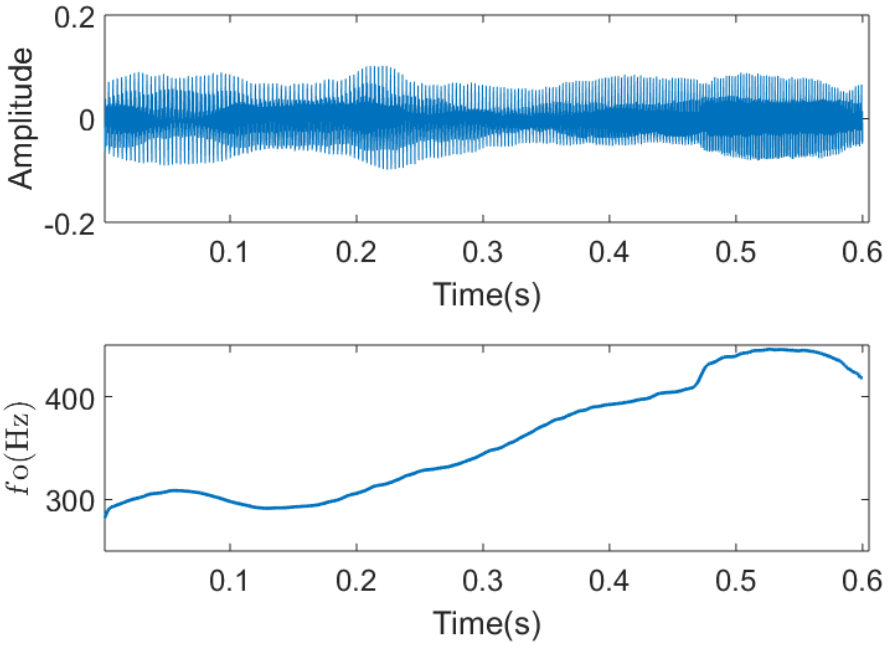
図5.3より，テノールAにおいて，O_qの滑らかな増加，α_mの減少が見られる．Q_aについてはあまり変化が見られない．表5.1のQ_a（0.08-0.09程度）と比較して，テノールAのQ_aの値は，0.1以上である．これは声門閉鎖がかなり弱いことを示している．したがって，テノールAの歌声データは，O_qの滑らかな増加，α_mの減少より，modalからfalsettoへの声区変換が起きていることを表し，modalの時点で声門閉鎖が十分弱いとQ_aは変化しないことがわかった．

図5.3における周期番号50までのα_m, Q_aのばらつきについては，図5.4より，周期番号50までの声道形状の推定結果のばらつきによるものであることがわかる．本研究の声道形状の推定方法は，ARXモデルを用いたものであるため，LPC法に見られるような倍音へのホルマントの引き寄せが起きる可能性がある．図5.4での現象は，これによるものであると考えられる．

図5.5より，テノールBにおいて，α_mの減少，Q_aの増加がみられる．O_qについてはあまり変化が見られない．表5.1のmodalのO_q（0.3）と比較して，テノールBのO_qの値は非常に大きい．Henrichらによれば，modalは0.3-0.8の範囲であること [11] から，テノールBのO_qはmodalの範囲内である．O_q0.5以上は，falsettoのO_qの範囲内でもある．したがって，テノールAの歌声データは，modalからfalsettoへの声区変換が起きているということができる．また，modalの時点でO_qが十分大きいとO_qは変化しないことがわかった．



(a) テノール A



(b) テノール B

図 5.2: 声区変換を含む歌声の分析で用いた歌声の音声波形と f_0

図 5.5 における周期番号 100 以降の α_m, Q_a のばらつきについては，図 5.4 の周期番号 100 以降のホルマントが一部を除いてばらついていないことから，α_m, Q_a のばらつきは確からしいものであるといえる．前述した知見より，falsetto において，声帯の開閉は安定しないことがわかっている [8,13]．よって，falsetto へ変換していくにつれた値のばらつきは，声帯の緊張度の変動，声門閉鎖の弱さの変動が起きていることと考えられる．テノール A，B の分析結果より，modal から falsetto への声区変換において，O_q, α_m, Q_a に急激な変動はなく，ほぼ滑らかに増減することがわかった．また，modal の時点で falsetto に十分な値を取っている場合，値に変化が見られなくなることがわかった．このことより，falsetto を終着点としていた場合，歌唱者は modal の時点で falsetto に近い発声をする可能性があるといえる．

5.5 まとめ

各声区の歌声と，声区変換を含む歌声の分析を行った．各声区の歌声の分析から，声区に関する声帯音源特性の知見と類似した結果を得た．声区変換を含む歌声の分析から，声帯音源波形の特性の滑らかな変化と，falsetto のような特性を modal の時点で持つ場合があることを確認した．

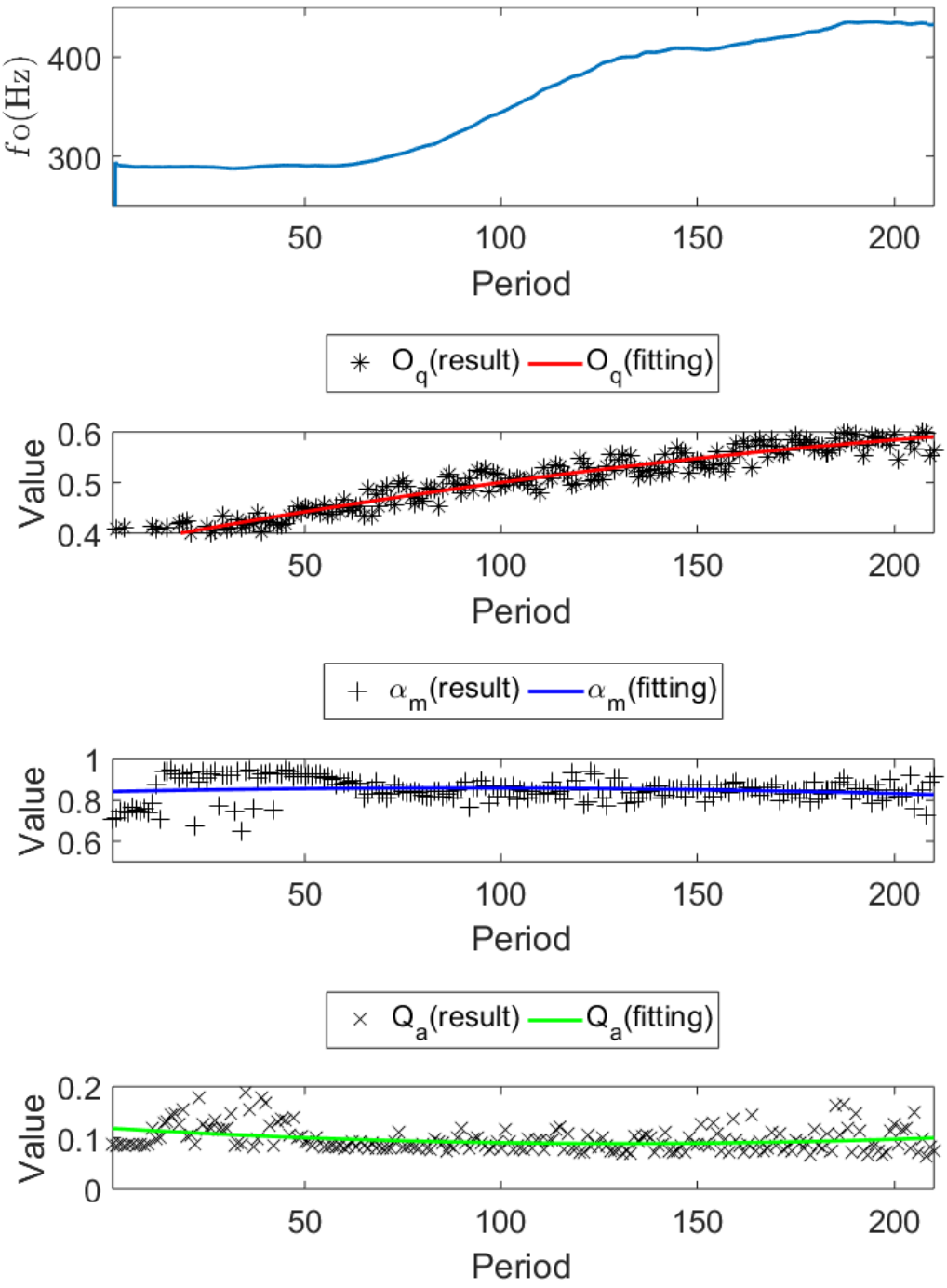


図 5.3: テノール A の O_q, α_m, Q_a の推定結果

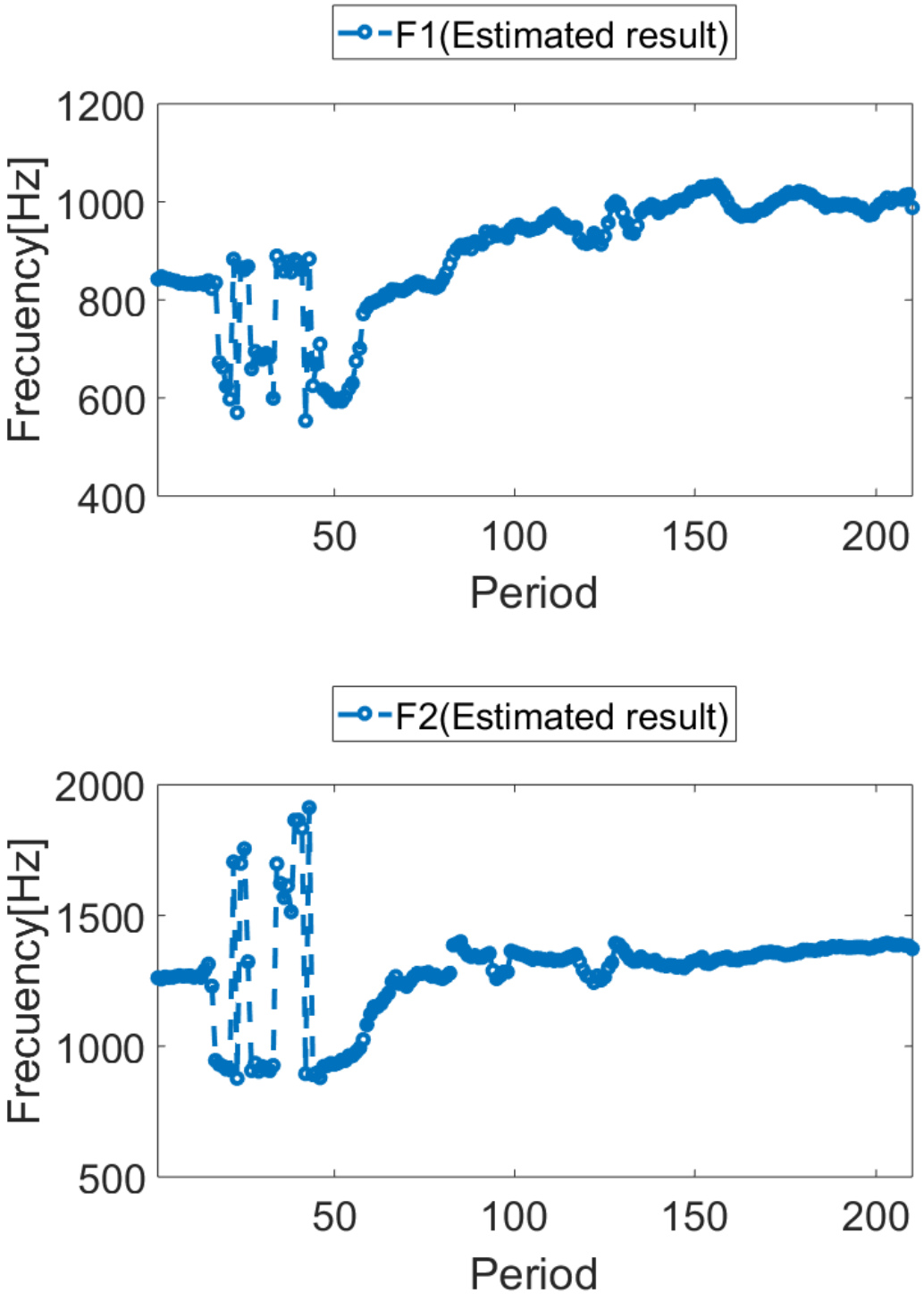


図 5.4: テノール A の第 1 ホルマント F1 と第 2 ホルマント F2 の推定結果

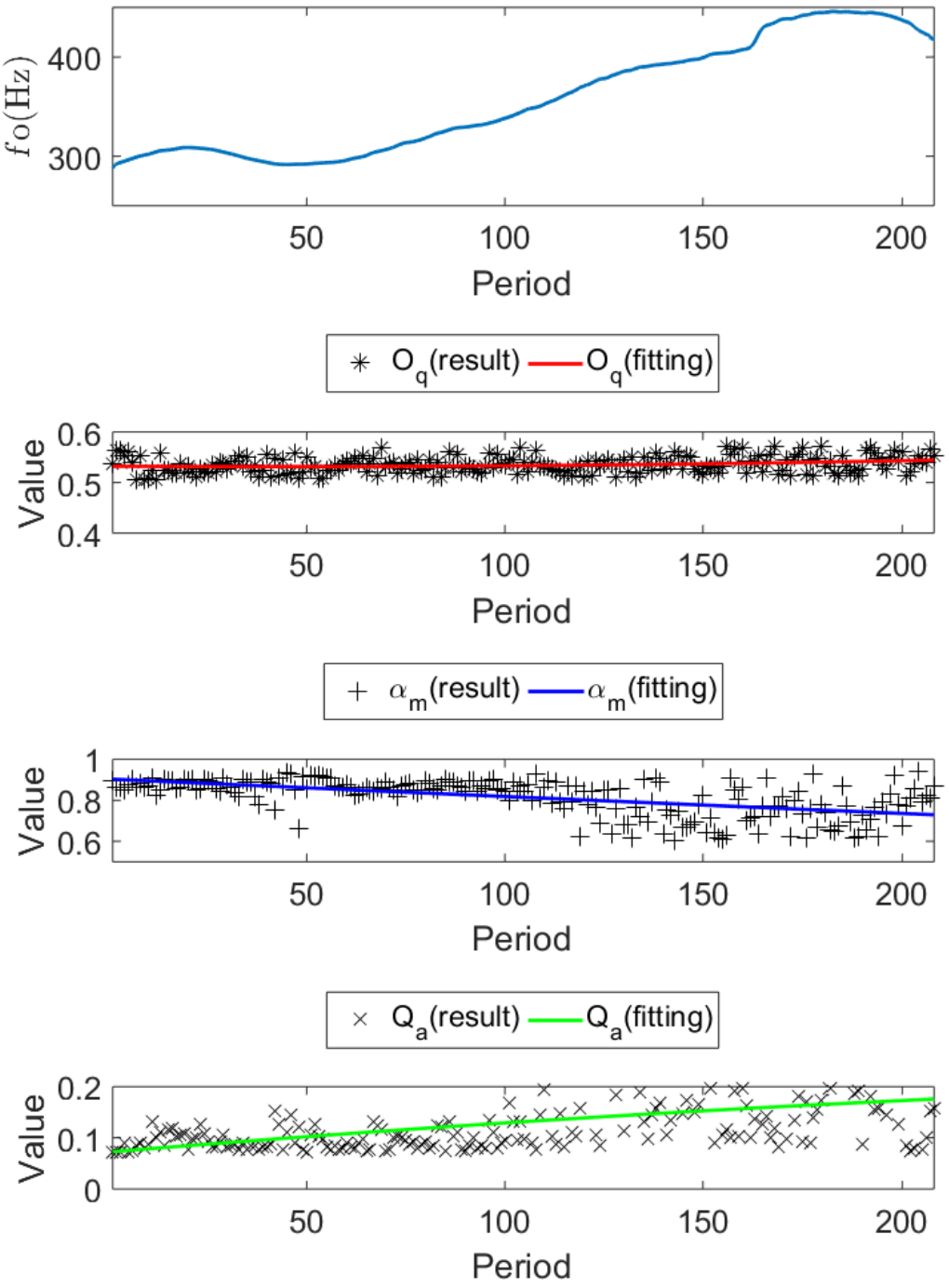


図 5.5: テノール B の O_q, α_m, Q_a の推定結果

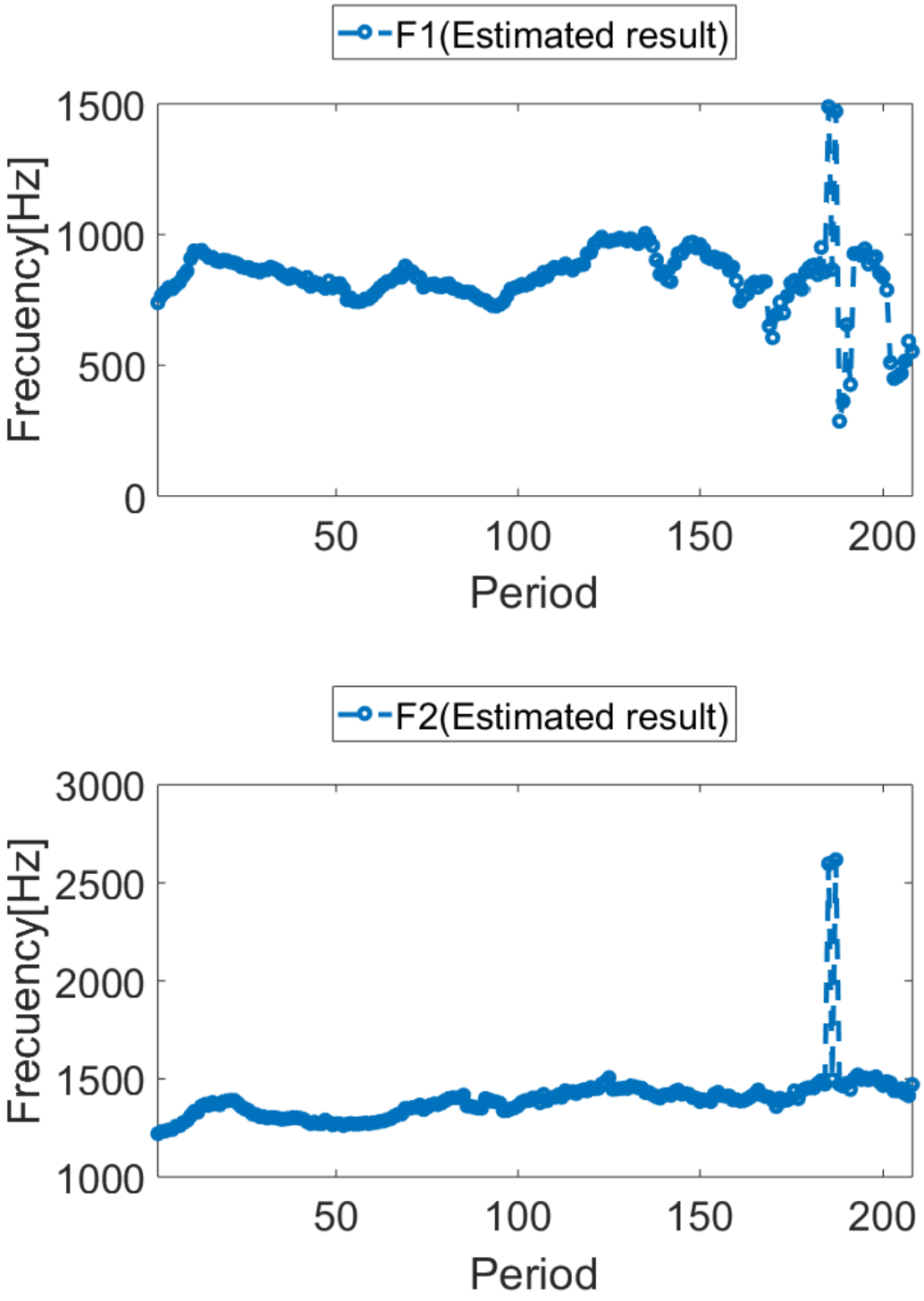


図 5.6: テノール B の第 1 ホルマント F1 と第 2 ホルマント F2 の推定結果

第6章　結論

6.1　本研究でわかったこと

ヒトの表現豊かな歌声の分析と計算機による模擬のために，様々な音域の歌声に対応可能な声帯音源波形と声道形状の推定方法を提案した．ヒトの歌唱において，声質表現の中で重要な要素として声区と声区変換がある．数々の知見から，声帯音源特性が声区を特徴づける要因であることがわかっている．歌声の声区変換の計算機による模擬のためには，声帯音源波形の時間変化を精度よく分析する必要がある．しかし，先行研究では，基本周波数が高い歌声における声帯音源波形と声道形状の推定精度の低さ，声帯音源波形中の周期成分と非周期成分の不完全な分離という 2 つの問題を抱えていた．本研究の推定方法では，1 つ目の問題に対して，最小二乗法を用いた声道フィルタフィッティング，前の周期の応答の加算分補正した歌声の再合成による声帯音源波形の時間変化への対応，2 つ目の問題に対して，サンプリング周波数 44.1 kHz での周期波形の推定，EGG 信号と全探索法と Simulated annealing 法を用いた最適化を用いたパラメータ値探索によって解決できることを示した．

本研究の推定方法について，先行研究が抱える 2 つの問題の解決を確認するため，シミュレーションデータと歌声データを用いた評価実験を行った．シミュレーションデータの分析実験の結果から，基本周波数の高い歌声データを含めたすべてのデータにおいて，声帯音源波形と声道形状の誤差の減少がみられた．この結果より，1 つ目の問題が解決されたことを確認した．歌声データの分析実験の結果から，声帯音源波形中の周期成分と非周期成分の明確な分離がみられた．この結果より，2 つ目の問題が解決されたことを確認した．これらの評価実験により，本研究の声帯音源波形と声道形状の推定方法は，先行研究の問題点を解決できることが実証された．

幅広い音域を持つ歌声の声帯音源波形と声道形状の時間的変動の推定を検証するために，本研究の推定方法を用いた歌声の声区と声区変換部の声帯音源特性の分析を行った．modal と falsetto の歌声の声帯音源波形の分析結果から，声区ごとの声帯音源特性に関する知見との一致が確認された．これは，本研究の推定方法によって，声区に関する声帯音源特性が観察可能であることを示している．modal から falsetto へ声区変換する歌声の声帯音源波形の分析から，声区変換での滑らかに時間変化する特性，条件によって変化しない特性が確認された．このことから，声区変換での声帯音源特性の滑らかな時間変化があること，声区変換する際，falsetto のような特性を modal の時点で持つ場合があることがわかった．

6.2 波及効果

本研究の推定方法は推定精度が高いため，非周期波形 $e(n)$ を歌声に含まれる声帯や口唇でのノイズとみなすことができる．よって，歌声に含まれる声帯や口唇でのノイズの推定が可能であるといえる．それにより，声区にとどまらず，その他の気息性の高い声質や子音の分析や合成に十分活用できる．また，ARX-LF モデルパラメータ値の推定誤差が十分小さいため，モデル作成時のデータ数も少数で済み，音声の声帯音源波形と声道形状のモデル化が非常に容易となる．そのため，自然な音声合成や高品質な歌声合成への寄与も大きいと考えられる．声帯音源波形と声道形状の高精度な推定は，声楽などの音楽教育への寄与や，音声学や音韻学への実験や研究への活用が期待される．

6.3 残された課題

より高精度な分析・表現力豊かな歌声合成を実現するためには，次のような問題点や課題がある．

6.3.1 声帯音源波形と声道形状の推定方法に関する問題点

計算量の多さ

ARX-LF モデルパラメータ値の探索において，全探索法と Simulated annealing 法を用いる．全探索法では，各パラメータ 1 個 1 個に対して全探索を行う．Simulated annealing 法では，全探索法からコスト最小のパラメータ値の組から上位最大 15 組の最適化を行い，その中で最小コストの組を最終解としている．このように，計算量が多いため，長いデータの分析ではかなりの分析時間を要する上，メモリの圧迫を引き起こす．計算量の削減が求められる．

ビブラートを含む歌声の声帯音源波形と声道形状の推定

ビブラートでは，T_0 の値が激しく変動する．T_0 の変動によって，推定精度の低下が確認されている．歌声の声帯音源波形と声道形状の推定の頑健性のために，T_0 の変動による推定誤差発生の原因究明と解決方策の検討が必要となる．

6.3.2 声区変換を含む歌声の合成へ向けた課題

声区変換部の分析結果

歌声の声区変換部の分析において，本研究では連続的に変換した歌声を用いた．声区の変換点で不連続になった歌声などとの比較が必要である．

分析データ数

歌声の声区変換部の分析において，本研究で用いたデータ数は 2 個である．声区変換部での声帯音源特性の存在は確認できたが，個人性による偏りがある可能性も少なくない．歌声合成のためのモデル化の作成には，歌声データを増やし，さらに分析を進める必要がある．

分析データの選出

歌声の声区変換については，ベルティング [45] という歌唱方法や，中声区を用いる歌唱方法 [8] がある．歌声データの定義や分類が必要である．また，データ収録の際に注意すべき事柄である．

声帯ノイズの合成

本研究の推定方法は，合成による分析（Analysis by synthesis）であるため，歌声の合成も可能である．しかし，声帯ノイズの合成の際には，推定された非周期波形 $e(n)$ から声帯からのノイズと口唇からのノイズを分離する必要がある．声帯ノイズの合成のためのモデルの作成，あるいは合成方法の検討が必要である．

謝辞

本研究を進めるにあたり，多大なる御指導ならびに御鞭撻を賜りました赤木 正人 教授に深く感謝致します.
本研究を進めるにあたり，日頃から熱心な御指導ならびに御鞭撻を賜りました鵜木 祐史 教授に心より感謝致します.
本研究を進めるにあたり，熱心に御討論頂き，また御助言を賜りました党 建武 教授に心より感謝致します.
歌声データを提供していただきました京都市立芸術大学 津崎 実 教授 および 博士後期課程2年 高橋 純 氏に心より感謝いたします.
本研究を進めるにあたり，日頃から熱心な議論と様々な御助言御助力をいただきました，博士後期課程3年 李 永偉 氏に深く感謝いたします.
また，本研究を進めるにあたり，日頃から熱心な議論と激励をいただきました，音情報処理分野の諸先輩方，及び諸氏に厚く御礼申し上げます.
最後に，本学での研究生活を支え，温かく見守ってくれた両親に心から感謝致します.

参考文献

[1] Johan Sundberg. *The Science of the Singing Voice*. Northern Illinois Univ Pr, 2 1987.

[2] 粕谷英樹, 楊長盛. 音源から見た声質 (小特集—声質:音声言語の多様性に迫る—). 日本音響学会誌, Vol. 51, No. 11, pp. 869–875, 1995.

[3] 今泉敏. 声質の計量心理学的評価 (小特集—計量心理学の音響学への応用—). 日本音響学会誌, Vol. 42, No. 10, pp. 828–833, 1986.

[4] J. Laver. *The Phonetic Description of Voice Quality*. Cambridge Studies in Linguistics. Cambridge University Press, 2009.

[5] Ilse Bernadette Labuschagne and Valter Ciocca. The perception of breathiness: Acoustic correlates and the influence of methodological factors. *Acoustical Science and Technology*, Vol. 37, No. 5, pp. 191–201, 2016.

[6] D. H. Klatt and L. C. Klatt. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Acoustical Society of America Journal*, Vol. 87, pp. 820–857, feb 1990.

[7] Hui-Ling Lu and JO Smith. Estimating glottal aspiration noise via wavelet thresholding and best-basis thresholding. In *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, pp. 11–14. IEEE, 2001.

[8] 榊原健一. 世界の歌唱法：様々な歌唱様式における supranormal な声 (小特集—歌声の科学—). 日本音響学会誌, Vol. 70, No. 9, pp. 499–505, 2014.

[9] Harry Hollien. On vocal registers. *Journal of Phonetics*, Vol. 2, pp. 125–143, 1972.

[10] 森下亮祐, 齋藤毅, 三好正人. 歌声の地声と裏声の切り替え方法の検討. 聴覚研究会資料, Vol. 43, No. 7, pp. 565–570, oct 2013.

[11] Nathalie Henrich, Christophed'Alessandro, Boris Doval, Michle Castellengo. Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency. *The Journal of the Acoustical Society of America*, Vol. 117, No. 3, pp. 1417–1430, 2005.

[12] Hiroshi Imagawa, Ken-Ichi Sakakibara, Isao T Tokuda, Mamiko Otsuka, and Niro Tayama. Estimation of glottal area function using stereo-endoscopic high-speed digital imaging. In *Eleventh Annual Conference of the International Speech Communication Association*, 2010.

[13] Ken-Ichi Sakakibara, Hiroshi Imagawa, Miwako Kimura, Hisayuki Yokonishi, and Niro Tayama. Modal analysis of vocal fold vibrations using laryngotopography. In *Eleventh Annual Conference of the International Speech Communication Association*, 2010.

[14] 大谷圭介. 声区転換部を含むオペラ歌唱の音響的特性スペクトル変動に見る音響的指標について. PhD thesis, 京都市立芸術大学, 2014.

[15] Takeshi Saitou, Masashi Unoki, and Masato Akagi. Development of an f0 control model based on f0 dynamic characteristics for singing-voice synthesis. *Speech communication*, Vol. 46, No. 3-4, pp. 405–417, 2005.

[16] Hideki Kenmochi and Hayato Ohshita. Vocaloid-commercial singing synthesizer based on sample concatenation. In *Eighth Annual Conference of the International Speech Communication Association*, 2007.

[17] 徳田恵一, 益子貴史, 小林隆夫, 今井聖. 動的特徴を用いた HMM からの音声パラメータ生成アルゴリズム. 日本音響学会誌, Vol. 53, No. 3, pp. 192–200, 1997.

[18] 大浦圭一郎, 絢美間瀬, 知彦山田, 恵一徳田, 真孝後藤. Sinsy:「あの人に歌ってほしい」をかなえる HMM 歌声合成システム. 情報処理学会研究報告, Vol. 86, No. 1, pp. 1–8, jul 2010.

[19] Gunnar Fant. The source filter concept in voice production. *STL-QPSR*, Vol. 1, No. 1981, pp. 21–37, 1981.

[20] Gunnar Fant, Johan Liljencrants, and Qi-guang Lin. A four-parameter model of glottal flow. *STL-QPSR*, Vol. 4, No. 1985, pp. 1–13, 1985.

[21] Wen Ding, Hideki Kasuya, and Shuichi Adachi. Simultaneous estimation of vocal tract and voice source parameters based on an arx model. *IEICE transactions on information and systems*, Vol. 78, No. 6, pp. 738–743, 1995.

[22] Hui-Ling Lu and Julius O Smith. Joint estimation of vocal tract filter and glottal source waveform via convex optimization. In *Applications of Signal Processing to Audio and Acoustics, 1999 IEEE Workshop on*, pp. 79–82. IEEE, 1999.

[23] Hui-Ling Lu. *Toward a high-quality singing synthesizer with vocal texture control*. PhD thesis, 2002.

[24] 元田紘樹, 赤木正人. 声区の違いによる声質の変化と声帯音源特性の関連性. Vol. 42, No. 7, pp. 585–590, 2012.

[25] 元田紘樹, 赤木正人. 声区表現可能な歌声合成を目的とした ARX-LF パラメータの制御法の検討. 聴覚研究会資料, Vol. 43, No. 1, pp. 37–42, feb 2013.

[26] 大塚貴弘. ARX 音声生成モデルに基づく音声分析合成法に関する研究. PhD thesis, 宇都宮大学, 2002.

[27] Gunnar Fant. The lf-model revisited. transformations and frequency domain analysis. *Speech Trans. Lab. Q. Rep., Royal Inst. of Tech. Stockholm*, Vol. 2, No. 3, p. 40, 1995.

[28] Qiang Fu and Peter 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.

[29] J.D. Markel and A.H. Jr. Gray. *Linear Prediction of Speech (Communication and Cybernetics)*. Springer, 3 2013.

[30] 大塚貴弘, 粕谷英樹. 音源パルス列を考慮した頑健な ARX 音声分析法. 日本音響学会誌, Vol. 58, No. 7, pp. 386–397, 2002.

[31] 粕谷英樹. 音声分析技術の最近の進歩. 喉頭, Vol. 14, No. 2, pp. 57–63, 2002.

[32] Takahiro Ohtsuka and Hideki Kasuya. An improved speech analysis-synthesis algorithm based on the autoregressive with exogenous input speech production model. In *Sixth International Conference on Spoken Language Processing*, 2000.

[33] Damien Vincent, Olivier Rosec, and Thierry Chonavel. A new method for speech synthesis and transformation based on an arx-lf source-filter decomposition and hnm modeling. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, Vol. 4, pp. IV–525. IEEE, 2007.

[34] Damien Vincent, Olivier Rosec, and Thierry Chonavel. Estimation of lf glottal source parameters based on an arx model. In *Ninth European Conference on Speech Communication and Technology*, 2005.

[35] Damien Vincent, Olivier Rosec, and Thierry Chonavel. Glottal closure instant estimation using an appropriateness measure of the source and continuity constraints. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, Vol. 1, pp. I–I. IEEE, 2006.

[36] Yannis Stylianou. Applying the harmonic plus noise model in concatenative speech synthesis. *IEEE Transactions on speech and audio processing*, Vol. 9, No. 1, pp. 21–29, 2001.

[37] Hui-Ling Lu and Julius O Smith III. Glottal source modeling for singing voice synthesis. In *ICMC*, 2000.

[38] Hiroki Motoda and Masato Akagi. A singing voices synthesis system to characterize vocal registers using arx-lf model. In *2013 International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP'13)*, pp. 93–96. 2013 International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP'13), 2013.

[39] Rudolf Emil Kalman. A new approach to linear filtering and prediction problems. *Trans. ASME-Journal of Basic Engineering*, Vol. 82, No. 1, pp. 35 – 45, 1960.

[40] Yongwei Li, Ken-Ichi Sakakibara, Daisuke Morikawa, and Masato Akagi. Commonalities of glottal sources and vocal tract shapes among speakers in emotional speech. In *The 11th International Seminar on Speech Production (ISSP 2017)*. The 11th International Seminar on Speech Production (ISSP 2017), 2017.

[41] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, Vol. 220, No. 4598, pp. 671–680, 1983.

[42] V. Černý. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of Optimization Theory and Applications*, Vol. 45, No. 1, pp. 41–51, Jan 1985.

[43] Hideki Kawahara, Ken-Ichi Sakakibara, Hideki Banno, Masanori Morise, Tomoki Toda, and Toshio Irino. Aliasing-free implementation of discrete-time glottal source models and their applications to speech synthesis and f0 extractor evaluation. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2015 Asia-Pacific*, pp. 520–529. IEEE, 2015.

[44] Hideki Kawahara. Straight, exploitation of the other aspect of vocoder: Perceptually isomorphic decomposition of speech sounds. *Acoustical Science and Technology*, Vol. 27, No. 6, pp. 349–353, 2006.

[45] Johan Sundberg, Patricia Gramming, and Jeanette Lovetri. Comparisons of pharynx, source, formant, and pressure characteristics in operatic and musical theatre singing. *Journal of Voice*, Vol. 7, No. 4, pp. 301 – 310, 1993.

研究業績

本研究に関する研究業績

国際会議における発表

(口頭，査読有)

1. Kyoko Takahashi and Masato Akagi, “Estimation of glottal source waveform and vocal tract shape for singing-voice analysis,” 2018 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP’18), 7PM2–1–6, Hawaii, USA, March, 2018.

その他の研究業績

学術雑誌に発表した論文

(査読有)

1. Kyoko Takahashi and Daisuke Morikawa, “Horizontal localization of sound image and sound source in monaural congenital deafness,” *Journal of Signal Processing*, Research Institute of Signal Processing, Vol. 21, No. 4, pp. 167–170, 2017.

国際会議における発表

(口頭，査読有)

1. Kyoko Takahashi and Daisuke Morikawa, “Horizontal localization of sound image and source in monaural congenital deafness,” 2017 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP’17), 2PM1–3–5, Guam, USA, March, 2017.

国内学会における発表

(口頭，査読無)

1. 高橋響子, 森川大輔, 「先天性単耳受聴者の水平面における音像定位と音源定位」, 日本音響学会 2017 年春季研究発表会, 2–1–7, 神奈川, 2017 年 3 月.

その他の業績

(受賞)

1. Kyoko Takahashi, Student Paper Award (2017 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing), Mar. 2017.

2. 高橋響子, 日本音響学会北陸支部優秀学生賞, 3 月, 2018.