

Title	歌声知覚・生成機構の解明に向けた歌声合成システム構築に関する研究
Author(s)	齋藤, 毅
Citation	
Issue Date	2006-03
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/983">http://hdl.handle.net/10119/983</a>
Rights	
Description	Supervisor:赤木 正人, 情報科学研究科, 博士

# 博士論文

## 歌声知覚・生成機構の解明に向けた歌声合成システム構築に関する研究

指導教官 赤木 正人 教授

北陸先端科学技術大学院大学  
情報科学研究科情報処理学専攻

齋藤 毅

2006年3月

## 要旨

本論文では、「人は如何にして歌声を聴き、発声するのか？」という問いに対する解答を得ることを目指して、歌声に共通して含まれる音響的特徴に着目し、高品質な歌声合成システムの構築を行うことで、歌声知覚において重要な役割を担っている音響的特徴を明らかにした。その結果、歌声の基本周波数変化においては、オーバーシュート、ヴィブラート、プレパレーション、微細変動の4つの成分が多く、歌声に共通して含まれる成分であり、且つ歌声知覚に大きな影響を与えていることが明らかとなった。また、歌声のスペクトルにおいては、singers' formant と呼ばれる 3 kHz 付近のスペクトル包絡の顕著なピークと、ヴィブラートに同期したホルマントの振幅変調が歌声知覚に影響を与えていることが明らかとなった。また、話声と歌声の違いを規定し、且つ歌声知覚に最も大きな影響を与えている音響的特徴は、メロディ変化中に存在する F0 動的変動成分であることが明らかとなった。

# 目次

<b>1</b>	<b>序論</b>	<b>1</b>
1.1	はじめに . . . . .	2
1.2	本研究の背景 . . . . .	3
1.2.1	歌声特有の基本周波数特性に関する研究 . . . . .	3
1.2.2	歌声特有のスペクトル特性に関する研究 . . . . .	6
1.2.3	歌声合成に関する研究 . . . . .	7
1.3	本論文の目的 . . . . .	9
1.4	本論文の構成 . . . . .	10
<b>2</b>	<b>歌声合成システムの構成</b>	<b>13</b>
2.1	はじめに . . . . .	14
2.2	歌声合成システム構築の方略 . . . . .	14
2.2.1	本論文で提案する歌声合成システムの構成 . . . . .	15
2.2.2	基本周波数制御法構築の方略 . . . . .	15
2.2.3	スペクトル制御法構築の方略 . . . . .	19
2.2.4	音韻長制御法構築の方略 . . . . .	19
2.3	STRAIGHT . . . . .	20
2.4	本研究で扱う歌声データ . . . . .	21
2.5	まとめ . . . . .	22
<b>3</b>	<b>歌声の基本周波数制御モデルの構築</b>	<b>24</b>
3.1	はじめに . . . . .	25
3.2	歌声の F0 動的変動成分の抽出 . . . . .	25
3.2.1	基本周波数推定 . . . . .	25

3.2.2	歌声データ	26
3.2.3	F0 動的変動成分	26
3.3	歌声の F0 制御モデルの構築	30
3.3.1	歌声 F0 制御モデルの枠組み	31
3.3.2	F0 動的変動成分の制御	32
3.3.3	基本周波数変化の記述過程	37
3.4	歌声 F0 制御モデルの検証	39
3.4.1	フィッティング処理	39
3.4.2	F0 動的変動成分の正確さに関する検証	41
3.5	まとめ	44
4	F0 動的変動成分と歌声知覚の関係の調査	46
4.1	はじめに	47
4.2	歌声合成システムの構築	48
4.2.1	Klatt formant synthesizer を用いた歌声合成システム	48
4.2.2	STRAIGHT を用いた歌声合成システム	49
4.3	F0 動的変動成分が歌声知覚に与える影響 1	
-	F0 動的変動成分を除去した場合の歌声知覚反応-	52
4.3.1	歌声合成音の作成	52
4.3.2	聴取実験	54
4.3.3	実験結果と考察	58
4.4	F0 動的変動成分が歌声知覚に与える影響 2	
-	F0 動的変動成分を付与した場合の歌声知覚反応-	60
4.4.1	Klatt formant synthesizer を用いた歌声合成システムによる 検証	60
4.4.2	STRAIGHT を用いた歌声合成システムによる検証	65
4.5	歌唱技術と F0 動的変動成分の関係	68
4.5.1	歌唱技術の違いによる F0 動的変動成分の特性の変化	68
4.5.2	歌唱技術の違いによる F0 動的変動成分の特性の変化が歌声 知覚に与える影響の調査	70

4.6	F0 動的変動成分の特性と歌声知覚の関係 . . . . .	72
4.6.1	歌声合成音の作成 . . . . .	74
4.6.2	聴取実験 . . . . .	75
4.6.3	実験結果・考察 . . . . .	75
4.7	Vibrato rate の時間変動と歌声知覚の関係の調査 . . . . .	78
4.7.1	Vibrato rate の分析 . . . . .	79
4.7.2	Vibrato rate の時間変動が歌声知覚に与える影響の調査 . . . . .	80
4.8	音高遷移方向の違いによるオーバーシュート特性の差異に関する検証	82
4.8.1	音高遷移方向の違いによるオーバーシュート特性の差異に関する検証 . . . . .	83
4.8.2	オーバーシュート特性の差異が歌声知覚に与える影響 . . . . .	84
4.9	まとめ . . . . .	86
<b>5</b>	<b>歌声特有のスペクトル特性と歌声知覚の関係の調査</b>	<b>89</b>
5.1	はじめに . . . . .	90
5.2	Singers' formant が歌声知覚に与える影響の調査 . . . . .	90
5.2.1	Singers' formant の制御法 . . . . .	91
5.2.2	Singers' formant が歌声知覚に与える影響 . . . . .	92
5.2.3	Singers' formant の特性と歌声知覚の関係の調査 . . . . .	96
5.3	ヴィブラートに同期したスペクトル制御 . . . . .	97
5.3.1	ホルマンントの振幅変調制御 . . . . .	98
5.3.2	ホルマンントの振幅変調が歌声知覚に与える影響の調査 . . . . .	101
5.3.3	振幅変調の偏移幅の大きさと歌声知覚の関係の調査 . . . . .	102
5.4	まとめ . . . . .	104
<b>6</b>	<b>高品質な歌声合成システムの実装</b>	<b>106</b>
6.1	はじめに . . . . .	107
6.2	音韻長制御法の構築 . . . . .	107
6.3	歌声知覚に重要な音響パラメータの検討 . . . . .	110
6.3.1	歌声合成システムの実装 . . . . .	110
6.3.2	歌声合成音の作成 . . . . .	111

6.3.3	聴取実験 . . . . .	113
6.3.4	実験結果と考察 . . . . .	113
6.3.5	全体考察 . . . . .	115
6.4	まとめ . . . . .	116
7	結論 . . . . .	118
7.1	本論文で明らかになったことの要約 . . . . .	119
7.2	今後の課題 . . . . .	124
	謝辞 . . . . .	127
	参考文献 . . . . .	128

# 第 1 章

## 序論

## 1.1 はじめに

「歌を歌う」ことは、「言葉を話す」ことと同様に人間の感情や想いを表出するための最も手近なコミュニケーション手段である。とりわけ、華やかな技巧をちりばめた歌声は、時として話声からは得られないほどの大きな感動を人間に与えることがある。そして、誰もが卓越した歌唱技術を習得し、自分の思い描く歌唱法や音色で歌を歌うことに憧れを抱くものである。では、我々人間は、なぜ歌声を聴くことで感動を覚えるのだろうか。なぜそのような歌声を発声することができるのだろうか。この疑問を突き詰めていくと、“人間は如何にして歌声を聴き、発声するのか？”という大きな問題に直面する。しかし現在、歌声知覚・生成機構を解明し、この問題に対する明確な解答を得るまでには至っていない。

「言葉を話す」ことへの探求は、長年に渡り心理学や生理学といった多角的立場から行われ、現在では多くの知見が得られてきている。そして現在、これらの知見を音声合成に応用する試みが積極的に行われるようになってきている。この取り組みは、単に心理学や生理学からの知見を音声合成に生かすことで合成システムを発展させるだけのものではなく、より高品質な音声の合成を目指すことにより「音声の知覚に重要な音響的特徴は何か?」、「それら音響的特徴を如何にして制御するのか?」といった音声知覚・生成についての知見を与え、それら機構の解明に大きく貢献している。実際に、人間の声帯制御機構の知見を考慮した自然性の高い音声合成や、感情知覚に関する知見を取り入れた感情音声合成の実現などにより、音声知覚・生成機構の一端が明らかになり始めている。

「歌を歌う」事に関しても、心理学や生理学、更には歌唱学からの取り組みにより様々な知見が得られてきている。ならば、これら知見を歌声合成に応用し、音声合成と同様の枠組みで研究を遂行することで、歌声合成の研究の発展はもちろん、歌声知覚・生成機構の解明に繋がる取り組みが可能となるはずである。しかし現在、歌声合成システムの構築はおろか、歌声を合成する為に必要な物理量が何であるのかも殆んど明らかになっていない状況にある。

では、歌声合成に必要な物理量とは何であろうか。歌声と一言にいても、歌唱技術の差や、オペラをはじめとする洋楽歌唱、演歌や長唄等の邦楽歌唱といった歌唱法の違いなどにより、その音響的特徴は多種多様な性質を示す。しかし我々

人間は、歌唱技術に大きな差がある歌声を聴いた時、その技術の差を容易に区別できる一方で、どちらも歌声として知覚している。また、馴染みの無い様な歌声を聴いた場合、その歌唱法を特定するのが困難であっても、それが話声でなく歌声として知覚することも可能である。つまり歌声には、歌唱技術や歌唱法の違いに依存しない歌声共通の音響的特徴が存在し、この特徴が、歌声が歌声として知覚される為に、更には歌声を合成する為に必要な物理量に他ならない。そして、この歌声に共通して含まれる音響的特徴に着目した歌声合成システムを構築し、その中で各特徴と歌声知覚の関係を明らかにする取り組みこそが、歌声知覚・生成機構の解明の足掛かりとなるのである。

以上から、本論文では、歌声に共通して含まれる音響的特徴に着目した歌声合成を行うことで、あらゆる歌声を知覚する上で重要な役割を担っている音響的特徴を明らかにすることを試みる。

## 1.2 本研究の背景

音源フィルタモデル (source-filter model) [19] に基づけば、人間の音声は、声帯振動の情報を表す基本周波数と、声道形状の情報を表すスペクトル形状の2つの音響パラメータによって表現される。そして、歌声(歌唱)と話声(発話)の違いは、これら音響パラメータの特性の違いとして現れる。実際に、話声と歌声それぞれにおける音響パラメータの特性が異なり、その違いを知覚することによって話声と歌声を容易に区別可能なことが先行研究で示されている[14]。その為、歌声知覚・生成機構の解明には、基本周波数とスペクトルの2つの音響パラメータに含まれる歌声特有の音響的特徴に着目した取り組みが重要となってくる。

そこで本節では、歌声特有の基本周波数特性、及びスペクトル特性に関する先行研究を示すと共に、これまで行われてきた歌声合成の研究について概説する。

### 1.2.1 歌声特有の基本周波数特性に関する研究

基本周波数は聴感上の音の高さ(ピッチ)に対応し、その変化はアクセントやイントネーションといった韻律的情報を表す。特に歌声においては、基本周波数の

値が音高値に対応し、連続的な基本周波数変化は楽曲のメロディを表すことから、話声に比べ大きく異なった特性を持つ。そして、この事は、歌唱における声帯振動の制御機構が、通常発話の場合と大きく異なること示している。その為、歌声特有の基本周波数特性に着目した取り組みは、高品質な歌声合成の実現はもちろん、話声に無い歌声特有の知覚・生成機構を解明する為の大きな足掛かりとなり、これまで数多くの研究が行われてきている。

歌声と話声の大きな違いの1つとして、基本周波数変化の概形が、多くの場合、その楽曲のメロディに対応していることが挙げられる。実際、矢田部らは、このメロディ変化を表すステップ状の基本周波数変化そのものが、歌声特有の成分として重要であることを示している [20]。しかし、メロディの変化のように急峻で且つ変動幅の広い基本周波数変化には、通常発話時には見られない動的な変動成分（以後、これを F0 動的変動成分と呼ぶ）が存在することが明らかとなってきた。

Sundberg は、音高変化時の基本周波数の遷移速度を測定することで、音高上昇時の方が下降時に比べ速い速度で基本周波数が変化する事を示している [77, 78]。また、de Krom らは、音高変化直後に目的の音高を越えて揺れる瞬時的な変動成分であるオーバーシュート (Overshoot) の存在を示し、この瞬時的な変動の大きさを測定している [31]。

Seashore らは、基本周波数変化における 4 ~ 7 Hz 程度の周期的変動成分であるヴィブラートの存在を示している [71, 72, 73, 74]。それ以来、多くの研究者によって、主にオペラ歌唱を対象としたヴィブラート分析が積極的に行われてきている [29, 37, 50, 58, 66, 89]。その結果、ヴィブラートは殆んどオペラ歌唱において共通して観測され、歌声の声質において重要な役割を果たしている事が明らかになってきている [78]。また、中山や Kojima は、邦楽歌唱を対象としたヴィブラート分析を行っている [49, 59]。その結果、様々な邦楽歌唱においてもヴィブラート（演歌や長唄等では“ゆり”と呼ぶ場合もある）が存在し、その重要性を示している。

これら歌声特有の F0 動的変動成分の存在を示す研究を受けて、各成分の制御方法を検討する試みや、各成分と歌声知覚との関係を調査する研究が行われてきている。

柏野らは、ボーカル音の分離システム構築の為に、歌声の基本周波数変化の生成

モデルを構築している [4] . ここでは , 話声の基本周波数変化を 2 次系モデルで構築した藤崎モデルに習い , ヴィブラートや音高の上昇・下降時の遷移スピード制御を考慮している .

宇都宮大学の研究グループでは , 様々な  $F_0$  動的変動成分の制御モデルの構築を試みている . その中で , 矢田部らは , 音高変化における基本周波数の上昇・下降の遷移スピード , 及びその直後に現れるオーバーシュート成分の制御によって , 歌声の基本周波数変化のモデル化を行っている [20] . また , 小田切らは , ARMA モデルを用いたヴィブラートの分析 , 及びモデル化を行っている [3] . 更に Mori らは , 上記の 2 つの研究を統合し , 歌声の基本周波数変化を生成するモデルを構築している [54, 55] . この一連の研究では , モデルの構築だけでなく , それを用いた合成音の作成によって , 各特徴の歌声知覚への影響も調査している .

Minematsu らは , 日本伝統歌謡である長唄を対象に , 基本周波数の分析 , 及びその制御法を提案している [52, 53] . ここでは , 長唄特有の変動成分である “あたり” の制御を考慮することで , 長唄の基本周波数変化の生成を実現させていると同時に , それら成分が長唄を知覚する上で重要な役割を担っていることを示している .

Akagi らは , 歌声の基本周波数からオーバーシュートやヴィブラートといった  $F_0$  動的変動成分を除去した後に残る不規則な変動成分に着目し , それらが歌声の自然性に影響を与えている事を示している [24] . また , この成分が , 歌声のみならず , 話声においても重要であることを示している [25] .

これらの研究によって , 話声にはない歌声特有の基本周波数特性として様々な  $F_0$  動的変動成分の存在が明らかになってきている . また , 近年になって , それら成分と歌声知覚との関係を明らかにする試みも行われるようになってきている . しかし , これらの研究の殆んどは , 特定の歌唱法や歌唱者の歌声 , とりわけオペラを中心とした洋楽歌唱を対象としたものである . その為 , 報告されている  $F_0$  動的変動成分が , 様々な歌声に共通して含まれるものなのかは定かではない . 更に , それら成分が特定の歌声を知覚する上で重要なのか , もしくはあらゆる歌声を知覚する上で重要なのかも明確になっていない . また , 個々の  $F_0$  動的変動成分の制御法は検討されているものの , 複数の  $F_0$  動的変動成分を制御することで歌声の基本周波数を自由に記述可能なモデルは構築されていない . この為 , 個々の  $F_0$  動的変

動成分と歌声知覚の関係を定量的に調査する試みは十分に行えていないのが現状である。

### 1.2.2 歌声特有のスペクトル特性に関する研究

歌声特有のスペクトル形状については，Sundberg が singers' formant（歌唱ホルマント）と呼ばれる 3 kHz 付近の顕著なスペクトルピークの存在を明らかにしている [78]．そこでは，singers' formant が男性のオペラ歌手やコンサート歌手の歌声に一貫して観測され，オーケストラ演奏における歌声の聞こえやすさに影響を与えてることが報告されている [77, 82]．また，中山らは，邦楽歌唱を対象とした音響分析の結果，3 ～ 4 kHz 付近において話声に無い顕著なスペクトルピークの存在を明らかにし，邦楽歌唱にも singers' formant が存在することを示している [15, 59]．

更に Sundberg は，singers' formant の生成機構を明らかにするべく，声門直上部の声道形状を円筒型のチューブで模擬したモデルを作成し，これを用いた音声合成を行っている [81]．この結果，singers' formant の生成は，広い下咽頭と比較的狭い喉頭管で形成される声道形状に起因していることを明らかにしている．また，今川らによって構築された声帯と声帯直上部の振る舞いを模擬した物理モデルや [2]，竹本らによる 3 次元 MRI 画像を基に構築した声道下部構造のモデルを用いた実験 [13] によって，Sundberg の知見を支持する結果が示されている．

また，歌声特有の基本周波数変化に連動し，スペクトル形状が変動する事も明らかになってきている．中でも，音声振幅全体がヴィブラートに連動して変調し，それによってホルマントが振幅変調することも多くの研究者によって報告されている [38, 39, 42, 65, 67]．また，Svec は，様々な歌唱技法において頻繁に使われる声区変化に着目し，急峻な基本周波数遷移に伴って，スペクトルの低域成分が強調され，高域成分が弱まることでスペクトルの傾きが急峻になることを示している [84]．

更に辻らは，話声を歌声らしく知覚される音声に変換させる為には，singers' formant と，ヴィブラートに同期したホルマントの振幅変調の成分が重要であることを示している [14]．

これらの取り組みによって、歌声特有のスペクトル特性の存在、更にはその生成機構が明らかになってきている。そしてこれらの特性が、多くの歌声において観測されることも明らかになってきている。しかし、各特徴を自由に操作可能な歌声合成法が提案されていない為、各特徴が歌声知覚に与える影響は定量的には調査されていない状況である。

### 1.2.3 歌声合成に関する研究

話声を対象とした音声合成は、これまでに様々な研究が行われて来ている。特に近年では、情報科学の枠組みを越え心理学や生理学といった多角的立場からの知見を考慮することによって、高品質な音声合成システムの実用化 [44] や、感情や態度といった非言語情報を考慮した多様な発話スタイルを表現可能な音声合成 [70, 41] の開発が進められている。一方、歌声合成の研究は、話声に比べると非常に少ない。その中で、実用化に向けた取り組みと、基礎的な取り組みそれぞれの立場で研究が行われている。

実用化に向けた取り組みとしては、波形接続方式音声合成 [9] や、HMM 等の確率モデルに基づいた音声合成方式 [12] を用いた歌声合成システムの研究がある。

波形接続方式とは、予め録音した多量の音声データから、必要となる音素を選択・結合させることで音声を合成する手法であり、現在市販されている歌声合成ソフトウェアの殆んどは、この方式を採用している。例えば、YAMAHA による VOCALOID [22] は、歌声のパートの旋律と歌詞を入力することで、高品質で自然性の高い歌声が合成できるシステムである。また、NTT アドバンステクノロジーのワンダーホルン [23] は、同様の入力で、こぶしやヴィブラート等の歌唱表現も制御された歌声の合成を可能にしている。

また、酒向らは、HMM 音声合成システムを基にした歌声合成法を提案している [10, 11]。これは、歌声データベースから得られる歌詞、音高、音長といった情報を学習することで歌声モデルを構築し、目的の楽譜情報に基づいた歌声モデルを選択することで、多様な歌声を合成可能とするものである。

これらは、任意のメロディ、歌詞による歌声合成を可能にしている。特に波形接続方式によるものは、録音された歌声データから歌声を作成する為、より肉声

に近い歌声を実現できる。しかし、どちらのシステムにおいても、合成される歌声の音韻性や声質は、予め録音された歌声データベースに大きく依存してしまう。その為、多様な歌声を合成する為には、多量の歌声データベースが必要となってくる。そして何より、歌声特有の音響的特徴を自由に操作することが困難であり、音響的特徴と歌声知覚の関係を定量的に評価するような歌声知覚・生成機構解明に向けた基礎研究に応用することが困難である。

一方、基礎的な取り組みとしては、近年になって音声生成過程モデルに基づいたボコーダ型分析・合成手法による研究が行われてきている。

河原らは、高品質な音声分析合成システム STRAIGHT[45, 46] を用いたスキヤットの生成を試みている [7]。ここでは、基本周波数、音源特性、スペクトル包絡の制御方法に関する予備的検討を行っている。また、ヴィブラート制御に重点を置いたアカペラ歌唱の聴覚モーフィングといった新しい取り組みによって、高品質な歌声合成を試みている [47]。

Minematsu らも STRAIGHT を用いることで、長唄の分析・合成に関する一連の研究を行っている [52, 53]。ここでは、長唄特有の音響的特徴に着目した基本周波数とパワーの制御を行うことで、高品質な長唄合成に成功している。

森らは、ARX モデル[61] を用いたバリトン歌唱の分析・合成を試みている。ここでは、音高上昇・下降時のオーバーシュートの特性の違いを考慮した基本周波数制御、及び歌声合成を試みている [54, 55]。

Takara らは、ケプストラム分析・合成法を用いて、古典琉球語謡曲「王府おもしろ」の合成に取り組んでいる [85]。ここでは、王府おもしろ特有の基本周波数特性に着目した合成法が提案されている。

これらボコーダ型の分析・合成手法に基づいた研究は、基本周波数やスペクトルに含まれる歌声特有の音響的特徴の分析・制御を検討する取り組みでもあり、歌声知覚・生成機構解明に向けた基礎的研究とも言える。そして、そのほとんどの研究では、主に歌声特有の基本周波数制御を試みており、歌声合成における基本周波数制御の重要性を示している。しかし、前述した歌声特有の音響的特徴、とりわけ F0 動的変動成分と歌声知覚の関係を明らかにする試みが不足している為、これらの研究は、ある特定の歌唱法や、歌唱法特有の音響的特徴に限定した取り組みとなっている。

## 1.3 本論文の目的

1.1 節で述べたように、話声を対象とした音声合成の研究が、心理学や生理学といった多角的立場から得られた知見を検証する為の最も効果的な手法であり、更には、高品質な音声合成システムを構築する過程で、音声知覚・生成についての有益な知見が得られることから、話声の知覚・生成機構の解明には必要不可欠なものになりつつある。これは、話声の合成だけでなく歌声合成に関しても同様であろう。しかし、前節で見てきたように、現在は歌声合成システムの構築はもとより、歌声を合成する上で必要な音響的特徴すら明らかになっていない状況である。なぜなら、これまでの歌声研究が、歌声を歌唱法や歌唱技術によって規定される多様な存在と捉えることで、特定の歌声に含まれる音響的特徴や、その特徴と歌声知覚の関係を検討することに特化していたからである。この現状を打破する為には、歌声を普遍的な1つの存在と捉えることで、歌声が歌声として知覚されるために必要な音響的特徴を明らかにし、それらを制御可能な歌声合成システムを構築する必要がある。そして、この取り組み中でこれまでの歌声研究で得られてきた知見を生かしていくことで、はじめて高品質な歌声合成システムが構築され、歌声知覚・生成機構の解明に繋がる取り組みが可能になると考えられる。

以上から、本論文では、歌声知覚・生成機構の解明に向けた基礎的取り組みとして、歌声に共通して含まれる音響的特徴に着目した歌声合成システムの構築を試みることで、歌声を知覚する上で重要な役割を担っている音響的特徴を明らかにする。

歌声知覚に必要な音響的特徴が明らかとなり、歌声合成システムが構築されれば、話声を対象とした音声合成と同様の枠組みで研究を遂行できる。つまりは、これまでに得られてきている「歌を歌う」ことに関する様々な知見を歌声合成に適用することで、歌声合成の研究の発展はもちろん、歌声知覚・生成機構の解明に繋がる取り組みが可能となる。歌声合成の発展は、計算機音楽の分野に新たな音楽創造の可能性を与えることができる。そして、歌声知覚・生成機構の解明は、理想の歌唱法や音色によって歌う為の優れた歌唱技術の習得の一助となる。

研究背景でみたように、歌声特有の音響的特徴は、基本周波数やスペクトル形状といった様々な音響パラメータに存在する。双方とも、歌声知覚・生成において

重要な意味を持っていることは間違いない．そこで，本論文では，高品質な歌声合成システムの構築する過程において，以下の3つの課題を遂行することで，歌声特有の音響的特徴と歌声知覚の関係を調査する．

1. 歌声の基本周波数に含まれる歌声知覚に必要な音響的特徴は何か？
2. 歌声のスペクトルに含まれる歌声知覚に必要な音響的特徴は何か？
3. 歌声知覚において重要な役割を担っている音響的特徴は何か？

はじめに，様々な歌声の基本周波数に共通して含まれる  $F_0$  動的変動成分の制御方法を検討することで，歌声の基本周波数制御モデルを構築する．次に，モデルを用いた歌声合成を行う事で，1番目の課題について検討する．次に，歌声特有のスペクトル特性を制御する方法を構築し，2番目の課題について検討する．そして最後に，基本周波数制御，スペクトル制御を組み合わせることで，歌詞の朗読音声から歌声を合成するシステムを構築し，3番目の課題について検討する．

## 1.4 本論文の構成

本論文の構成を以下に示す．

第1章では，本論文が対象としている研究分野の背景と問題点を示し，本論文の位置付と目的を示す．

第2章では，本論文の研究方針を示す．ここでは，本論文において使用する分析合成系 STRAIGHT と，歌声データについて概説する．そして，本論文で構築を試みる歌声合成システムの構成を示し，それを構築する過程において如何に歌声知覚に必要な音響的特徴を明らかにしていくかについて述べる．

第3章では，歌声の基本周波数制御モデルを構築する．はじめに，様々な歌声に共通して観測される  $F_0$  動的変動成分を抽出する．そして，それらの成分を制御可能な歌声の基本周波数制御モデルを構築する．最後に，モデルにおける制御パラメータ値の抽出を行い，モデルの有効性を検証する．

第4章では， $F_0$  動的変動成分と歌声知覚の関係について詳細に調査する．はじめに，前章で提案した基本周波数制御モデルを用いた歌声合成法を提案する．次に，個々の  $F_0$  動的変動成分を原音声から除去した場合，及び矩形的なメロディ変

化に各成分を付与した場合の歌声知覚の反応をそれぞれ調査する．次に，歌唱技量の違いと各成分の特性の違いについて検討する．次に，個々の F0 動的変動成分の特性と歌声知覚の関係を調査する．次に，ヴィブラート特性の時間変動と歌声知覚の関係を調査する．最後に，音高遷移方向の違いによるオーバーシュート特性の差異が歌声知覚に与える影響を調査する．

第 5 章では，歌声特有のスペクトル特性の制御方法を構築し，それら特性と歌声知覚の関係について調査する．最初に，singers' formant の制御方法を提案し，singers' formant と歌声知覚の関係を調査する．次に，基本周波数の周期的な変動成分であるヴィブラートに伴った音声の振幅変調，及びホルマントの振幅変調の制御方法を提案し，それらの特性が歌声知覚に与える影響を明らかにする．

第 6 章では，第 2 章で提案した歌声合成システムを実装として，歌詞の朗読音声から歌声に変換する歌声合成法を提案する．ここでは，前章までで構築した基本周波数制御モデルとスペクトル制御法に加え，朗読音声の音韻長を歌声のものに変換する音韻長制御法を提案する．そして，それら制御法を STRAIGHT に組み込んだ朗読音声から歌声を変換・合成可能なシステムを実装し，その合成音の評価を行うことで，個々の音響的特徴量の歌声知覚における重要性について検討する．

第 7 章において，本論文で得られた結果を要約し，今後の展望を述べる．

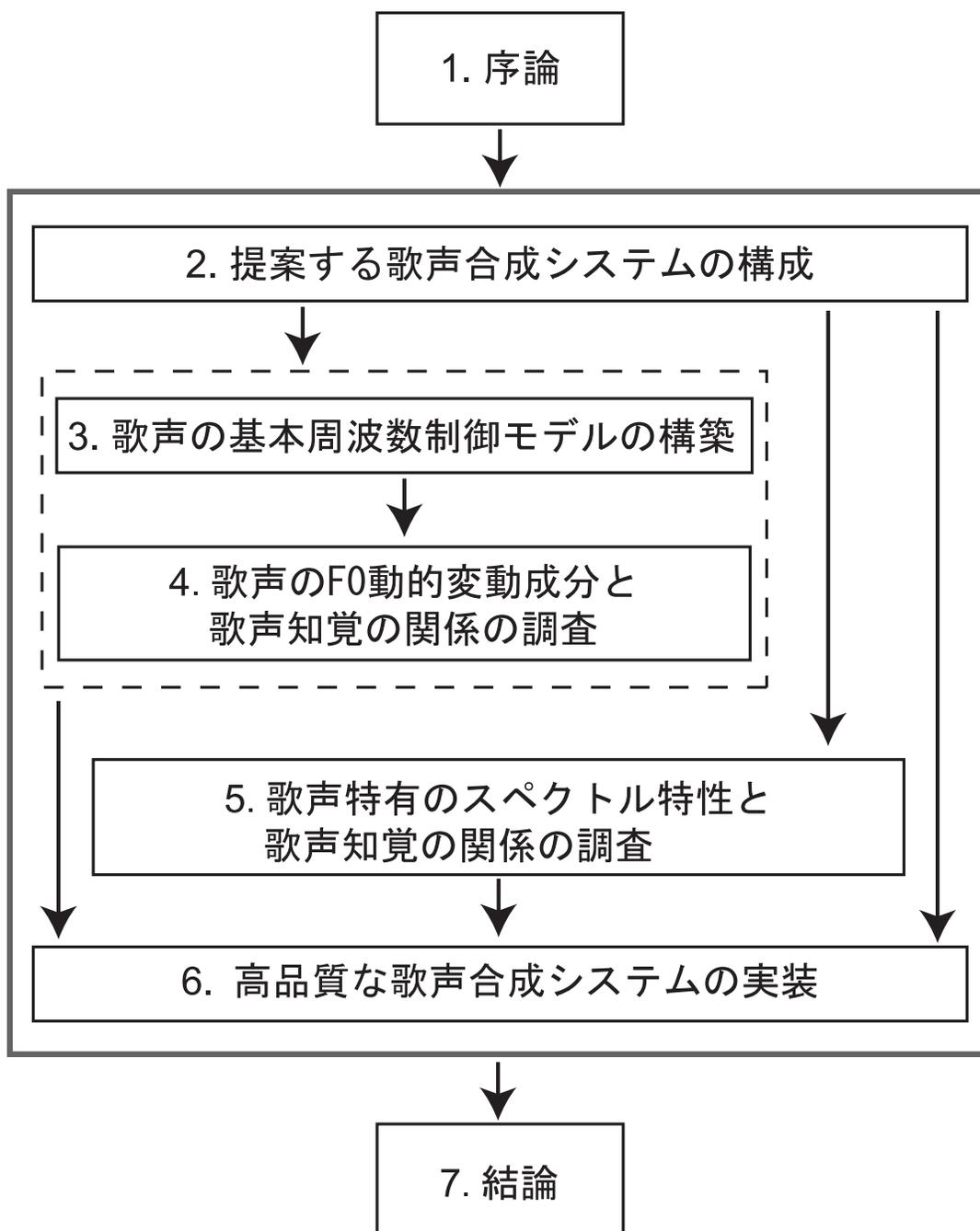


図 1.1: 各章の相互関係

## 第 2 章

### 歌声合成システムの構成

## 2.1 はじめに

本章では，本論文で構築を試みる歌声合成システムの構成を示し，システムを構築する過程において歌声知覚に必要な音響的特徴を解明していく為の方針を示す．

はじめに，本論文で構築を試みる STRAIGHT[45, 46] を用いた音声合成システムの構成を示す．ここでは，歌声合成システムを構築する為に必要な基本周波数とスペクトルの制御方法を検討する過程において，如何にして各音響パラメータに含まれる歌声特有の音響的特徴の歌声知覚における重要性を検証するかについて述べる．

次に，本論文で使用する音声分析合成系 STRAIGHT について概説する．そして，本論文で分析・合成対象とする歌声データについて説明する．

## 2.2 歌声合成システム構築の方略

前章の研究背景でみたように，近年になってボコーダを用いた歌声合成の研究が行われてきている．ボコーダとは，音声生成過程モデルに基づき，音声から声帯振動情報の基本周波数と，声道形状情報のスペクトルの音響パラメータを抽出し，これらを用いて音声を合成（再生）するシステムであり，音声分析合成系とも言われる [19]．中でも，近年河原によって提案された STRAIGHT[45, 46] は，高精度な音声分析，自由な音響パラメータ操作，そして高品質な音声合成が可能なボコーダであり，これを用いた歌声合成の研究では，高品質なオペラ歌唱 [7] や長唄の合成 [53] に成功している．また，これらの研究は，対象とする音響的特徴や歌唱法は限定した取り組みであるものの，STRAIGHT を用いた歌声分析・合成の取り組みが高品質な歌声合成はもとより，歌声知覚・生成機構を解明する為の有効な手法であることを示している．

以上から，本論文においても STRAIGHT を用いた歌声分析・合成の枠組みで歌声合成システムの構築を試みる事で，音声生成機構に対応した音響パラメータに含まれる歌声知覚に重要な音響的特徴について検討する．

### 2.2.1 本論文で提案する歌声合成システムの構成

歌声知覚において重要な役割を担っている音響的特徴を明らかにするためには、以下の3つを検討する必要がある。

1. 歌声の基本周波数に含まれる歌声知覚に必要な音響的特徴は何か？
2. 歌声のスペクトルに含まれる歌声知覚に必要な音響的特徴は何か？
3. 歌声知覚において重要な役割を担っている音響的特徴は何か？

そこで本論文では、歌詞の朗読音声歌声に変換する歌声合成システムを提案する。そして、このシステムを構築する過程において、上記3つの課題を検討していく。

図 2.1 に本論文で提案する歌声合成システムの構成を示す。このシステムは、STRAIGHT による歌詞の朗読音声分析・合成過程において、基本周波数制御 (F0 control)、スペクトル制御 (Spectrum control)、音韻長制御 (Duration control) の3つの制御を組み込むことで構成される。そして、各制御法を構築する上で、上記3つの課題を検証する。その方略を次項から説明する。

### 2.2.2 基本周波数制御法構築の方略

歌声は話声に比べ、より複雑で動的な特性を持っている。とりわけ、基本周波数の変化においては、その特性が顕著に現れていることが知られている [15]。実際に、先行研究においては、歌声特有の基本周波数特性の存在が数多く報告され、それら成分が歌声合成を行う上で重要であることも指摘されてきている。更には、話声と歌声の違いを知覚する上で、歌声特有の音響的特徴、なかでも基本周波数における特徴量が重要な役割を果たしていることが報告されている。その為、基本周波数における歌声特有の音響的特徴と歌声知覚の関係を明らかにすることが、高品質な歌声合成の実現はもちろん、歌声知覚・生成機構に向けて最も重要な課題であると考えられる。

図 2.2 に、日本童謡「七つの子」の歌い出し2小節“からすなぜなくの”を朗読・歌唱した際の基本周波数変化を示す。ここからも分かるように、発話内容が同じであっても、基本周波数変化の振る舞いが、話声と歌声で大きく異なることが分

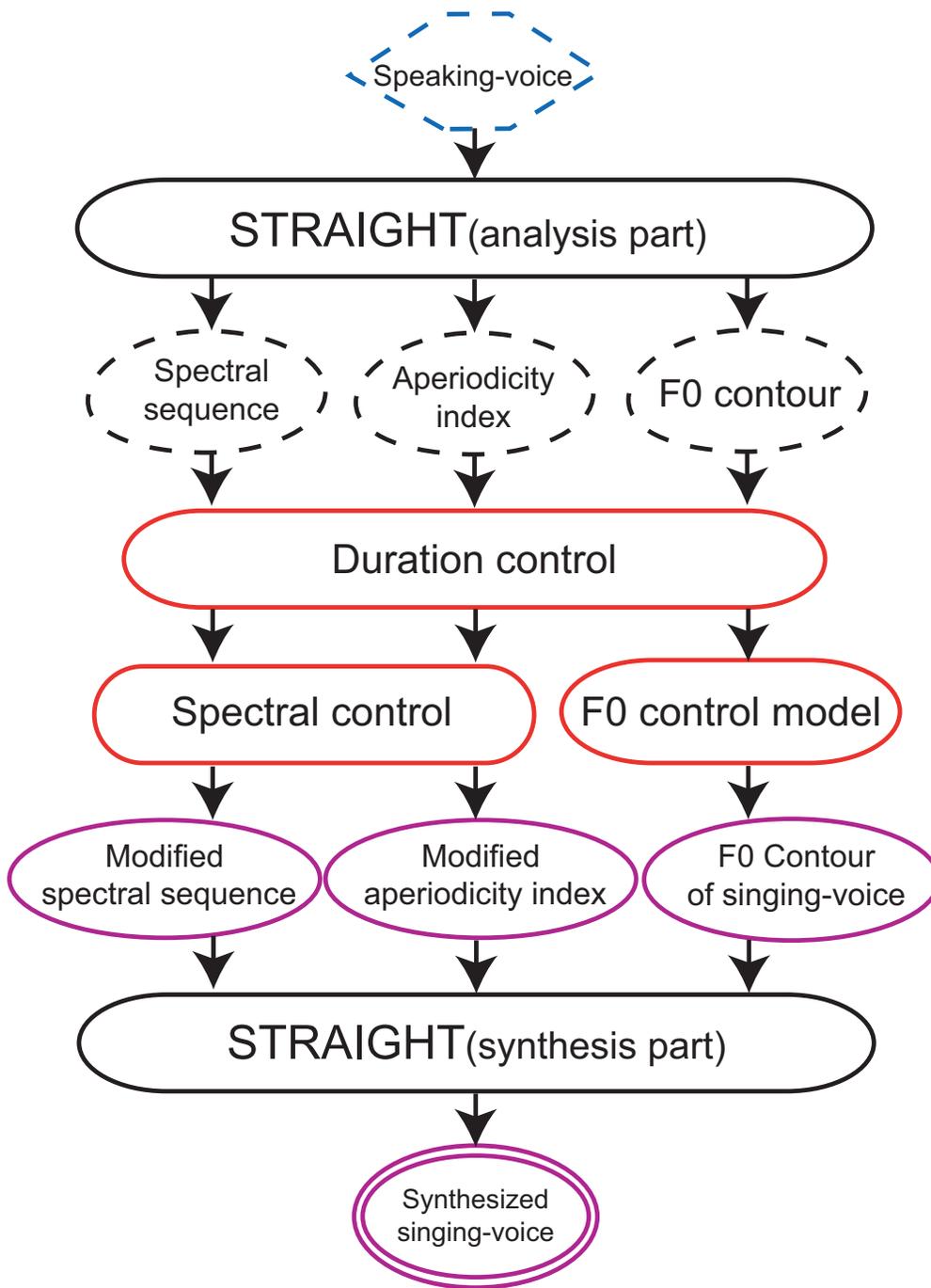


図 2.1: 歌声合成システムの概要.

かる．そして，歌声特有の基本周波数特性としては，以下の3つが挙げられる．

1. 基本周波数は音階に，その変化の概形は楽曲のメロディに対応する

12平均律において440 Hzの基本周波数を持つ音は，A4（八長調のラ）の音を表す．よって基本周波数は，楽譜に記されるメロディ変化を直接表す事になる．このような特性は，メロディを持たない話声の基本周波数変化には当てはまらない．

2. 同一の音高区間の基本周波数変化は，音高値からの大きな逸脱がなく安定した特性を示す

話声の場合，1つのフレーズにおける基本周波数は，時間と共に減少するのが通常である．これは基本周波数変化が肺からの空気圧によって変化することに起因している．歌声の場合は，基本周波数が一定に保たれることで音高が生まれ，その連続がメロディを構成する．その為，同一音高区間においては，基本周波数が大きく下降するような変化は示さない．

3. メロディ変化中にF0動的変動成分が存在する

これは，メロディに従った安定した基本周波数変化において，ほぼ規則的なタイミングでメロディ変化から逸脱する変化を示す成分である．その変化は，話声の基本周波数変化においては観測されないような短時間の間に起る急峻なものである．

1，2番目の特性は，共にメロディ変化を構成する為に必然的に生成されるものであり，話声には無く歌声特有の特性で，且つ歌声知覚において重要であることは当然考えられる．一方で，3番目の特性は，メロディ変化から逸脱する成分であり，一見して歌声知覚において不必要な特性として捉えることもできる．しかし，先行研究でみたように，主にオペラ歌唱を対象とした音響分析の結果，複数のF0動的変動成分が存在することが確認され，その重要性が指摘されている．また，我々が馴染みの無いメロディの歌声や単独発話母音の歌声を聴いた場合でも，それらを歌声と知覚できることを考えると，メロディ変化以外の基本周波数特性であるF0動的変動成分が，歌声知覚において重要な役割を担っていることが十分考えられる．しかし，あらゆる歌声に共通してF0動的変動成分は存在するのか，またそれらは歌声知覚において重要であるのかという問題に関しては，未だ

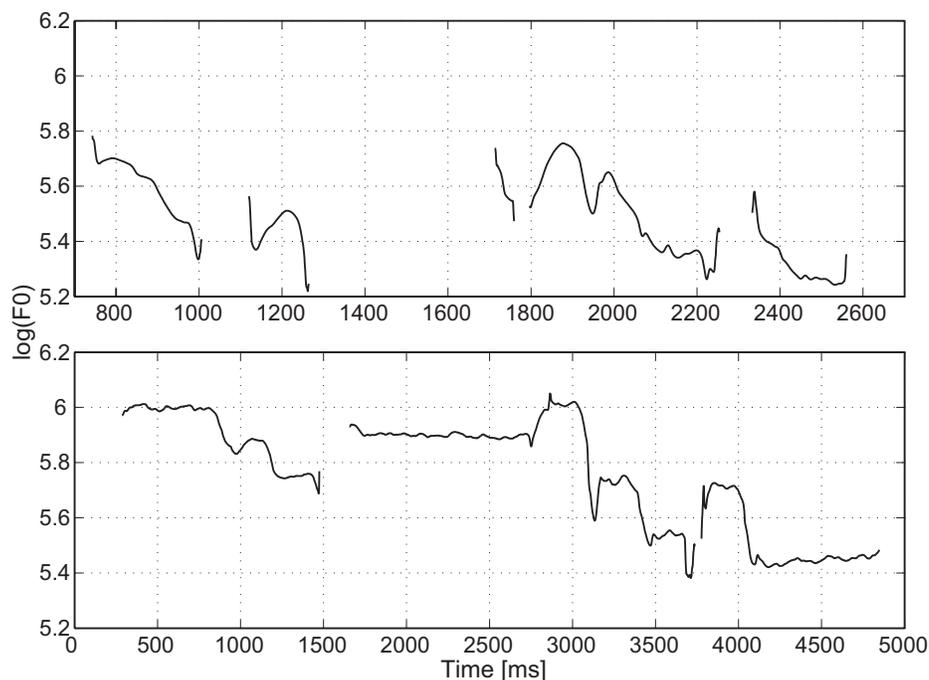


図 2.2: STRAIGHT-TEMPO で抽出した“からすなぜなくの”発話・歌唱時の基本周波数変化（上:朗読音声，下：歌唱音声）。

明らかとなっていない。

そこで，基本周波数制御方法を構築する為の方略として，歌声特有の基本周波数特性である  $F_0$  動的変動成分に着目し，その成分を制御可能な歌声の基本周波数制御モデルを構築する．そしてこのモデルを用いることでスペクトル制御を考慮しない歌声合成システムを構築する．このシステムによって， $F_0$  動的変動成分を操作した歌声合成音を作成し，心理物理実験で評価することで，各成分と歌声知覚の関係の定量的な調査を行う．以上の取り組みにより，最初に挙げた 1 番目の課題「歌声の基本周波数に含まれる歌声知覚に必要な音響的特徴は何か？」に対する明確な解答を得ることを試みる．

歌声の基本周波数制御方法の構築に関しては第 3 章で， $F_0$  動的変動成分と歌声知覚の関係に関しては第 4 章においてそれぞれ詳細に述べる．

### 2.2.3 スペクトル制御法構築の方略

前章の先行研究でみたように、歌声特有のスペクトル特性に関する研究は、基本周波数特性に比べ多くのことが明らかとなってきた。中でも、singers' formant や、ヴィブラートに連動したホルマントの振幅変調成分は、様々な歌唱において共通して存在することが明らかになってきている。また、それら特性の生成機構解明に向けた生理学的側面からの取り組みや、歌声知覚における役割を明らかにする試みも行われている。しかし、歌声合成の取り組みを通じてそれら特性を自由に操作した合成音を作成し、心理物理実験によって評価することで、歌声知覚との関係を定量的に調査する試みはほとんど行われていない現状にある。

そこで、スペクトル制御法を構築する為の方略として、先行研究で報告されている singers' formant と、基本周波数の周期変動成分であるヴィブラートに連動したホルマントの振幅変調成分に着目する。はじめに、各成分を制御可能な方法を提案する。そしてこの方法を先に構築した基本周波数制御のみを考慮した歌声合成システムに組み込むことで、各成分と歌声知覚の関係について検討する。以上の取り組みより、2番目の課題「歌声のスペクトルに含まれる歌声知覚に必要な音響的特徴は何か？」に対する明確な解答を得ることを試みる。

これら歌声のスペクトルに関する取り組みは、第5章において詳細に述べる。

### 2.2.4 音韻長制御法構築の方略

音韻長の制御は、先に述べた基本周波数とスペクトルの制御方法が確立された後、最終的に話声から歌声を変換・合成するシステムを実装する上で必要となる。歌声の音韻長は、音韻に割り当てられた音符の種類と曲のテンポによって一意に決定される。その為、音韻長制御法構築の方略として、テンポ情報に基づいた音韻長の時間伸張処理方法の構築を試みる。その後、基本周波数、スペクトル、音韻長の制御方法を統合することで図 2.1 のシステムの実装を行う。そして、基本周波数とスペクトルにおける歌声特有の音響的特徴を操作した歌声合成を行い、心理物理実験によって合成音の評価を行うことで、3番目の課題「歌声知覚において最も重要な音響的特徴は何であるか？」に対する検討を行う。

音韻長制御の構築、及び歌声知覚と複数の音響的特徴との関係の調査は第6章

で行う．

## 2.3 STRAIGHT

本論文で使用する高品質なボコーダ STRAIGHT について概説する．

STRAIGHT は，図 2.3 に示すように，STRAIGHT-core，SPIKES，TEMPO の 3 つの主要な部分から構成される．

STRAIGHT-core は，音声の励振の周期性による干渉の影響のない時間周波数表現，つまりはスペクトル情報を抽出する方法である．その中心的なアイデアは基本周期，基本周波数を節点とする区分的線形関数による補間と等価な時間周波数領域の平滑化を行なうことにある．

TEMPO は，2 つのフィルタ出力の微分の特性を基に，音声の基本周波数を推定する方法である．特別なフィルタ設計と搬送対雑音比（C/N 比）の組み合わせにより，正確な基本周波数の推定が可能になっている．

SPIKES は，合成に用いる駆動音源の位相特性を操作することにより，ボコーダ特有の buzzy な音色を軽減する方法である．ここでは，同一のパワースペクトルであっても群遅延を操作して時間的な微細構造を変えることで音色が変化することを利用している．

STRAIGHT で抽出される主な音響パラメータは，TEMPO で抽出される基本周波数と，STRAIGHT-core で抽出されるスペクトル包絡の時間系列，及び非周期成分割合の時間系列である．ここで，非周期成分の割合とは，音声の非周期成分の比率を表すパラメータであり，非周期成分に対応するスペクトルの下側包絡と周期成分に対応する上側包絡の差分によって記述される．これら音響パラメータは高精度に且つ自由に制御可能な状態で抽出される．そして SPIKES における音源特性の操作によって，合成音特有のバズ音が無い合成を可能にしている．よって，STRAIGHT を用いた歌声の分析，加工，そして再合成の処理体系に基づいた取り組みにより，高精度に抽出された音響パラメータを対象とした歌声特有の音響的特徴の分析ができ，更にはそれら特徴を操作した合成音を作成することによって，各特徴と歌声知覚の関係を定量的に調べることができる．

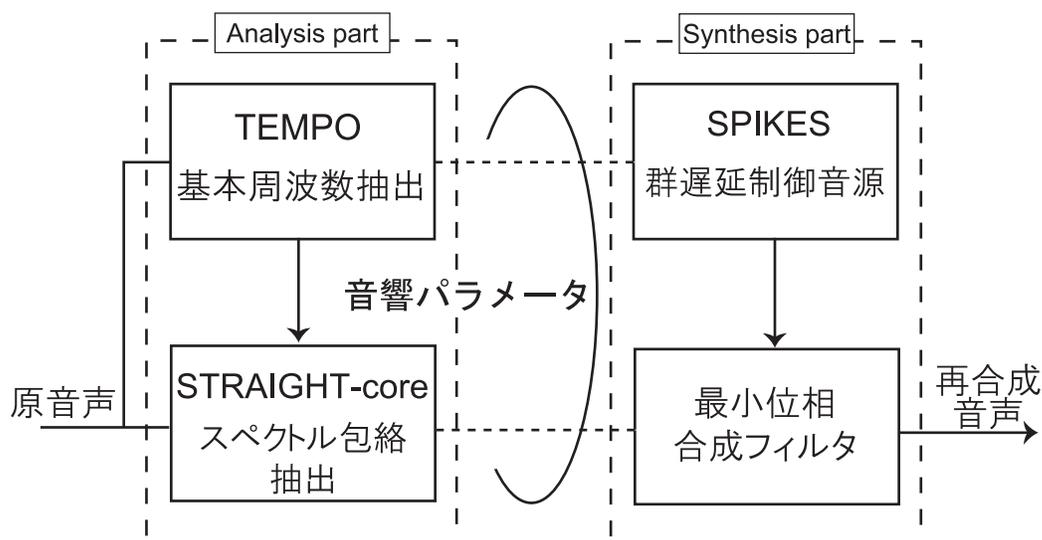


図 2.3: STRAIGHT の構成 [6]

## 2.4 本研究で扱う歌声データ

本論文で明らかにしたい歌声知覚に重要な音響的特徴は、ある特定の歌唱法や歌唱者に特化したものや、卓越した歌唱技術によって発声された歌声のみに観測されるものではなく、あらゆる歌声に共通して含まれるものである。その為、分析・合成対象にする歌声データとして、アマチュアとプロ両方の歌唱者、更には様々な歌唱法による歌声データを用いる。歌声データの詳細を以下に記す。

### 歌声データセット A

このデータセットは、主としてアマチュア歌唱者による歌声を収録したものである [8]。その詳細は以下の通りである。

歌唱者 : ボーカールスクールに通う女性 3 名, オペラ歌唱歴 18 年の男性 1 名

歌唱法 : 女性 3 名は特に無し。男性はオペラ歌唱 (テノール)

発声内容 : 日本童謡「七つの子」日本語母音 /a/ のみによる歌唱と、歌詞を付けた歌唱それぞれを収録 (但し男性は母音歌唱のみ)。

尚，各歌唱者に対しては，歌唱法に関する特別な指示は与えず，譜面に記されている音階で歌唱するという制限も与えていない．その理由として，歌唱者の発声可能な音域を考慮に入れたためである．また，録音は防音室内で行われ，歌声は，サンプリング周波数 48 kHz，量子化ビット数 16 bit で DAT にデジタル録音されたものである．また，録音された 16 小節からなる歌を 4 小節ごとに分割した計 12 個の歌声データとなっている．尚，録音に使用した機器を以下に列挙する．

単一指向性コンデンサマイクロホン : SONY C-536P

マイクロホン用パワーサプライ : SONY AC-148F

ポータブル DAT 録音機 : SONY TCD-D10 DAT

## 歌声データセット B

これは，大規模な歌声データベース「日本語を歌・唄・謡う」[16]の中から選定した洋・邦楽様々な歌唱法による歌声を収録するデータセットである．詳細を以下に記す．

歌唱者 : 洋・邦楽のプロ歌唱者 24 名

歌唱法 : 洋楽: ソプラノ (5 名), メゾ・ソプラノ (2 名), アルト (1 名), テノール (3 名), バス・バリトン (3 名)

邦楽: 長唄 (1 名), 民謡 (6 名), わらべうた (1 名), 演歌 (1 名), ポピュラー (1 名)

発声内容 : 共通歌詞 “かえでいろづくやまのあさは” (メロディは歌唱者の自由)  
日本語 5 母音の歌声と話声

## 2.5 まとめ

本章では，歌声合成システムを構築する過程において，歌声知覚・生成機構において重要な音響的特徴を明らかにする為の方略を示した．

はじめに，本論文で提案する歌声合成システムとして，話声から歌声を変換合成するシステムの構成を示した．ここでは，歌声合成システムを実装する為に必要な基本周波数とスペクトルの制御法を構築する過程において，各音響パラメータにおいて着目する音響的特徴を示し，その特徴と歌声知覚の関係を調査するための研究方針を提案した．

次に，本論文で利用する音声分析合成系のシステム，及び歌声データセットについて概説した．音声分析合成系には STRAIGHT を採用することで，高品質な歌声の音響分析，及び合成が可能であることを示した．また，歌声データセットに関しては，歌唱技量や歌唱法の違いに依存しない歌声特有の音響的特徴を明らかにする為の 2 種類の歌声データセットについて概説した．

## 第 3 章

# 歌声の基本周波数制御モデルの構築

## 3.1 はじめに

本章では、歌声を知覚する上で重要な役割を担っている F0 動的変動成分を明らかにするための最初の手続きとして、歌声の基本周波数制御モデル（以後 F0 制御モデルと呼ぶ）を構築する。

最初に、アマチュア歌手と、プロの洋・邦楽歌手の歌声から抽出された基本周波数軌跡を対象に、先行研究で報告された F0 動的変動成分や、それ以外の特徴的な成分が存在するか否かを調査し、本研究で対象とする F0 動的変動成分を定義する。

次に、歌声の F0 制御モデルを構築する。ここでは、F0 動的変動成分を自由に操作するための方法を検討することで、歌声の基本周波数変化を制御・記述可能なモデルの構築を行う。

最後に、歌声 F0 制御モデルの有効性を検証する。ここでは、モデルで記述される基本周波数変化と原音声の基本周波数変化のフィッティング誤差を求めることで、各 F0 動的変動成分が正確に制御可能か否かについて検証する。

## 3.2 歌声の F0 動的変動成分の抽出

歌声特有の F0 動的変動成分の存在については、これまでに多くの研究で報告されている。しかし、そのほとんどがオペラ歌唱のみを対象とした取り組みであった。そこで、本節では、以下の2点について検証することで、本論文で対象とする F0 動的変動成分を定義する。

- 過去に報告された F0 動的変動成分が、オペラ歌唱も含め、それ以外の歌声においても観測されるのか？
- 過去に報告された以外には F0 動的変動成分は存在しないのか？

### 3.2.1 基本周波数推定

前章に記したように、本論文では、歌声の基本周波数推定法として STRAIGHT の TEMPO[45] を用いる。基本周波数の抽出には、音声の周期性の特徴を利用した自己相関処理法 [86] や、調波性の特徴を利用したケプストラム法 [26] といった

種々の方法が提案されているが，その中で TEMPO は，音声の瞬時周波数を利用した手法の一つである．TEMPO の基本的なアイデアは，フィルタバンク出力の瞬時周波数とフィルタの中心周波数が一致する周波数が調波の存在する周波数として考えられることを利用することであり，高精度な基本周波数推定を実現している．実際，様々な手法の中で，TEMPO による基本周波数推定精度が雑音を含まない音声に対して非常に高い事が示されている [30] ．

### 3.2.2 歌声データ

本節で対象とした歌声データは，歌声データセット A に収録されたすべての歌声と，歌声データセット B に収録された共通歌詞を歌唱したすべての歌声である．

### 3.2.3 F0 動的変動成分

図 3.1 , 3.2 に，TEMPO によって抽出した歌声データセット A に収録されたアマチュア歌手の歌声と，歌声データセット B に収録されたテノールと民謡歌唱の基本周波数変化を示す．すべての歌声データの基本周波数変化を観測した結果，以下に示す 4 つの成分の存在が確認された．

- オーバーシュート (Overshoot)

音高変化時から変化直後にかけて観測され，以下の 2 つの成分から構成される．

ポルタメント (Portamento) : 音高変化時の傾斜を持った滑らかな基本周波数変化

オーバーシュートエクステンツ (Overshoot Extent) : 音高変化直後に目的音の音高値を越えて振れる瞬時的な振動成分

- ヴィブラート (Vibrato)

同一音高区間での 4 ~ 7 Hz の周期的な振動成分

- 微細変動成分 (Fine-fluctuation)

上記の変動成分を除去した後に残る不規則で細かい振動成分

- プレパレーション (Preparation)

音高が変化する直前に、変化とは逆の方向に振れる瞬時的な振動成分

ポルタメントに関しては、Sundberg によって行われたオペラ歌唱を対象とした音響分析結果、その存在が示されている [80]。そこでは、ポルタメントの傾き（音高変化の速さ）に着目し、音高下降時の方が上昇に比べ、傾きが急峻（変化が速い）なことを示している。また、Myers や Ohala によっても同様の取り組みが行われている [58, 60]。Mori らは、バリトン歌唱を対象に分析を行うことで、ポルタメントはもちろん、オーバーシュートエクステント成分の存在を示している [54, 55]。de Krom らも、オペラ歌唱を対象にオーバーシュートエクステントの大きさを測定することで、音高変化直後に必然的に存在する成分であることを示している [31]。しかし、今回様々な歌声を分析した結果、洋楽歌唱だけでなく、邦楽歌唱やアマチュア歌唱においても共通して観測される成分であることが確認された。

ヴィブラートに関しては、Seashore らの研究 [71] を皮切りに、様々な取り組みが行われてきている。その結果、ヴィブラートが歌声だけでなく楽器音においても観測されることが示されている [91]。また、その存在は、オペラ歌唱やコンサート歌唱 [28] といった洋楽歌唱だけでなく、様々な邦楽歌唱においても確認されている [49]。更にこの成分が歌声中に存在するか否かが、歌唱技術を評価する上で一つの指標になるとまで報告されている [51]。その中で、今回の歌声分析では、ヴィブラートの存在がアマチュア歌唱においても確認され、歌声に共通して含まれる F0 動的変動成分であることが明らかとなった。

微細変動は、Akagi らのアマチュア歌唱を対象とした分析の結果、その存在が報告されている [24]。更にこの研究では、微細変動と歌声知覚の関係も定量的に調査されており、歌声知覚における重要性を示している。また、今回の歌声分析では、プロ歌唱においてもその存在が確認された。

以上から、最初に挙げた 3 つの成分は、先行研究において存在が報告されていると同時に、今回の分析においてもほとんどの歌声において観測されることが分かった。これにより、3 つの成分が、歌声共通の基本周波数特性である可能性が強いことが確認された。しかし、これらの成分が歌声を知覚する上で必要な成分なのか、また、仮に必要であるのならどの成分が最も重要な役割を担っているのかという問題に対しては明確な解答が得られていない状況にある。その為、歌

声合成を通じて、個々の成分と歌声知覚の関係を定量的に調査し、歌声知覚における各成分が担う役割について調査する必要がある。

一方で、4番目に挙げた成分のプレパレーションの存在は、先行研究では報告されて来ていない。しかし今回の歌声分析では、他の成分と同様に多くの歌声データにおいて観測された。よって、歌声を知覚する上で重要な役割を担っている可能性が高く、歌声知覚との関係を調査する必要がある。

以上から、本論文では、上記4つの成分を歌声特有のF0動的変動成分として定義し、その歌声知覚における重要性について検証する。次節では、その為の最初の手続きとして、各成分の制御方法を検討することで、歌声のF0制御モデルを構築する。

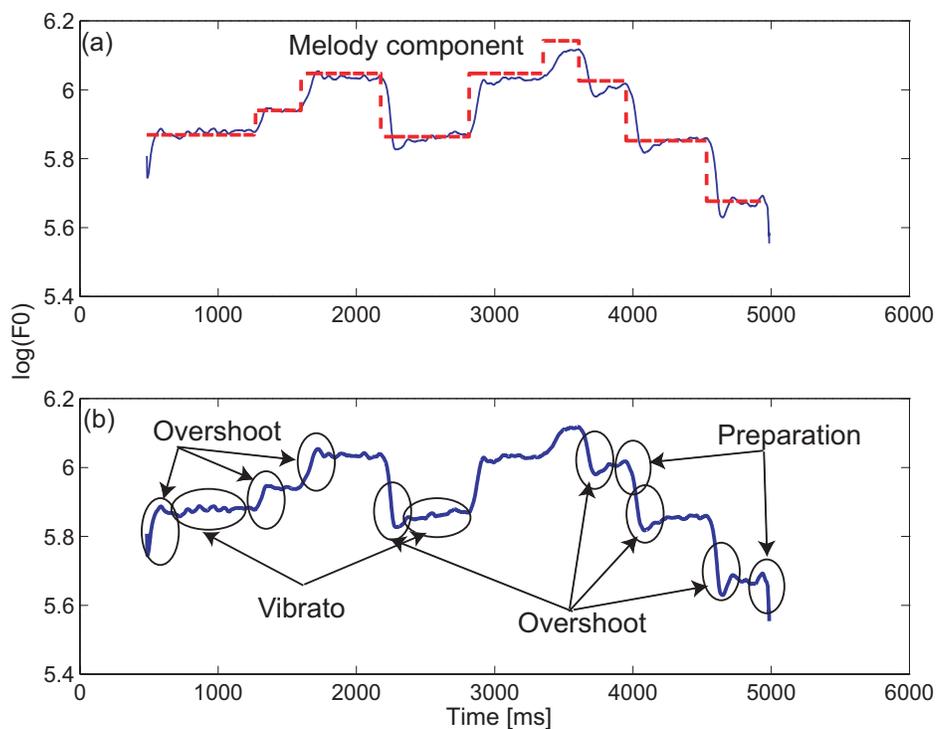


図 3.1: アマチュア歌手の基本周波数変化における (a) メロディ変化, (b) 及び F0 動的変動成分。

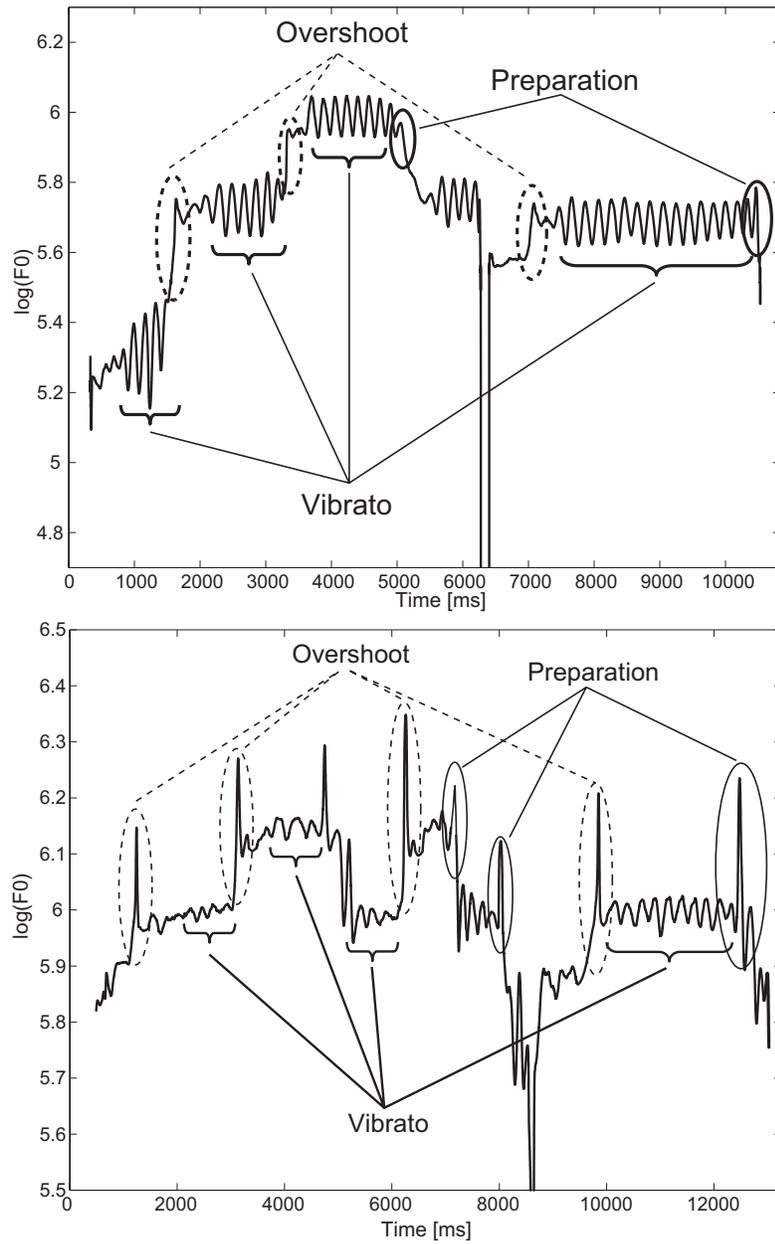


図 3.2: 基本周波数変化に含まれる  $F_0$  動的変動成分 (上: テノール歌唱, 下: 民謡歌唱)。

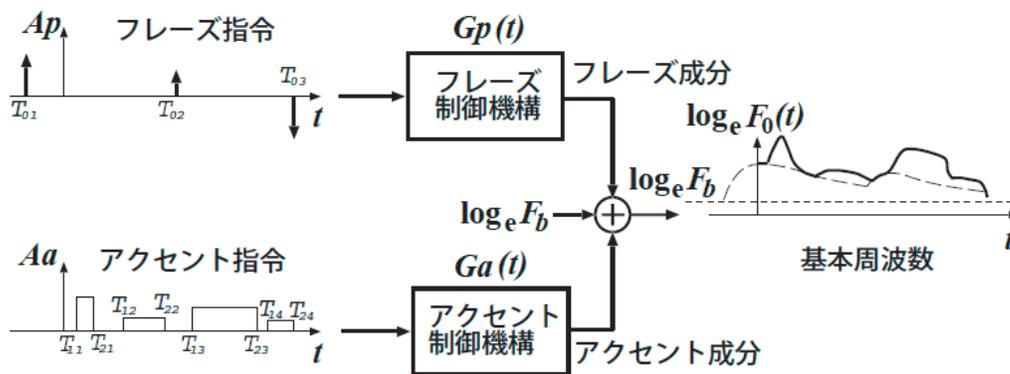


図 3.3: 藤崎モデルの概要 .

### 3.3 歌声の F0 制御モデルの構築

話声を対象とした F0 制御モデルに藤崎モデル [18, 35] がある . このモデルでは , 話声の基本周波数の標準的な変化パターンを , 単語を発声するための呼気の自然な流出に伴うパターン (ほぼ「へ」の字型の , やや前高の山形) であるフレーズ成分に , アクセントの存在を示す強調による部分的上昇パターンであるアクセント成分を重畳することで記述している (図 3.3 参照) . フレーズ成分とアクセント成分は , それぞれ臨界制動 2 次系のインパルス応答とステップ応答で表現されており , 少数のパラメータ設定によって話声の基本周波数特性を自由に操作でき , 正確な基本周波数変化の記述が可能なモデルである . その為 , 音声合成への適用はもとより , 話声における感情や態度といった非言語情報と音声知覚の関係を調査する有効な手段として様々な研究で用いられている [36] .

一方で歌声の基本周波数制御を考えた場合 , メロディに伴った急峻な基本周波数の上昇・下降の制御 , 及びメロディ変化中に存在する F0 動的変動成分の制御が必要となるが , 藤崎モデルでそのような制御を正確に行うのは困難である . 近年では , 歌声の F0 制御モデル構築に向けたいくつかの取り組み [34, 55, 52] が行われているものの , 先に示した複数の F0 動的変動成分を操作することで , 歌声の基本周波数を制御・記述可能なモデルは提案されていない . その為 , 各成分を操作した歌声合成 , 更には各成分が歌声知覚に与える影響を調査する試みはほとんど行われていない状況にある .

歌声の基本周波数制御を考えた場合，メロディ変化を表す矩形的な変化と，F0 動的変動成分のような動的变化を如何に制御・記述するかが重要となる．そこで，本節では，以下の2つ条件を満たすシステムを考える事で，歌声のF0 制御モデルの構築を行なった．

1. メロディ変化，及び4つのF0 動的変動成分-オーバーシュート，プレパレーション，ヴィブラート，微細変動-を独立に制御可能なこと
2. 個々のF0 動的変動成分を正確に制御可能なこと

### 3.3.1 歌声 F0 制御モデルの枠組み

我々人間が歌唱する際には，譜面に記載された音符の羅列や歌う曲に関する知識などから，メロディに関する情報は予め持っていることがほとんどである．そして，そのメロディ変化に従って声帯制御を行う過程において，前章で示したF0 動的変動のような種々の変化が生じていると考えられる．そこで，本節の冒頭で示した1つ目の条件を満たすF0 制御モデルを構築する為の方策として，矩形的なメロディの変化に対して4つのF0 動的変動成分を付加することで歌声の基本周波数変化を記述するシステムとして体系化した．

図 3.4 に提案する歌声 F0 制御モデルの概要を示す．メロディ変化及びF0 動的変動成分は，以下のように記述される．

メロディ成分 (Melody Component) :

ステップ関数の重ね合わせとして記述される．譜面上から作成され，矩形的なメロディ変化を表す．

オーバーシュート (Overshoot) :

制動2次系の減衰振動モデルで記述される．傾斜をもった音高変化 (ポルタメント)，及びその後の目標音高を振り切る振動成分を制御する．

ヴィブラート (Vibrato) :

制動2次系の定常振動モデルで記述される．同一音高が持続する際に生じる周期的な振動を制御する．

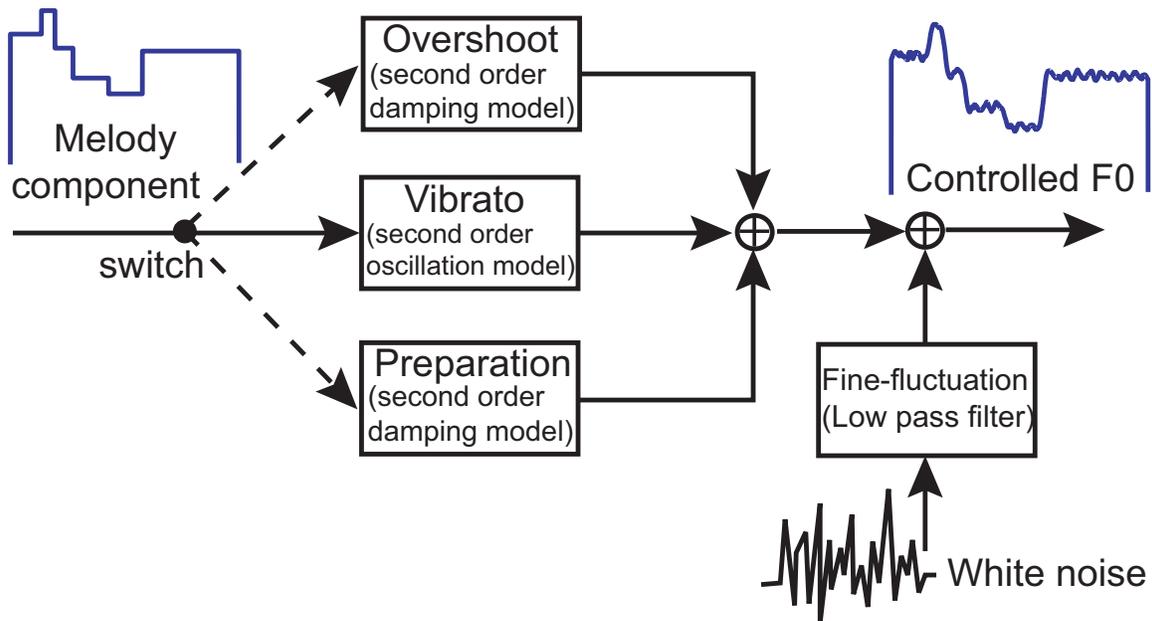


図 3.4: 本論文で提案する F0 制御モデルの概要 .

プレパレーション ( Preparation ) :

制動 2 次系の減衰振動モデルで記述される . 音高変化直前の変化とは逆方向へ瞬時的に変動する成分を制御する .

微細変動 ( Fine fluctuation ) :

白色雑音を元に記述する . 基本周波数変化全体に不規則で細かい振動を付与する .

### 3.3.2 F0 動的変動成分の制御

F0 動的変動成分の制御を行う上での基本的なアイデアは , オーバーシュートとプレパレーションを 2 次系減衰振動モデル , ヴィブラートを 2 次系定常振動モデルで記述することである . これにより , 臨界制動 2 次系で構成される藤崎モデルでは記述が困難であった F0 動的変動成分が制御可能となる .

制動 2 次系伝達関数は , 次式で記述される .

$$H(s) = \frac{K}{s^2 + 2\zeta\Omega s + \Omega^2}, \quad (3.1)$$

ここで、 $\Omega$  は固有角周波数、 $\zeta$  は減衰項、そして  $K$  は振幅項 ( $= \alpha\Omega^2$ ) である。そして、このシステムのインパルス応答  $h(t)$  は、 $\zeta$  の値に従って次式のように表される。

- $|\zeta| > 1$  のとき：指数減衰

$$h_1(t) = \frac{K}{2\sqrt{\zeta^2 - 1}} (\exp(\lambda_1\Omega t) - \exp(\lambda_2\Omega t)), \text{ 但し } \lambda_{1,2} = -\zeta \pm \sqrt{\zeta^2 - 1}, \quad (3.2)$$

- $|\zeta| < 1$  のとき：減衰振動

$$h_2(t) = \frac{K}{\sqrt{1 - \zeta^2}} \exp(-\zeta\Omega t) \sin(\sqrt{1 - \zeta^2}\Omega t), \quad (3.3)$$

- $|\zeta| = 1$  のとき：臨界制動

$$h_3(t) = Kt \exp(-\Omega t), \quad (3.4)$$

- $|\zeta| = 0$  のとき：定常振動

$$h_4(t) = \frac{K}{\Omega} \sin(\Omega t), \quad (3.5)$$

## オーバーシュートの制御

オーバーシュートを制御する機構は、(3.3) 式の減衰振動モデルのインパルス応答  $h_2(t)$  で与えられる。そして、オーバーシュートを含んだ基本周波数変化 (overshoot

component)  $F0_{os}$  は、メロディ成分のステップ関数  $M_d$  を入力とした、このシステムの過渡応答特性として次式で表現される。

$$F0_{os}(t) = M_d(t) * h_2(t), \quad [\text{Hz}] \quad (3.6)$$

そして、この式で記述されるオーバーシュートの特性は、パラメータ  $\Omega$  と  $\zeta$  の値によって制御される。

$\Omega$  は、主に変化速度（ポルタメントの傾き）を制御する。 $\Omega$  の値が大きくなれば、目標音高値を越えた振動（オーバーシュートエクステント）がピークに達するまでの時間  $T_p$  が早くなる。その際、 $T_p$  は次式で近似される。

$$T_p = \frac{\pi}{\Omega \sqrt{1 - \zeta^2}}, \quad [\text{ms}] \quad (3.7)$$

ここでは、 $|\zeta| < 1$  なので、 $T_p$  の大きさは、 $\Omega$  の値に大きく依存する形となる。

$\zeta$  は、オーバーシュートエクステントの大きさを制御する制御パラメータ値である。 $\zeta$  の値が小さければ振動は大きくなり、小さければ臨界制動に近い変化を表す。振動がピークに達する時刻  $T_p$  における変動の大きさ  $A_{os}$  は次式で与えられる。

$$A_{os} = Ftr_i \exp\left(\frac{-\zeta\pi}{\sqrt{1 - \zeta^2}}\right), \quad [\text{Hz}] \quad (3.8)$$

ここで、 $Ftr_i$  は、メロディ開始から  $i$  番目の音高変化における基本周波数の変化量 [Hz] を示す。

また、両者のパラメータ値によって、オーバーシュートの持続時間  $T_{os}$  が制御され、次式で与えられる。

$$T_{os} = \frac{2\pi - \varphi}{\Omega \sqrt{1 - \zeta^2}}, \quad [\text{ms}] \quad (3.9)$$

以上述べたオーバーシュート特性と制御パラメータ値の関係を図 3.5 に示す。

## ヴィブラートの制御

ヴィブラートを制御する機構は、(3.5) 式の定常振動のインパルス応答  $h_4(t)$  によって記述され、ヴィブラート成分（vibrato component） $F0_{vb}$  は次式によって表現される。

$$F0_{vb}(t) = F_n h_4(t), \quad [\text{Hz}] \quad (3.10)$$

ここで、 $F_n$  はヴィブラートを付与する区間の音高値を表す。

定常振動で表現されるヴィブラートの特性は、パラメータ  $\Omega$  の値で振動の速さを表す vibrato rate が決定される。また、 $\Omega$  に対して  $K$  の値を調整することで、振動の大きさを表す vibrato extent が制御される。その為、記述されるヴィブラートの速さと大きさは、1つのヴィブラート区間で一定となる。また、ヴィブラート制御開始時の初期位相は、0 としている。ヴィブラート特性とパラメータの関係を図 3.6 に示す。

## プレパレーションの制御

F0 制御モデルでは、プレパレーションを音高変化に伴う急峻な基本周波数遷移に備えた予備的変動と考えることで制御を行う。その為、その制御機構は、オーバーシュートと同様に (3.3) 式の減衰振動のインパルス応答  $h_2(t)$  で記述され、プレパレーション直後に起る音高変化量のステップ関数入力に対する過渡振動成分として生成される。図 3.7 にプレパレーション成分 (preparation component)  $F0_{pre}$  を示す。

このシステムへの入力は、メロディ成分を過渡応答の立ち上がり時間  $T_r$  だけ遅らせ、時間反転させたものである。 $T_r$  は次式に従う。

$$T_r = \frac{\pi - \varphi}{\Omega \sqrt{1 - \zeta^2}}, \quad [\text{ms}] \quad (3.11)$$

また、生成されるプレパレーションの大きさ  $A_{pre}$  と持続時間  $T_{pre}$  は次式に従う。

$$A_{pre} = Ftr_{i+1} \exp\left(\frac{-\zeta\pi}{\sqrt{1 - \zeta^2}}\right), \quad [\text{Hz}] \quad (3.12)$$

$$T_{pre} = \frac{\pi}{\Omega \sqrt{1 - \zeta^2}}, \quad [\text{ms}] \quad (3.13)$$

ここで、 $Ftr_{i+1}$  は、プレパレーション直後に起る音高変化時の基本周波数変化量を表す。

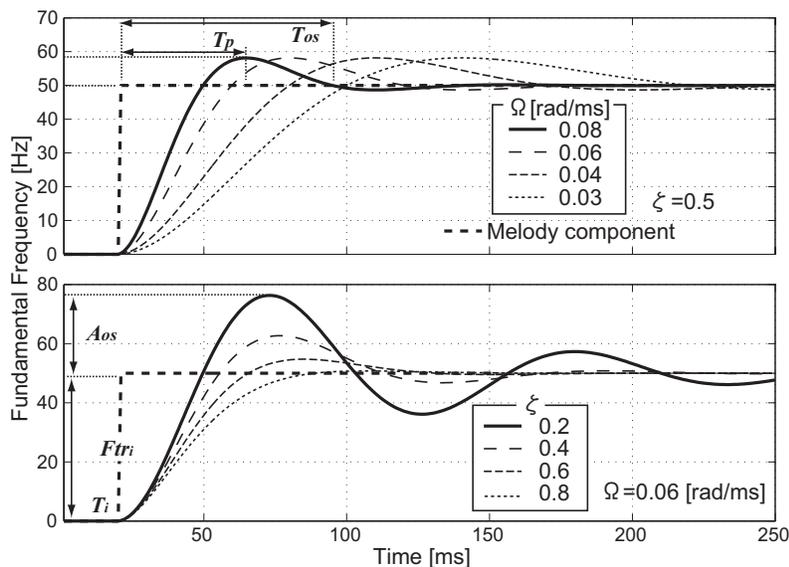


図 3.5: オーバースhootの特性と制御パラメータ値の関係 .

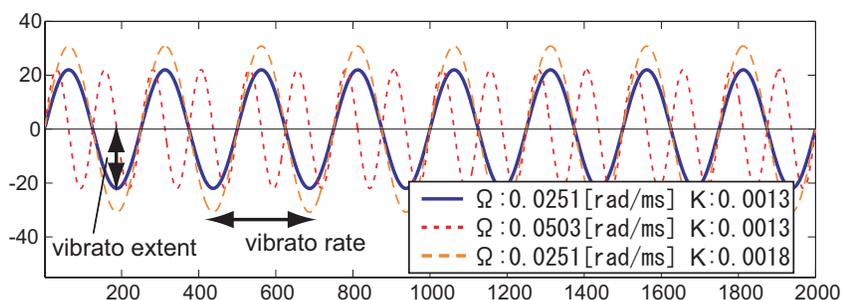


図 3.6: ヴィブラートの特性と制御パラメータ値の関係 .

## 微細変動の制御

微細変動の制御システムは、2次系システムに従わず、白色雑音を基に作成し、基本周波数変化全体に付与する制御を行っている。その特性は、Akagiらの先行研究[24]を参考に決定した。それによると、歌声に含まれる微細変動成分の変調周波数は最大で15～20 Hzで、その偏移幅(振幅)は平均で±20 centである。そこで、本研究では白色雑音をカットオフ周波数10 Hz、減衰特性-20 dB/octのローパスフィルタに通し、最大振幅が5 Hzになるように調整することで微細変動を表現した。

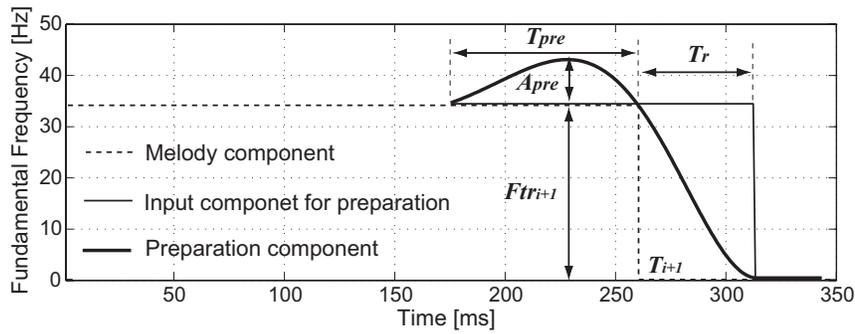


図 3.7: プレパレーション制御の概要 .

### 3.3.3 基本周波数変化の記述過程

前項で述べた各 F0 動的変動成分を制御する機構は，入力メロディ成分に対し，ある特定の区間において駆動するものである．図 3.8 に，各 F0 動的変動成分の生成過程を記す．図中の  $T_{mi}$  は，メロディ成分における  $i$  番目の音高変化時刻を表し，この時間を基準に各 F0 動的変動成分を付与する区間が決定される．

オーバーシュートが制御される区間は，音高変化時刻  $T_{mi}$  から  $T_{os}$  の区間である． $T_{os}$  はオーバーシュートの持続時間長を表し，(3.9) 式で求められる．

$i$  番目の音高変化直前のプレパレーションは， $T_{mi} - T_{pre}$  から  $T_{mi}$  の区間で制御される． $T_{pre}$  はプレパレーションの持続時間長を表し，(3.13) 式で求められる．

ヴィブラートに関しては，オーバーシュートとプレパレーションが制御される以外の区間  $T_{vb}$  において，以下の条件が満たされる場合のみ制御される．

$$T_{vb} \geq 1.5 \times f_{vr} \quad (3.14)$$

ここで， $f_{vr}$  は制御するヴィブラートの振動周期を示すものであり， $f_{vr} = \frac{2\pi}{\Omega}$  で表される．つまり，1.5 周期以上の振動を付与できる区間を  $T_{vb}$  として設定し，ヴィブラートを付与している．

最後に，微細変動成分に関しては，上記 3 つの F0 動的変動成分の制御が行われた後に基本周波数変化全体に付与することで制御している．

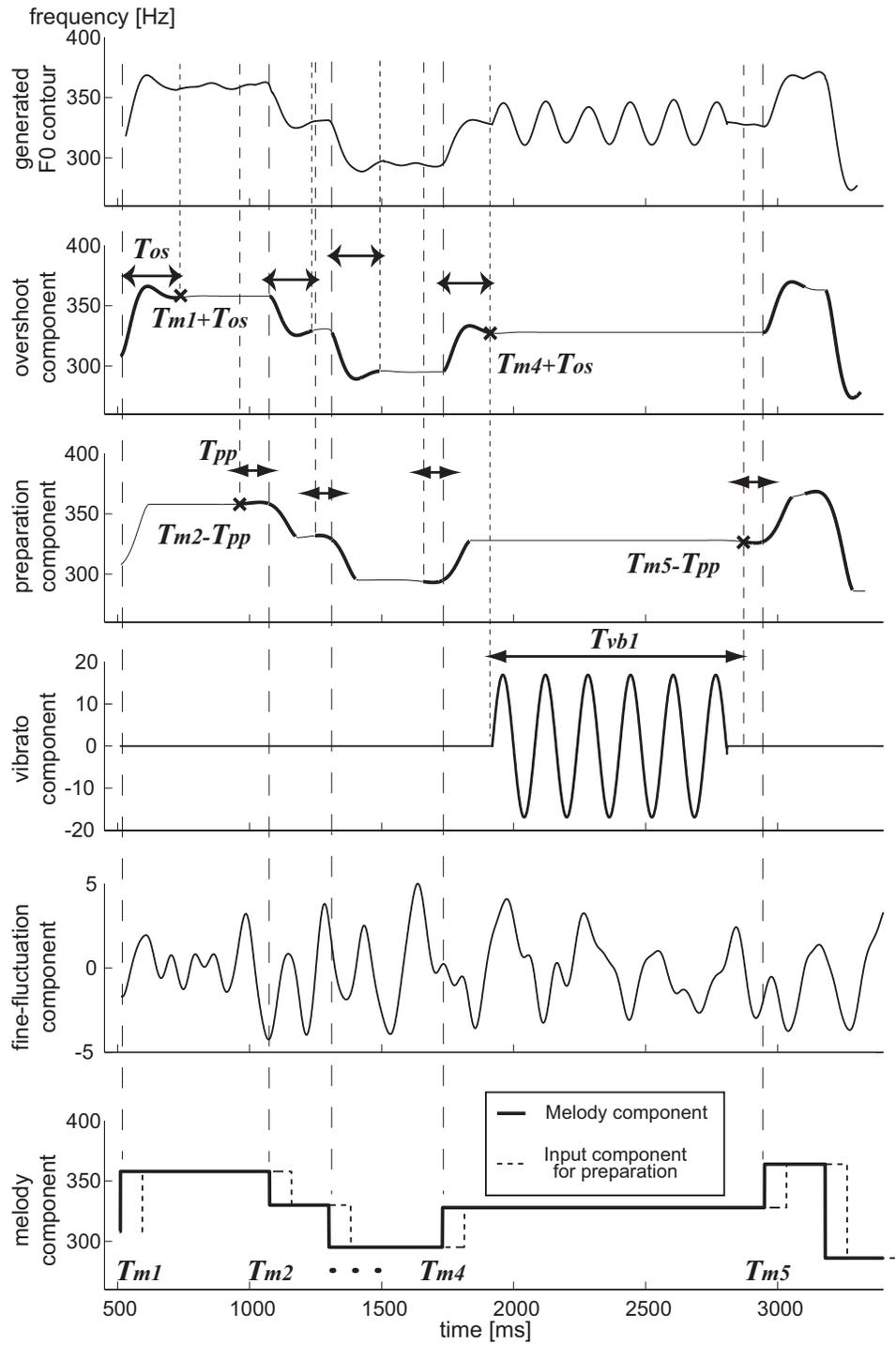


図 3.8: 歌声 F0 制御モデルによる基本周波数変化の記述過程の概要 .

### 3.4 歌声 F0 制御モデルの検証

本論文では、前節で提案した歌声 F0 制御モデルによって F0 動的変動成分を変化させ、その変化に対する歌声知覚の反応を調べることで各成分と歌声知覚の関係を調べる。この枠組みで実験を行うためには、歌声 F0 制御モデルが前節の冒頭で記した 2 つ目の条件「個々の F0 動的変動成分を正確に制御可能なこと」を満たす必要がある。

そこで本節では、F0 制御モデルで記述される基本周波数変化と原音声の基本周波数変化のフィッティング誤差を求めることで、個々の F0 動的変動成分が正確に記述可能か否かを検証する。またこれによって、各成分を記述する為の制御パラメータ値を抽出する。

#### 3.4.1 フィッティング処理

F0 制御モデルで記述される基本周波数変化の原音声の基本周波数変化に対するフィッティングは、非線形最小二乗法 [64] によって行った。その際の平均二乗誤差は、次式で与えられる。

$$E = \sqrt{\frac{a}{N} \sum_{m=M}^{M+N} (x(mT) - y(mT))^2}, \quad (3.15)$$

ここで、 $x(t)$  は F0 制御モデルによって生成された基本周波数変化を、 $y(t)$  は TEMPO で抽出した基本周波数変化を示す。また、 $T = 1/f_s$  ( $f_s$  はサンプリング周波数) である。

ここでは、F0 制御モデルによって記述される個々の F0 動的変動成分の精度を検証するために、各 F0 動的成分ごとにフィッティング処理を行っている。つまりは、各成分が存在する区間においてのみ平均二乗誤差を計算し、その誤差値を元に各成分の制御精度について検証を行う。その為、式中の  $M$  と  $N$  は、フィッティング対象とする個々の F0 動的変動成分が存在する区間の開始時刻とその区間長を示している。

各 F0 動的変動成分に対するフィッティングは、歌声データセット A に収録されている女性 3 名の歌声と、歌声データセット B に収録されているすべての洋楽歌唱及び演歌とわらべ歌を対象に、以下の条件で行った。

## オーバーシュートとプレパレーションに関して

- フィッティング処理を行う前に，原音声の基本周波数変化をローパスフィルタ（カットオフ周波数 10 Hz）に通す：  
この操作により，フィッティング区間に含まれる微細変動成分を除去する．
- フィッティング対象区間として，オーバーシュートが  $M = T_{mi}[\text{ms}]$ ,  $N = 250 [\text{ms}]$ ，プレパレーションが  $M = T_{mi} - 200[\text{ms}]$ ,  $N = 200 [\text{ms}]$ ．
- $K$  と  $\Omega$  を同じ値として設定する：  
経験的に設定した制約である．しかし，この制約によって，設定する制御パラメータ値の削減を可能にしている．
- 音高上昇・下降時のどちらにおいても同じパラメータを用いて制御する：  
オーバーシュートに関する先行研究において，その特性が音高上昇時と下降時で異なることが報告されている．しかしここでは，オーバーシュートの平均的な特性を制御するパラメータ値を導出する．
- 各制御パラメータの変化幅は， $\Omega$  が  $0.01 \sim 0.1 \text{ rad/ms}$ ， $\zeta$  が  $0.01 \sim 0.99$ ，刻み幅はそれぞれ  $0.0001 \text{ rad/ms}$ ， $0.001$ ．

## ヴィブラートに関して

- フィッティング処理を行う前に，原音声の基本周波数軌跡をローパスフィルタ（カットオフ周波数 10 Hz）に通す：
- フィッティング対象区間は， $M = T_{mi} + 250 [\text{ms}]$ ,  $N = T_{mi+1} - T_{mi} + 50 [\text{ms}]$ ．  
つまり，オーバーシュートとプレパレーションのフィッティングを行う以外の区間．
- 各制御パラメータの変化幅は， $\Omega$  が  $0.003 \sim 0.08 \text{ rad/ms}$ ， $K$  が  $0.0003 \sim 0.015$ ，刻み幅はそれぞれ  $0.0001 \text{ rad/ms}$ ， $0.0001$ ．

表 3.1: 各 F0 動的変動成分に対するフィッティング処理における平均二乗誤差 (RMSE:Root Mean Squared Error) .

F0 fluctuation	RMSE [Hz]
overshoot	3.28
vibrato	12.80
preparation	2.15

表 3.2: 歌声データセット A, B 全体に対するフィッティング処理によって導出された F0 制御モデルの制御パラメータ値 .

F0 fluctuation	$\Omega$ [rad/ms]	$\zeta$	$K$
overshoot	0.0348	0.5422	0.0348
vibrato	0.0345	—	0.0013
preparation	0.0292	0.6681	0.0292

### 3.4.2 F0 動的変動成分の正確さに関する検証

上記の方法でフィッティングを行った結果, 各 F0 動的変動成分の制御区間において表 3.1 に示す二乗誤差が得られた. また表 3.2 に, 表 3.1 の二乗誤差が得られた際の各 F0 動的変動成分の制御パラメータ値を示す. 更に, 制御パラメータを用いて記述された基本周波数変化の一例を図 3.9 に示す.

#### オーバーシュートについて

表 3.1 に示すように, オーバーシュートに関するフィッティング誤差は 3.28 Hz と非常に小さい結果となった. これより, F0 制御モデルがオーバーシュートを正確に記述可能なことが確認された. また, 様々な歌声データに対してフィッティングした結果誤差が小さかったことから, オーバーシュートが歌声に共通して存在する成分であることが定量的に示され, 表 3.2 の制御パラメータ値によって記述される変動が平均的なオーバーシュート特性であることが明らかとなった. その特

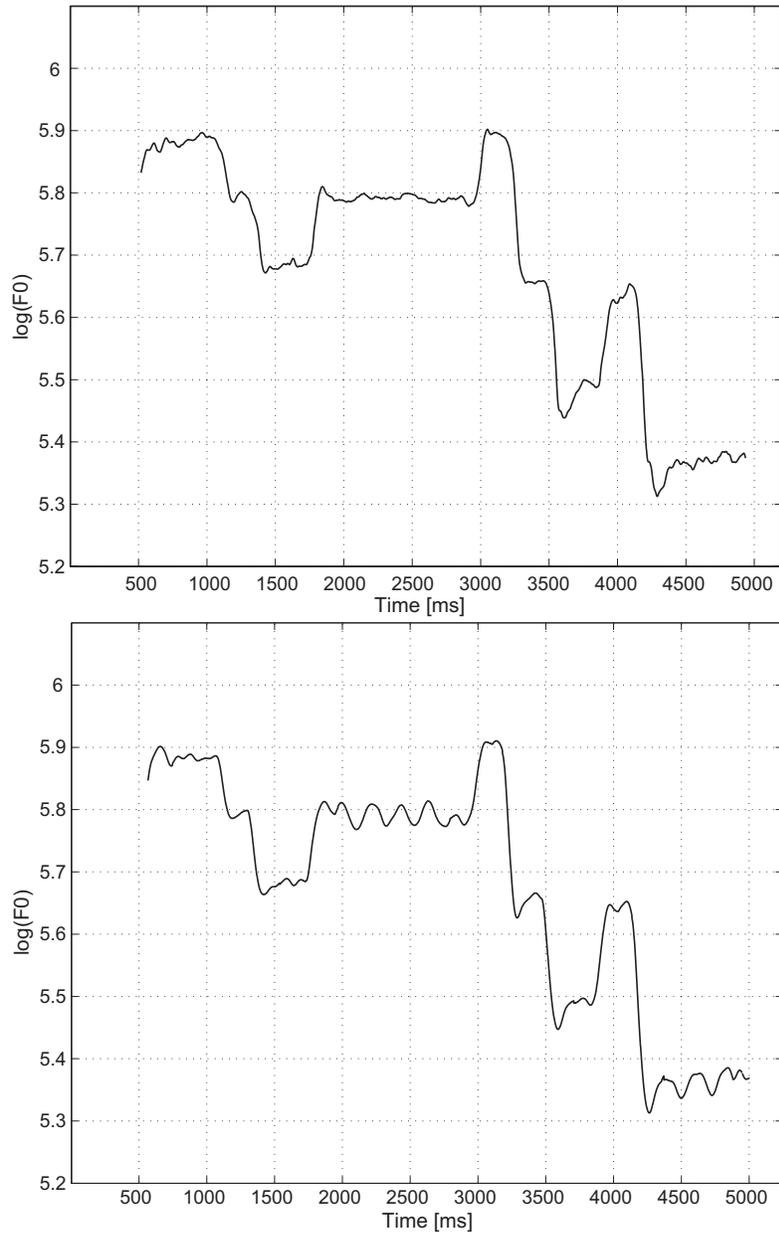


図 3.9: F0 制御モデルで記述された基本周波数変化（上：TEMPO で抽出した原音声の基本周波数変化，下：表 3.2 のパラメータ値によって記述された基本周波数変化）。

性としては，目的音高を越えてピークに達するまでに要する時間  $T_p$  は約 107 ms，そのピーク値と目的音高値の差が音高変化量の約 13 %，そして，変動が収束するまでに要する時間が 180 ms である．

## ヴィブラートについて

ヴィブラートに関しては，フィッティング誤差が収束しなかった為，歌声データに含まれるヴィブラートを対象に vibrato rate と extent の分析を行うことで，平均的なヴィブラート特性を求めた．そして，分析で得られた特性を F0 制御モデルによって記述した場合の二乗誤差を求めたが，他の成分に比べ誤差が大きい結果となった．この原因としては，以下の 2 つのヴィブラート特性の存在が考えられる．

1. 歌唱技術の違いによってヴィブラートの特性が大きく異なる：

この為，フィッティング処理や平均的な特性を制御することでは誤差が小さくならないことが考えられる．

2. 実際のヴィブラートが定常振動ではなく時間的に特性が変動する：

この為，F0 制御モデルによって定常振動で記述されたヴィブラートとの間に誤差が生じることが考えられる．

それぞれのヴィブラート特性を裏付ける報告として，Bjorklund は，ヴィブラートが歌唱トレーニングを積み重ねることで自然に発生する成分であることを示し [28]，ある程度の歌唱技術が無いとヴィブラートが存在しないことを示唆している．また，Prame は，vibrato rate が時間的に一定でなく，変動していることを示している [62]．しかし，どちらの特性についても，今だ明確な解答は得られていないのが現状である．そこで，次章においては，先ず本章で提案した手法によってヴィブラート制御を行い，ヴィブラートを定常振動と仮定した上で歌声知覚との関係を調査する．その後，上記 2 つのヴィブラート特性について検討することで，正確にヴィブラートを制御する方法を構築すると同時に，これまで議論の余地が残っていた上記 2 つのヴィブラート特性を明らかにしていく方略をとる．尚，分析の結果得られた平均的なヴィブラート特性は，vibrato rate が 5.4 Hz，vibrato extent がヴィブラート区間の音高値の 3.7 % (約 62 cent) である．これらは，先行研究 [71, 63] で報告されている平均的なヴィブラート特性，vibrato rate: 4 ~ 7 Hz，vibrato extent: 71

cent に近いものとなっている。

## プレパレーションについて

プレパレーションにおけるフィッティング誤差は、3つの成分の中で最も小さい結果となった。これより、F0 制御モデルがプレパレーションを正確に記述可能であることが確認され、更にプレパレーションも多くの歌声に共通して含まれる成分であることが実証された。また表 3.2 から、音高が変化する約 109 ms 前から、音高変化量の約 5 % の大きさを振動する特性が、平均的なプレパレーションの変動であることが確認された。

## 3.5 まとめ

本章では、歌声を知覚する上で重要な役割を担っている F0 動的変動成分を明らかにするための最初の手続きとして、F0 動的変動成分を操作することで歌声の基本周波数変化を制御・記述可能な F0 制御モデルを構築した。

最初に、本論文で対象とする F0 動的変動成分を定義した。ここでは、アマチュア歌手と、プロの洋・邦楽歌手の歌声から抽出された基本周波数変化を対象に、先行研究で報告された F0 動的変動成分や、それ以外の特徴的な成分が存在するか否かを調査した。その結果、先行研究で報告されているオーバーシュート、ヴィブラート、微細変動が多くの歌声に共通して含まれる成分であることが確認された。更に、新たな F0 動的変動成分として、音高変化直前に変化とは逆方向に振動する成分のプレパレーションが多くの歌声において観測されることを確認した。以上から、これら 4 つの成分を本論文で対象とする F0 動的変動成分として定義した。

次に、歌声の F0 制御モデルを構築した。ここでは、矩形的に変化するメロディ成分に対して、前節で定義した 4 つの F0 動的変動成分を付与することで歌声の基本周波数軌跡を記述可能なモデルを構築した。臨界制動 2 次系で話声の基本周波数を制御する藤崎モデルに習い、オーバーシュート、プレパレーション、ヴィブラートを制動 2 次系のインパルス応答で記述することで、各成分を独立に操作し、歌声の基本周波数変化を制御・記述可能なモデルとして体系化した。

最後に、歌声 F0 制御モデルで F0 動的変動成分を正確に記述可能か否かを検証した。ここでは、F0 制御モデルで記述される基本周波数変化と、原音声の基本周波数変化のフィッティング誤差値を求めることで、各成分が正確に制御可能であるか調査した。その結果、オーバーシュートとプレパレーションに関しては、制御パラメータ値を適切に設定することで、原音声に含まれる変動に極めて近い特性で記述可能なことが確認された。一方ヴィブラートに関しては、本手法で表現される定常振動と先行研究で報告されているヴィブラート特性が相違している為、フィッティング誤差が他の成分に比べ大きい結果となった。この問題に関しては、次章で検討していくことにする。

以上から、歌声共通の F0 動的変動成分としてオーバーシュート、ヴィブラート、プレパレーション、微細変動の存在が確認された。更に、これら F0 動的変動成分を記述可能な歌声 F0 制御モデルが構築された。これにより、次章では、歌声 F0 制御モデルを用いた歌声合成システムを構築し、制御パラメータ値を変化させた時の歌声知覚の反応を調査することによって、F0 動的変動成分と歌声知覚の関係について詳細に検討する。

## 第 4 章

# F0 動的変動成分と歌声知覚の関係の 調査

## 4.1 はじめに

本章では、基本周波数制御のみを考慮した歌声合成システムを提案し、各成分を操作した歌声合成音の作成、及び心理物理実験による合成音の評価を行うことで、 $F_0$  動的変動成分と歌声知覚との関係を調査する。これにより、歌声知覚における各成分の重要性について検証する。

はじめに、前節で提案した歌声  $F_0$  制御モデルを用いた歌声合成システムを提案する。ここでは、 $F_0$  制御モデルを Klatt formant synthesizer に適用した合成法と、STRAIGHT に適用した合成法を提案する。

次に、提案した歌声合成システムを基に、各  $F_0$  動的変動成分の有無による歌声知覚の反応を調べる。はじめに、原音声の基本周波数変化から各  $F_0$  動的変動成分を除去した場合の歌声知覚の反応を調べる。次に、メロディ成分のみを持つ矩形的な基本周波数変化に各  $F_0$  動的変動成分を付与した場合の歌声知覚の反応を調べる。これらの実験により、 $F_0$  動的変動成分が歌声知覚に与える影響を定量的に明らかにする。また、各成分が歌声知覚へ与える影響の度合いについても検証する。

次に、歌唱技術と  $F_0$  動的変動成分の特性の関係について調査する。はじめに、アマチュアとプロそれぞれの歌唱者による歌声データから制御パラメータ値を抽出し、歌唱技術の違いによって各成分の特性に差異があるか否かについて検証する。次に、各歌声データから抽出した制御パラメータ値を用いた歌声合成を行い、これら合成音を心理物理実験によって評価する。この実験により、前章で課題として挙げた歌唱技術の違いとヴィブラート特性の関係を明らかにする。また、オーバーシュートとプレパレーションに関しても、歌唱技術の違いと各成分の特性の対応関係についても検証する。

次に、個々の  $F_0$  動的変動成分の特性を変化させた場合の歌声知覚の反応を調査する。ここでは、オーバーシュート、プレパレーション、ヴィブラートの制御パラメータ値を様々に変化させた歌声合成音を評価する。これにより、各成分の特性と歌声知覚の関係について検証する。

次に、前章で課題として挙げたヴィブラート特性の時間変動について検証する。ここでは、vibrato rate の時間変動特性に関する先行研究の知見を検証することで、ヴィブラートを正確に制御する方法を構築し、これまで明確にされてこなかった

vibrato rate の時間特性と歌声知覚の関係について検証する。

最後に、音高遷移方向とオーバーシュート特性の関係について検証する。ここでは、先行研究においてオーバーシュート特性が音高上昇時と下降時で特性が異なるか否かで議論が分かれている問題に対して、それら特性の違いと歌声知覚の関係を調査することで解答を得ることを試みる。

## 4.2 歌声合成システムの構築

本節では、歌声 F0 制御モデルを用いた 2 種類の歌声合成システムを構築する。1 つは、F0 制御モデルと Klatt formant synthesizer[48] を用いたシステムである。そしてもう 1 つが、F0 制御モデルを STRAIGHT の分析合成過程に適用したものである。

### 4.2.1 Klatt formant synthesizer を用いた歌声合成システム

図 4.1 に Klatt formant synthesizer を用いた歌声合成システムの構成を示す。このシステムでは、基本周波数変化を歌声 F0 制御モデルで記述し、残りの入力パラメータであるホルマント周波数とその帯域幅、そしてゲイン（パワー）を一定の値で固定することで、基本周波数の変化のみが反映された歌声合成音が作成可能である。その為、スペクトルやパワーの情報に干渉されずに、F0 動的変動成分と歌声知覚の関係について検討可能となる。尚、Klatt formant synthesizer の合成条件は、以下に示す通りである。

- ホルマント：6 ホルマント母音/a/

中心周波数 :800, 1200, 2500, 3500, 4500, 5500 Hz

帯域幅 :中心周波数の 10%

- 励振パルス列： $t_{n+1} = t_n + 1/f_m(t_n)$

基本周波数： $f_m(t)$

パルス位置： $t_n$

- 励振波形：Rosenberg wave
- パワーレベル：80 dB で固定

#### 4.2.2 STRAIGHT を用いた歌声合成システム

図 4.2 に STRAIGHT を用いた歌声合成システムの構成を示す。このシステムでは、STRAIGHT によって歌声データを基本周波数とスペクトル包絡と非周期成分割合の時間系列に分解し、F0 制御モデルで記述した基本周波数変化と入力歌声データのスペクトル包絡と非周期成分割合を用いて STRAIGHT で再合成を行うことで歌声合成音を作成する。このシステムの利点としては、STRAIGHT を用いることで高品質な歌声合成が可能となり、F0 制御モデルで記述した基本周波数を用いた歌声合成音と原音声との比較が可能となる。

しかし、STRAIGHT で歌声合成を行う場合、SPIKES における駆動音源生成について検討を行う必要がある。SPIKES では、駆動音源となるオールパスフィルタの群遅延特性を毎回異なった乱数から作成することにより、ボコーダで問題になっていた合成音声特有のバズ音を軽減させている。しかし、この群遅延操作によるオールパスフィルタの設定は、例えば対象とする音声は女性と男性のように基本周波数が異なる場合には変更する必要がある [5]。これは、基本周波数が大きく変化する歌声にも言える事である。その為、本研究では、歌声合成を行う前に、歌声に適應する群遅延操作のパラメータを決定した。設定に必要なパラメータとしては、群遅延の標準偏差と、群遅延の固定領域と変動領域が遷移する境界周波数の 2 つである。パラメータの決定方法としては、1 つの歌声データ（歌声データ 1 に収録された発声者 MS による歌声）に対して、標準偏差を 3 ms から 0.5 ms まで 0.5 ms 刻みの 6 段階で、境界周波数を 1 kHz から 6 kHz まで 1 kHz 刻みの同じく 6 段階で変化させた全部で 36 個の歌声を作成し、被験者 6 人を対象に聴いてもらった印象から最適なパラメータの組を決定した。その結果から、群遅延の標準偏差を 1 ms 以下に、境界周波数を 5 kHz 以上にするのが理想であることが分かった。よって、この歌声合成システムの群遅延パラメータとして、標準偏差 1 ms、境界周波数を 5.5 kHz に設定して歌声合成を行う。

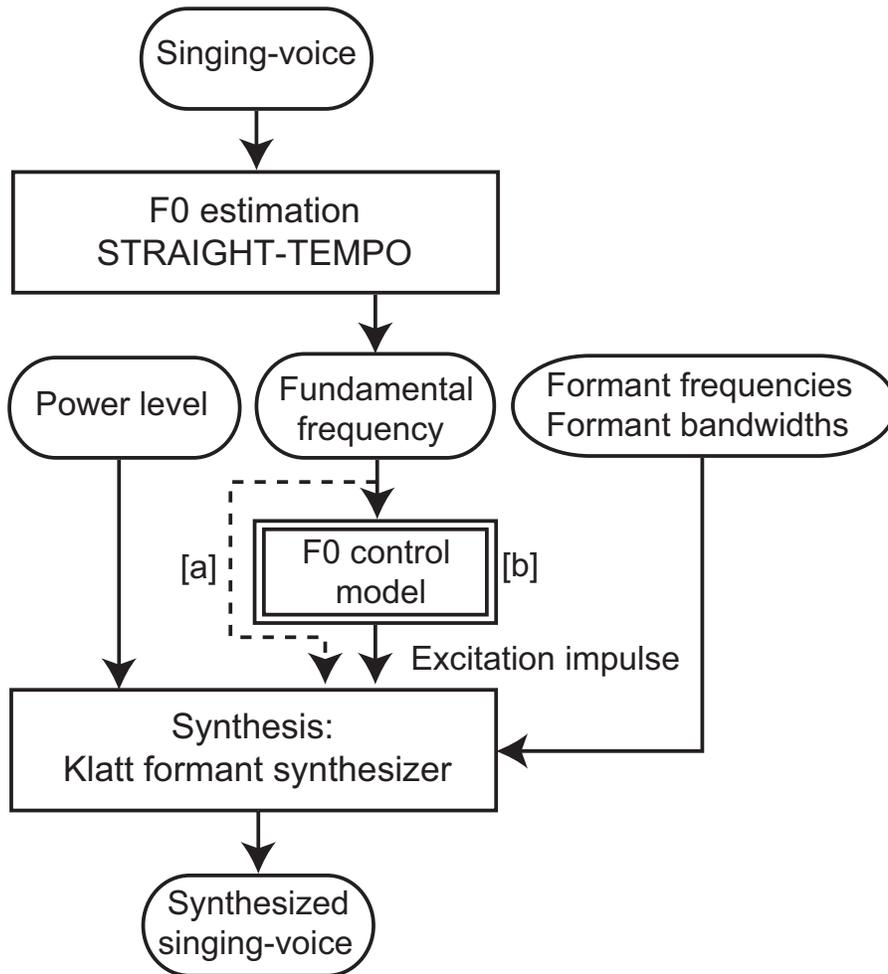


図 4.1: 歌声 F0 制御モデルと Klatt formant synthesizer を用いた歌声合成システム .

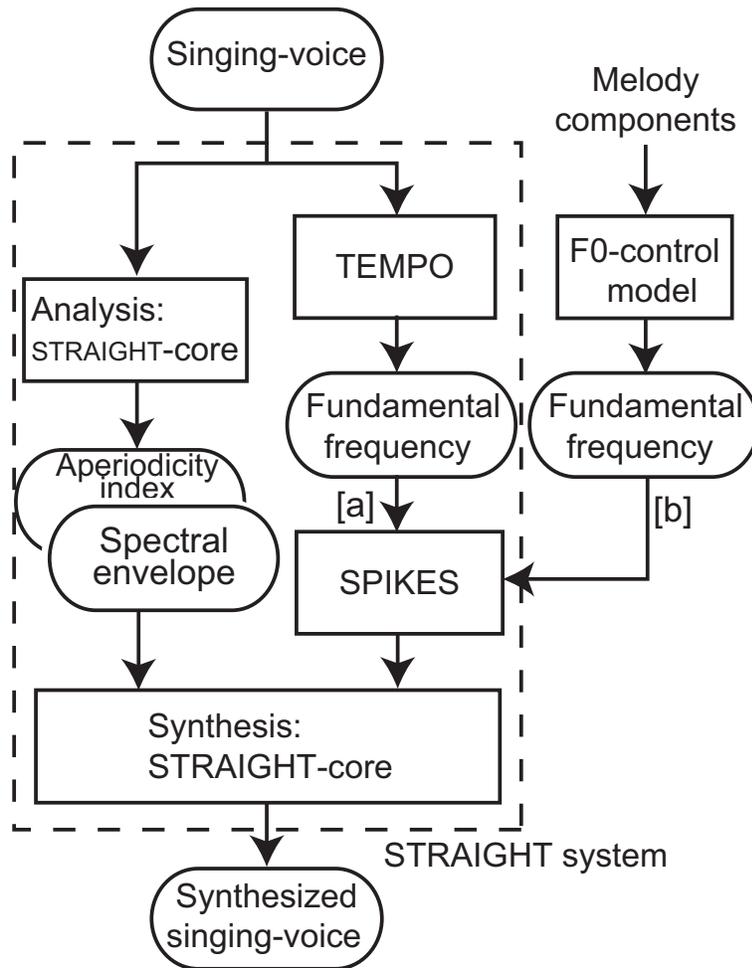


図 4.2: 歌声 F0 制御モデルと STRAIGHT を用いた歌声合成システム .

## 4.3 F0 動的変動成分が歌声知覚に与える影響 1

### —F0 動的変動成分を除去した場合の歌声知覚反応—

本節では、F0 動的変動成分と歌声知覚の関係の最初の検討として、原音声から各 F0 動的変動成分を除去した歌声合成音を作成し、各合成音を心理物理実験によって評価する。この実験では、各成分の除去による歌声の自然性劣化を調査することで、歌声知覚における各成分の重要性を検証する。

#### 4.3.1 歌声合成音の作成

F0 動的変動成分を除去した歌声合成音を作成する為に、図 4.3 に示すような TEMPO による基本周波数推定と、Klatt formant synthesizer を組み合わせた歌声変換合成系を構築した。

この方法は、最初に歌声データの基本周波数を TEMPO で推定し、その基本周波数に対して F0 動的変動成分を除去するための加工を施す。その後、加工した基本周波数を Klatt formant synthesizer に入力することで、個々の F0 動的変動成分が除去された歌声合成音が生成される。

#### 歌声データ

歌声合成の為に用いたデータは、歌声データセット A に収録された女性、男性それぞれ 1 名によって日本語母音/a/のみで歌唱された日本童謡「七つの子」(最初の 8 小節)である。尚、データは 2 小節ごとに分けて使用した。

#### 歌声合成音

上記の方法によって作成した歌声合成音の種類を示す。

- NORMAL

TEMPO で抽出した基本周波数変化に加工を加えず、Klatt formant synthesizer によって合成したもの。(図 4.3 における [a] の行程)

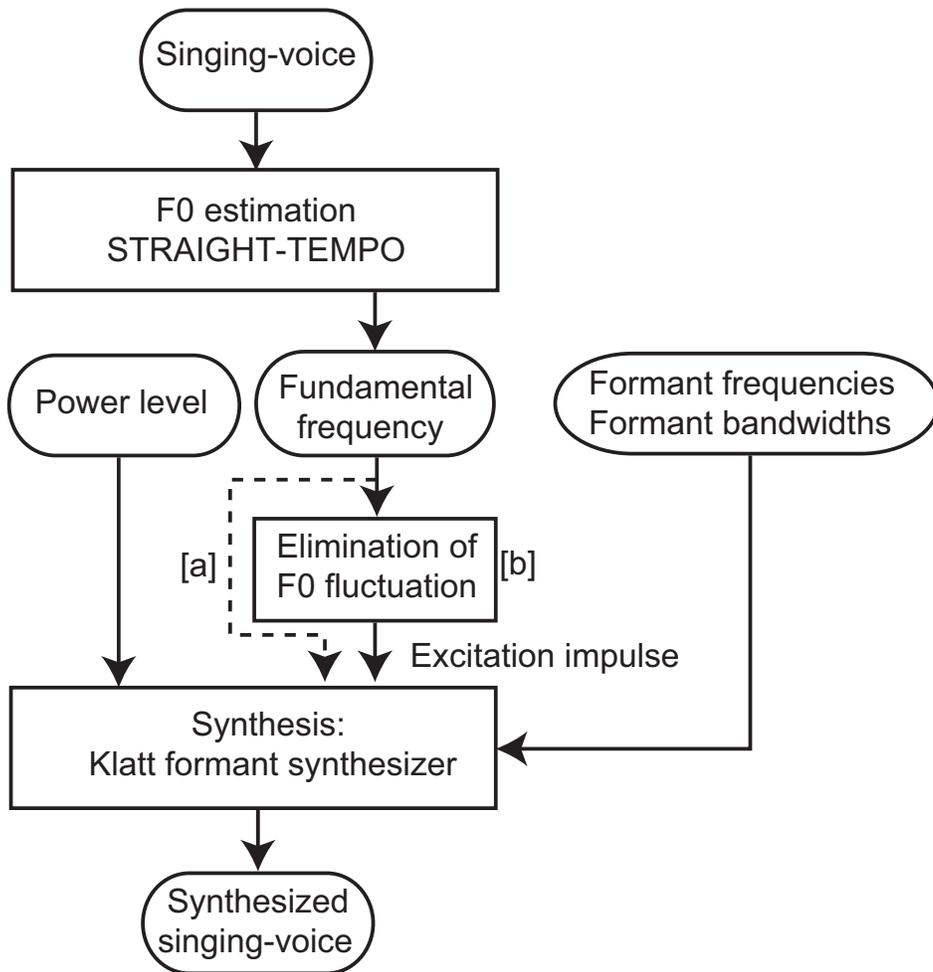


図 4.3: Klatt formant synthesizer を用いた F0 動的変動成分を操作した歌声合成音の作成手順.

- NO-OS  
オーバーシュートエクステントを音高区間の平均値で置き換える事によって除去した合成音 ([b] の行程)
- NO-VB  
音高安定区間を FIR 型のローパスフィルタ (カットオフ周波数 5 Hz) でスムージングし, ヴィブラートと微細変動を除去した合成音 ([b] の行程)
- NO-PRE  
プレパレーションを音高区間の平均値で置き換える事によって除去した合成音 ([b] の行程)
- SMS  
基本周波数全体をローパスフィルタ (カットオフ周波数 5Hz) でスムージングし, 上記の F0 動的変動成分を除去した合成音 ([b] の行程)

それぞれの合成音の基本周波数変化の一例を図 4.4, 4.5 に示す.

#### 4.3.2 聴取実験

聴取実験はシェッフェの一对比較法 [17] によって行った.

心理物理実験において, 一对比較法は, 数個の刺激を 2 つずつ対にして判断を求める方法である. 被験者にとって判断が比較的やさしいので, 判断の信頼性も高く, 適用範囲が広い方法である. また, 実験の所要時間は比較的短くて済み, 2 つの刺激に対して比較判断を求めるので, 刺激間の差が微妙な場合にも適用できる. その中で, 一对比較法にカテゴリー判断を取り入れたシェッフェの一对比較法は, 被験者が対にして提示される刺激を比べて, どちらがどれだけ好きかなどの判断を求めるものである. 一对比較法では, 正規分布の仮定に基づいて序数尺度を間隔尺度に変化する手続きを行うため, 多くの被験者を必要とするが, シェッフェの一对比較法では, 被験者が判断した評価点を序数尺度のまま統計的検討を行うので, 多くの被験者を必要としないという利点がある.

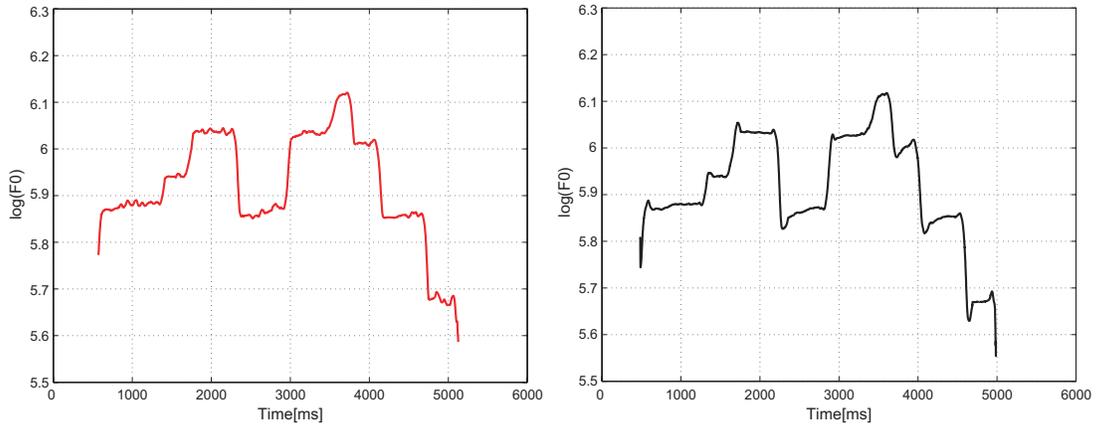


図 4.4: 左:オーバーシュートエクステントを除去した基本周波数変化 (NO-OS) , 右: ヴィブラート・微細変動を除去した基本周波数変化 (NO-VB) .

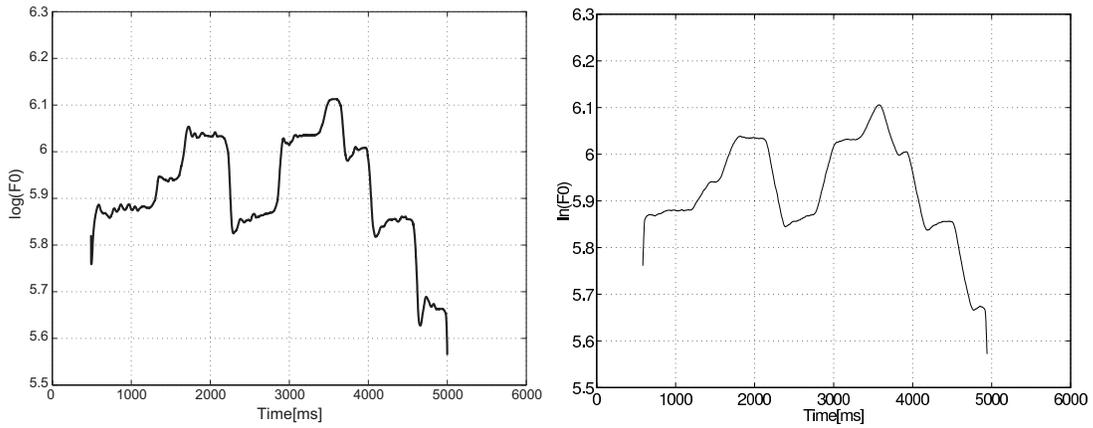


図 4.5: 左:プレパレーションを除去した基本周波数変化 (NO-PRE) , 右:スモーディング処理によりすべての F0 動的変動成分を除去した基本周波数変化 (SMS) .

## 刺激条件

実験で用いる聴覚刺激は、先に示した NORMAL, NO-OS, NO-VB, NO-PRE, SMS を 2 つずつ対したものである。刺激対の数は、1 つの歌声データについて、刺激順序の違いも考慮した  $5 \times 4 = 20$  対である。図 4.6 に刺激の呈示順序を示す。

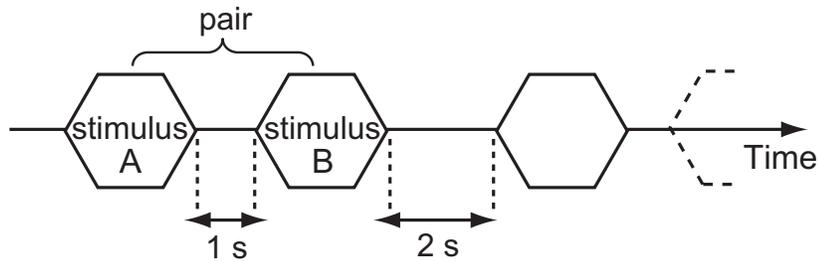


図 4.6: 刺激の呈示順序。

## 被験者

実験に参加した被験者は、オーディオメータを用いた聴力検査で両耳とも健常な聴力を持つとみなされた大学院生 6 名（男性 5 名，女性 1 名）である。

## 実験環境

実験は、図 4.7 に示すように、防音室におけるヘッドホンを介しての両耳受聴で行った。また、その際使用された主な機器を以下に列挙する。

刺激呈示用サーバ : DAT+LINK & Linux(Vine ver1.2)

D/A 変換器 : STAX DAC-TALENT BD

ヘッドホンアンプ : STAX SRM-1/MK-2

ヘッドホン : STAX SR-404

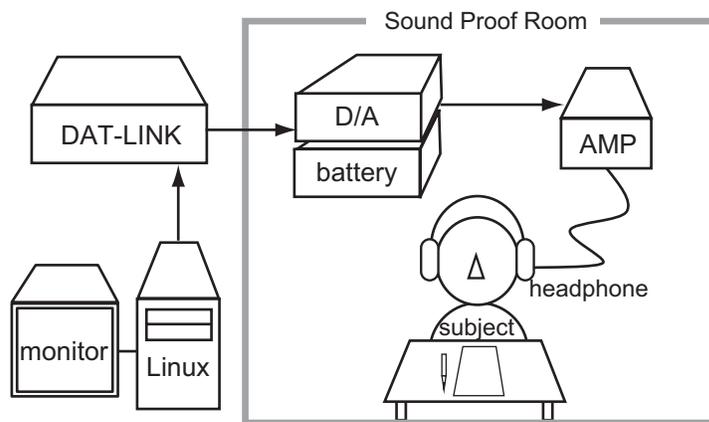


図 4.7: 実験環境の概要 .

### 実験手続き

被験者には次のような教示を与え、歌声の自然性に関して評価してもらった .

ヘッドホンから 2 つの歌を対にして流します . 前の歌と後の歌を聴き比べて、どちらが人間の歌声として自然であるかを、下に記した 7 段階の評価尺度 ( 図 4.8 ) に従って判断してください . 前の歌の方が自然と聴えたら正の値 ( 1 ~ 3 ) に、後の歌が自然と聴えたら負の値 ( -3 ~ -1 ) の当てはまる値にチェックをしてください . どちらも同程度の自然な歌だと判断した場合は 0 を選択してください .

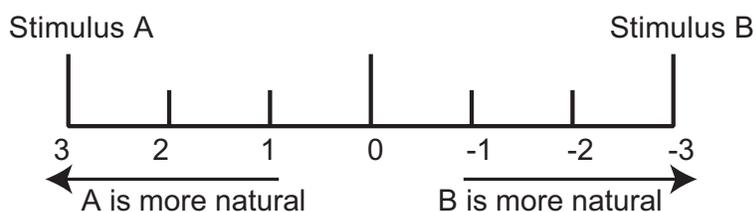


図 4.8: シェッフエの一対比較実験で用いた歌声の自然性に関する七段階評価尺度 .

ここで扱う自然性とは、人間が歌唱した歌声として自然であるか否かを表す心理量である . その為、2 つの刺激それぞれに対して、人間が歌唱した歌声にどれだけ近いかを聴き比べてもらい、距離が近い方をより自然な歌声として判断して

表 4.1: 母数の推定 (自然性) .

歌声	母数
NORMAL	0.72
NO-OS	-0.03
NO-VIB	0.27
NO-PRE	0.25
SMS	-1.21

もらうようにしている．その為，本来なら多次元的な心理量として扱われる自然性を，本研究では1次元まで落とし込むことで扱っている．

### 4.3.3 実験結果と考察

上記の実験方法で得られたデータを，浦の変法 [1] によって処理した結果を表 4.1 に示す．また，表に示した母数の値に従って，5つの刺激の距離関係を直線上で示したものが，図 4.9 になる．母数の値は，刺激がどれだけ人間の歌声として自然に聴えたかを表す値であり，正の大きな値を示せば，その刺激がより自然な歌声である事を表す．尚，F 検定法による各歌声刺激間の有意差検定を行った結果，各刺激間に 5% の水準で有意な差が確認された．

図 4.9 から分かるように，各 F0 動的変動成分を除去することで，歌声合成音の自然性が劣化する結果となった．これにより，本論文で着目している個々の F0 動的変動成分が，歌声知覚に影響を与えていることが確認された．更に，すべての成分を除去した歌声 SMS の自然性が大きく劣化することが明らかとなった．これらの結果から，歌声を知覚する為には，基本周波数変化の概形がメロディに対応しているだけでなく，その変化中に F0 動的変動成分が存在することが必要であることが明らかとなった．

次に，個々の成分の歌声知覚に与える影響について考察する．

オーバーシュートに関しては，目的音高値を越えた過振動成分（オーバーシュートエクステンツ）のみの除去にもかかわらず，自然性劣化が最も大きい結果となっ

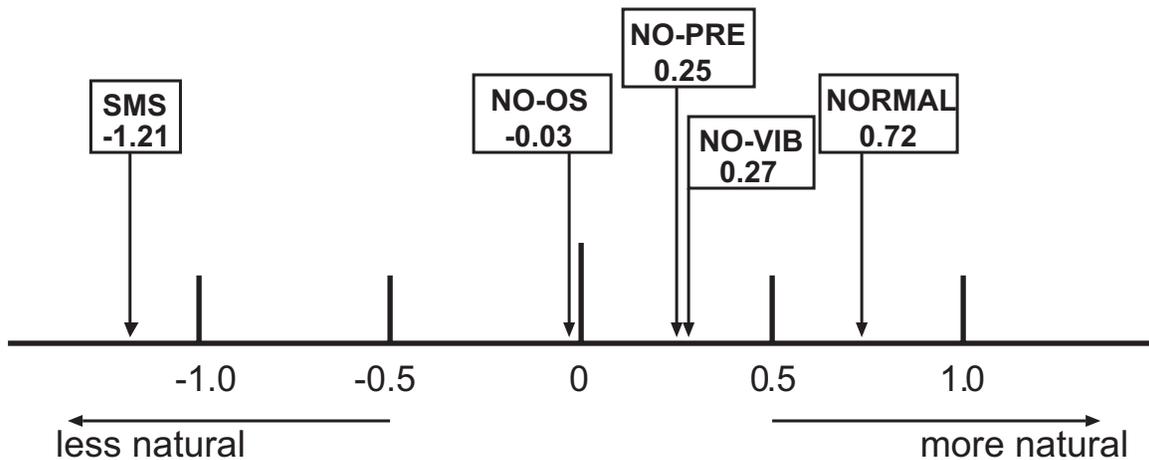


図 4.9: 歌声の自然性の関係 .

た．ここから，傾斜を持った音高変化成分であるポルタメント成分の影響も考慮した場合，オーバーシュート成分が歌声知覚に与える影響はさらに大きくなることが予想される．

本研究で新たな F0 動的変動成分として着目したプレパレーションに関しても，オーバーシュートの次に影響を与えている結果となった．この場合も，オーバーシュートの場合と同様に除去による基本周波数の変化が小さいにも関わらず，自然性劣化は明確に知覚され，歌声知覚におけるプレパレーションの重要性を示す結果となった．

上記 2 つの成分に対して，ヴィブラートや微細変動成分の除去による歌声知覚への影響は，それほど大きくない結果となった．これには，2 つの理由が考えられる．1 つは，本実験で使用したアマチュア歌唱者のヴィブラートが限られた音高区間においてのみ存在していた為に，音高変化直前・直後に必ず存在するオーバーシュートやプレパレーションを除去した場合に比べ，自然性の劣化が小さかったと考えられる．また，NO-VIB を作成する際の同一音高区間のみを対象としたスムージング処理により，それ以外の区間に含まれる微細変動の影響が NO-VIB に含まれていることも，もう 1 つの理由として考えられる．しかし，すべてを除去した SMS の自然性が NORMAL に比べ大きく劣化していることから，ヴィブラート，微細変動共に歌声を知覚する上で重要な音響的特徴であることは間違い無いと考えられる．

## 4.4 F0 動的変動成分が歌声知覚に与える影響 2

### —F0 動的変動成分を付与した場合の歌声知覚反応—

前節では、原音声から F0 動的変動成分を除去した場合の歌声知覚の反応を調査した。本節では、メロディを表す矩形的な基本周波数変化（歌声 F0 制御モデルにおけるメロディ成分）に個々の F0 動的変動成分を付与した歌声合成音を作成し、各合成音を心理物理実験によって評価する。この実験では、各成分を付与することで歌声合成の自然性が向上するか否かを調査することで、歌声知覚における各成分の重要性を検証する。

本節では、図 4.1 に示した Klatt formant synthesizer を用いた歌声合成システム、及び図 4.2 に示した STRAIGHT を用いた歌声合成システムそれぞれを用いて個々の F0 動的変動成分を付与した歌声合成音を作成し、聴取実験によって評価する。

#### 4.4.1 Klatt formant synthesizer を用いた歌声合成システムによる検証

4.2 節で記したように、Klatt formant synthesizer を用いた歌声合成システムでは、ホルマントとパワーの情報を固定した条件で合成を行うことで、合成された歌声は基本周波数変化の違いのみが反映されたものになる。この為、F0 動的変動成分の付与による影響のみが歌声合成の音質の違いとなって現れ、各成分と歌声知覚の関係を詳細に調査することが可能となる。また、前節の実験と同じ合成条件に設定することで、両者の実験結果を比較検討することが可能となる。

##### 歌声合成音の作成

図 4.1 に示した歌声合成システムにより、各 F0 動的変動成分を付与した歌声合成音を作成した。その際の合成条件は、4.2 節に示したものと同様である。また、合成に使用した歌声は、歌声データセット A に収録されている 1 名の女性歌唱者による歌声（4 データ）である。以下、この方法によって作成した歌声合成音の種類を示す。

- **NORMAL**

TEMPO によって歌声データから抽出した基本周波数を用いて ,Klatt formant synthesizer によって合成したもの ( 図 4.1 における [a] の行程 )

- **SYN-ALL**

F0 制御モデルによってすべての F0 動的変動成分を制御した基本周波数変化を用いた合成音 ([b] の行程)

- **SYN-OS**

F0 制御モデルによってオーバーシュートのみ制御した合成音 ([b] の行程)

- **SYN-VB**

F0 制御モデルによってヴィブラート・微細変動成分のみ制御した合成音 ([b] の行程)

- **SYN-PRE**

F0 制御モデルによってプレパレーションのみ制御した合成音 ([b] の行程)

- **SYN-BASE**

F0 制御モデルの入力であるメロディ成分のみを用いた合成音 ([b] の行程)

尚 , F0 制御モデルの制御パラメータ値は , 前章の表 3.2 に示したものを使用した . 図 4.10 に , これら合成音の基本周波数変化の一例を示す .

### 聴取実験

前述の歌声合成音を聴覚刺激とし , 心理物理実験を行った . 実験はシェッフエの対比較法で行い , 前節と同じ 7 段階の評価尺度を用いることで歌声の自然性について評価した . また , 実験環境と実験手続きに関しては , 前節の実験と同じである .

実験で用いる聴覚刺激は , NORMAL , SYN-ALL , SYN-OS , SYN-PRE , SYN-VB , SYN-BASE を 2 つずつ対にしたものである . 刺激対の数は , 4 つある歌声データそれぞれについて , 刺激順序の違いも考慮した  $6 \times 5 = 30$  対である .

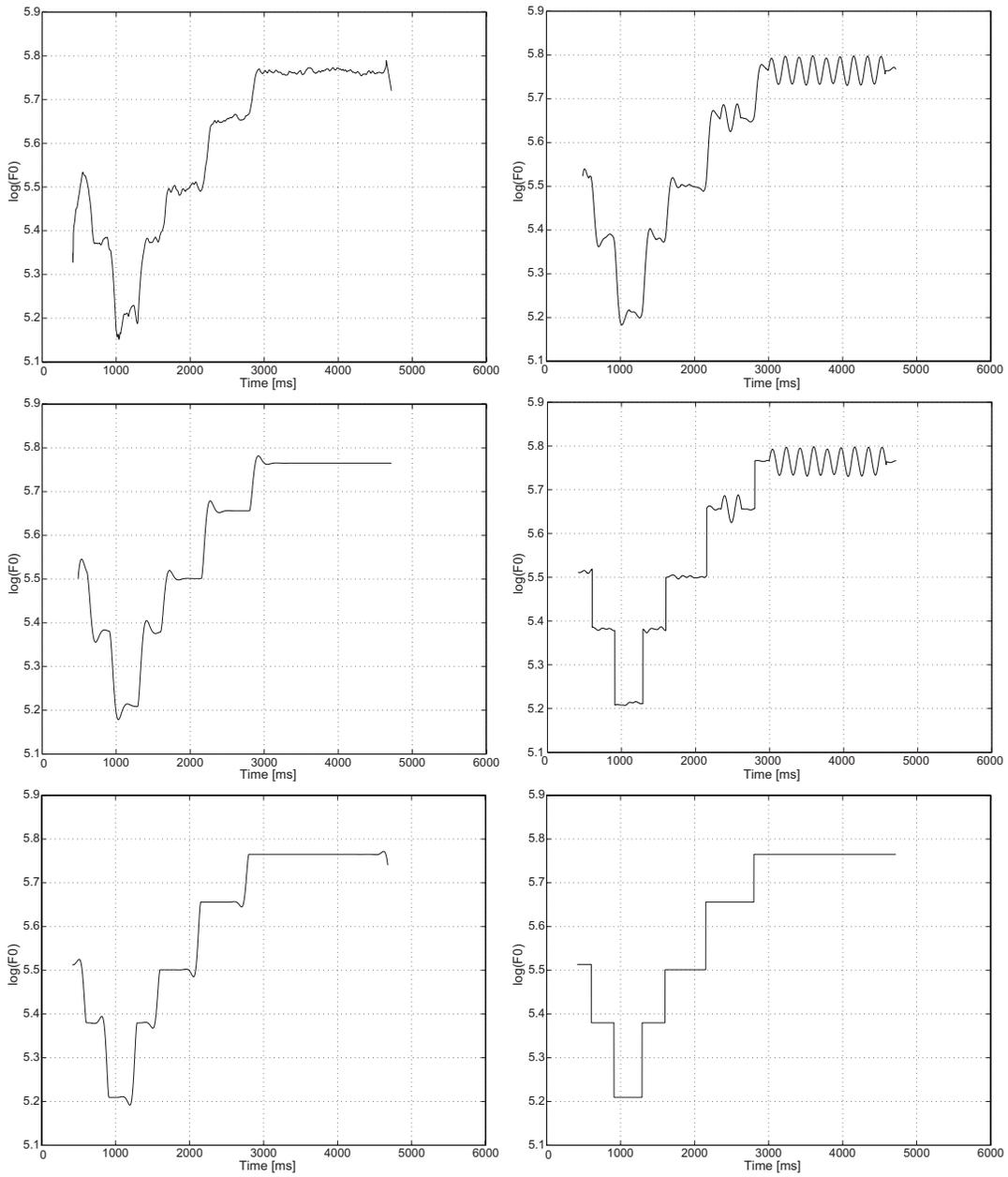


図 4.10: 歌声合成に用いた基本周波数変化の一例．上段の左から順に NORMAL, SYN-ALL. 中段 SYN-OS, SYN-VB. 下段 SYN-PRE, SYN-BASE の基本周波数変化を表す．

表 4.2: 母数の推定 (自然性) ( Klatt formant synthesizer による歌声合成の場合 ) .

歌声	母数
NORMAL	1.16
SYN-ALL	1.18
SYN-OS	0.56
SYN-VB	-0.73
SYN-PRE	-0.06
SYN-BASE	-1.15

刺激ペアはランダムに呈示し，被験者によってその順番は異なる．刺激呈示の条件は，図 4.6 と同様である．

実験に参加した被験者は，オーディオメータを用いた聴力検査で両耳とも健常な聴力を持つことが確認された大学院生 7 名（男性 6 名，女性 1 名）で，この内の 6 名は前節の実験と同じ者である．

#### 実験結果と考察

聴取実験で得られた評価値から，浦の変法によって処理した結果を表 4.2 に示す．また，図 4.11 に，表に示した母数の値に従って，6 つの刺激の心理距離関係を直線上に表したものを示す．尚，各刺激間の中の有意差検定を有意差水準 5 %で行った結果，いずれの刺激間にも有意な差が確認された．

図 4.11 から分かるように，メロディ成分のみの合成音 SYN-BASE に対し個々の F0 動的変動成分を制御することで，どの場合も歌声合成音の自然性が向上する結果となった．更に，すべての F0 動的変動成分の制御を行った合成音 SYN-ALL の自然性が，実音声の基本周波数変化を用いた合成音 NORMAL に比べ高い値を示す結果となった．以上の結果より，歌声を歌声として知覚する為には，メロディ変化だけでは十分でなく，その変化中に F0 動的変動成分の存在が必要であることが明らかとなった．これは，前節の実験結果を支持するものでもあり，これらの結果から，我々があらゆる歌声を歌声として知覚可能にしている要因の 1 つが F0

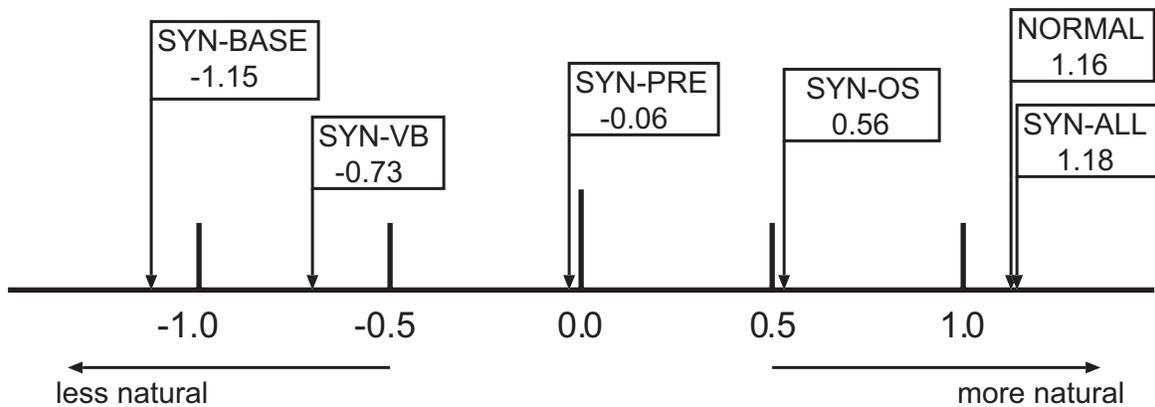


図 4.11: Klatt formant synthesizer による歌声合成音の自然性の関係。

動的変動成分であることが明らかとなった。

次に、SYN-ALL の自然性が NORMAL よりも高かった要因について考察する。SYN-ALL を含め、NORMAL と BASE 以外の基本周波数制御に用いた制御パラメータ値は、多くのプロ歌唱による歌声と少数のアマチュア歌唱の歌声を対象に抽出したものであった。その為、各 F0 動的変動成分の特性は、アマチュア歌唱の合成音 NORMAL の基本周波数に含まれる特性とは異なっている可能性が考えられる。その結果、各成分の特性の違いが、歌声合成音の自然性の差となって知覚され、その結果 SYN-ALL の自然性が最も高くなったと考えることができる。

次に、個々の F0 動的変動成分が歌声知覚に与える影響について考察する。

オーバーシュートの歌声知覚への影響が最も大きかったことは、前節の実験においてオーバーシュートを除去した時の自然性劣化が最も大きかったことと対応が取れる。しかし、前節の実験では、ポルタメント成分を残し、オーバーシュートエクステント成分のみを除去した場合の歌声知覚への影響を調べることは出来なかった。だが今回の実験では、2つの成分が制御されており、その影響の大きさは、他の成分を制御した場合と比べて顕著であった。これにより、間接的ではあるが、ポルタメント成分も歌声知覚に大きな影響を与えていることが確認された。

プレパレーションの影響は、オーバーシュートほど大きいものではない結果となった。しかし、前節の実験と同様に、ヴィブラートや微細変動とほぼ同程度の影響を与えることが明らかとなった。これにより、プレパレーションが、歌声知

覚において重要な役割を果たしている成分であることが再確認された。

ヴィブラートの影響は、前節の実験と同様、あまり大きくない結果となった。しかし、先に述べたような SYN-ALL が最も高い自然性を示した要因が歌唱技術の違いから生じた F0 動的変動成分の特性の差異であり、且つ前章で示した歌唱技術の違いによってヴィブラート特性が異なるという知見が正しいとすれば、SYN-ALL が最も自然な歌声として知覚されたことには、ヴィブラートの存在が大きく貢献していることが可能性として考えられる。しかし、この結果からだけでは、この問題に対して明確な解答を得ることができない。そこで、次節において、歌唱技術の差による各 F0 動的変動成分の特性の違い、及びその特性の違いが歌声知覚に与える影響を調査することで、SYN-ALL の高い自然性の要因を追求すると同時に、前章で課題として挙げた歌唱技術とヴィブラート特性の関係について検討を行う。

#### 4.4.2 STRAIGHT を用いた歌声合成システムによる検証

先の Klatt formant synthesizer を用いたシステムでは、スペクトルとパワーの情報を固定していた為、原音声の基本周波数を用いて合成された歌声の音質でさえ決して良いものでは無かった。そこで、STRAIGHT に F0 制御モデルを組み込んだ図 4.2 の歌声合成システムを用いることで、スペクトルやパワーの時間的な変化も考慮した高品質な歌声の合成を行う。この実験により、原音声との比較による各 F0 動的変動成分の歌声知覚に与える影響が調査可能となる。

##### 歌声合成音の作成

図 4.2 に示した歌声合成システムにより、各 F0 動的変動成分を付与した歌声合成音を作成した。その際の合成条件は、4.2 節に示したものと同様である。また、合成に使用した歌声は、前項の実験と同じものである。以下に作成した歌声合成音の種類を示す。

- NORMAL

STRAIGHT によって分析・再合成した合成音。SPIKES における群遅延操作により、その音質は原音声とほぼ同じである（図 4.2 の [a] の行程）

- SYN-ALL  
F0 制御モデルによってすべての F0 動的変動成分を制御した基本周波数を用いた合成音 ([b] の行程)
- SYN-OS  
F0 制御モデルによってオーバーシュート成分のみ制御した合成音 ([b] の行程)
- SYN-VB  
F0 制御モデルによってヴィブラート・微細変動成分のみ制御した合成音 ([b] の行程)
- SYN-PRE  
F0 制御モデルによってプレパレーションとポルタメントのみ制御した合成音 ([b] の行程)
- SYN-BASE  
メロディ成分のみ用いた合成音 ([b] の行程)

## 聴取実験

前述の歌声合成音を聴覚刺激とし、聴取実験による評価を行った。実験はシェッフェの対比較法で行い、実験条件は前節と同じである。

## 実験結果と考察

聴取実験で得られた評価値を、浦の変法によって処理した結果を表 4.3 に示し、図 4.12 に歌声刺激の心理距離関係を示す。尚、各刺激間の有意差検定を有意差水準 5% で行った結果、いずれの刺激間にも有意な差が確認された。

この実験においても、メロディ成分に対して各 F0 動的変動成分を付与することで、歌声合成音の自然性が向上する結果となった。また、各成分が歌声知覚に与える影響は、オーバーシュート、プレパレーション、ヴィブラート・微細変動の順に大きい結果となり、前項の実験結果と同じ傾向を示した。更に、すべての F0 動的変

表 4.3: 母数の推定 (自然性) (STRAIGHT による歌声合成の場合) .

歌声	母数
NORMAL	0.94
SYN-ALL	0.85
SYN-OS	0.42
SYN-VB	-0.71
SYN-PRE	-0.23
SYN-BASE	-1.04

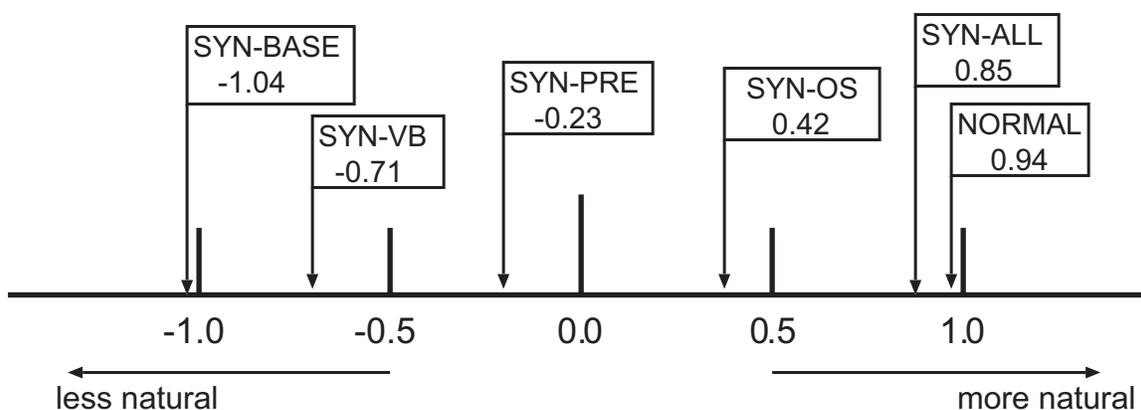


図 4.12: STRAIGHT による歌声合成音の自然性の関係 .

動成分を付与した SYN-ALL の自然性が，原音声 NORMAL と比較しても殆んど遜色ない程に高いものであった．これらの結果から，個々の F0 動的変動成分が歌声を知覚する上で重要な役割を担っていることを確認できた．また STRAIGHT を用いてスペクトルやパワーの特性も考慮した合成音が原音声と同程度の自然性を持つことが明らかとなり，F0 動的変動成分だけでなく，スペクトル等に含まれる音響的特徴が歌声知覚に影響を与えていることを間接的に示唆する結果となった．

## 4.5 歌唱技術と F0 動的変動成分の関係

前章では、4つの F0 動的変動成分が、多くの歌声において共通して含まれる歌声特有の音響的特徴であることが確認された。そして、4.3, 4.4 節においては、各成分が歌声を知覚する上で必要な音響的特徴であることも明確になった。しかしその一方で、4.3 節の Klatt formant synthesizer を用いた実験では、歌唱技術の違いによって F0 動的変動成分の特性が異なり、且つその特性の違いが歌声知覚に影響を与えている可能性を示唆する結果が得られた。

そこで本節では、歌唱技術の差によって生じる各 F0 動的変動成分の特性の違いを分析し、その特性の違いが歌声知覚に与える影響を調査する。これにより、前章で課題として挙げた歌唱技術の違いとヴィブラート特性の関係を明らかにし、他の成分においても同様の関係が存在するかについても検証を行う。

### 4.5.1 歌唱技術の違いによる F0 動的変動成分の特性の変化

歌唱技術の違いによってどれだけ F0 動的変動成分の特性が異なるかを調べるため、アマチュアとプロの歌唱者による歌声データを対象に、歌声 F0 制御モデルの制御パラメータ値を抽出した。

パラメータ値の抽出法は、第3章で示した最小二乗誤差に基づくフィッティング処理を採用した。また、フィッティング対象とした歌声データは、アマチュア歌唱に関しては歌声データセット A に収録されるすべての歌声、プロ歌唱に関しては歌声データセット B に収録されるすべての洋楽データと演歌、民謡、及びポピュラーである。抽出した制御パラメータ値を、表 4.4 に示し、制御パラメータ値を用いて F0 制御モデルによって記述された基本周波数変化の一例を図 4.13 に示す。

表 4.4 の制御パラメータ値で制御される各 F0 動的変動成分の特性について考察する。

ヴィブラート制御では、 $\Omega$  によって vibrato rate (振動の速さ) を、 $K$  によって vibrato extent (振動の大きさ) が制御される。表 4.4 の制御パラメータ値から、アマチュア歌唱で vibrato rate, extent がそれぞれ 4.8 Hz と約 2%、プロ歌唱で 6.2 Hz と 5.4% が平均的なヴィブラート特性である結果となった。これより、プロの歌唱におけるヴィブラートの方が速くて大きい振動で、その差が顕著であること

表 4.4: 歌声データセット A, B それぞれを対象に抽出した F0 制御モデルの制御パラメータ値 (歌声データセット A / 歌声データセット B の順) .

F0 fluctuation	$\Omega$ [rad/ms]	$\zeta$	$K$
overshoot	0.0312/0.0363	0.5183/0.5701	0.0312/0.0363
vibrato	0.0302/0.0390	—	0.0006/0.0021
preparation	0.0287/0.0308	0.6604/0.6746	0.0287/0.0308

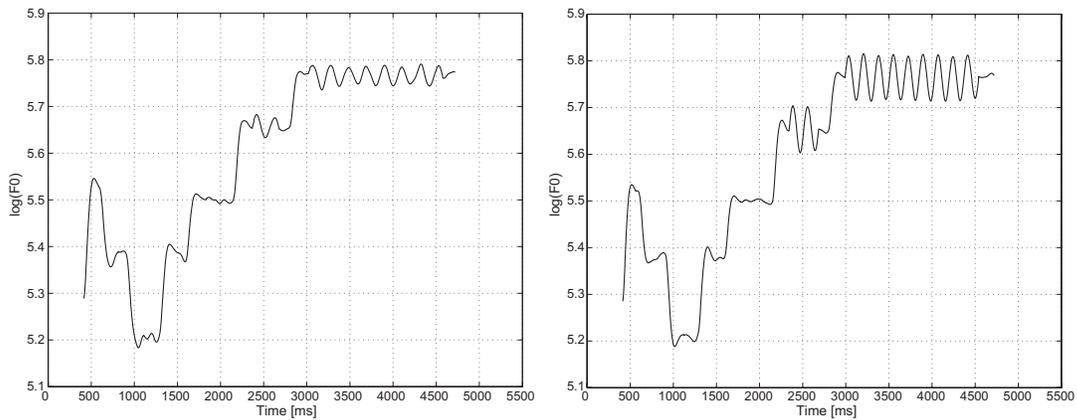


図 4.13: F0 制御モデルで記述された基本周波数変化 (左: アマチュア歌唱データから抽出した制御パラメータ値を用いた場合, 右: プロ歌唱データから抽出した制御パラメータ値を用いた場合) .

が明らかとなった . これにより , 前節で課題として挙げた歌唱技術の違いとヴィブラート特性の関係を明確にすることができた .

オーバーシュート制御では ,  $\Omega$  によって音高変化の速さ (ポルタメントの傾き) が ,  $\zeta$  によって目的音高値を越える過振動 (オーバーシュートエクステント) の大きさが , そして両者によって振動の持続時間が制御される . これより , プロ歌唱におけるオーバーシュートの方がアマチュア歌唱に比べ , ポルタメントの傾きは急峻 (変化が速い) で , オーバーシュートエクステントは小さく , そして変動の持続時間は短いことが明らかとなった . しかし , 各特性の差はどれも非常に小さく , オーバーシュート特性は歌唱技術の違いによって大きく変化しないことを示す結果となった .

プレパレーションの制御は、 $\zeta$ によって振動の大きさが、そして $\Omega$ と $\zeta$ によって振動の持続時間が制御される。これより、プロ歌唱におけるプレパレーションの方がアマチュアに比べ短時間に起こる小さい振動ではあるものの、その差は非常に小さいこと示す結果となった。これにより、プレパレーションの特性に関してオーバーシュートと同様に、歌唱技術の違いによって大きく変化しないことが明らかとなった。

#### 4.5.2 歌唱技術の違いによる F0 動的変動成分の特性の変化が歌声知覚に与える影響の調査

ここでは、前項で抽出した制御パラメータ値を用いた歌声合成を行い、聴取実験によって歌声合成音を評価する。これにより、歌唱技術の違いによる F0 動的変動成分の特性の差異が歌声知覚に影響を与えているか否かについて検証する。

##### 合成音の作成

歌声合成音は、4.2 節で提案した 2 種類のシステムそれぞれを用いて作成した。合成に使用した歌声は、歌声データセット A に収録されている 1 名の女性歌唱者による歌声 (4 データ) である。以下、この方法によって作成した歌声合成音を示す。

- REAL

TEMPO で抽出した原音声の基本周波数変化を用いた合成音 (STRAIGHT の場合は分析・再合成音) (図 4.1, 4.2 の [a] 行程)

- AMA

アマチュア歌手の歌声データ (歌声データセット A) から抽出した制御パラメータ値によって F0 制御モデルで制御した合成音 ([b] の行程)

- PRO-OS

AMA のオーバーシュート制御のみをプロの歌手の歌声データ (歌声データセット B) から抽出した制御パラメータ値によって制御した合成音 ([b] の

行程)

- PRO-VB

AMA のヴィブラート制御のみをプロの歌手の歌声データ (歌声データセット B) から抽出した制御パラメータ値によって制御した合成音 ([b] の行程)

- PRO-PRE

AMA のプレパレーション制御のみをプロの歌手の歌声データ (歌声データセット B) から抽出した制御パラメータ値によって制御した合成音 ([b] の行程)

- PRO-ALL

プロの歌手の歌声データから抽出した制御パラメータ値をすべての F0 動的変動成分制御に適用させた合成音 ([b] の行程)

制御パラメータ値は、表 4.4 に示した値を採用した。

## 聴取実験

前述の歌声合成音を聴覚刺激とし、シェッフエの対比較法による聴取実験によって自然性の評価を行った。その際の被験者は、正常な聴力を有する大学院生 8 名 (男性 6 名, 女性 2 名) である。それ以外の実験条件は、4.3 節の実験と同じである。

## 実験結果と考察

聴取実験で得られた評価値を、浦の変法によって処理した結果を表 4.5 に示し、図 4.14 に歌声刺激の心理距離関係を示す。尚、各刺激間の有意差検定を有意差水準 5 % で行った結果、Klatt formant synthesizer を用いた実験では、REAL-AMA, AMA-PRO-PRE, AMA-PRO-OS, PRO-PRE-PRO-OS の刺激対間に、STRAIGHT を用いた実験では、AMA-PRO-PRE, PRO-PRE-PRO-OS の刺激対間に有意な差は無く、それ以外の刺激間には有意差が確認された。

両実験において、アマチュア歌唱の F0 動的変動成分の特性を持つ合成音 AMA に対して、各成分をプロ歌唱の特性に変更することで歌声の自然性が向上する結果と

表 4.5: 母数の推定 (自然性) (Klatt formant synthesizer , STRAIGHT それぞれを用いた場合) .

歌声 刺激	母数	
	Klatt	STRAIGHT
REAL	-0.3239	0.1149
AMA	-0.2559	-0.1308
PRP-OS	-0.0052	-0.0156
PRO-VB	0.2348	0.0504
PRO-PRE	-0.0729	-0.0312
PRO-ALL	0.4375	0.1149

なった . また , プロ歌唱の F0 動的変動成分の特性を持った合成音 PRO-ALL の自然性は , Klatt formant synthesizer の実験では最も高い値を示した . 一方 STRAIGHT の実験では , PRO-ALL の自然性は原音声 NORMAL に比べわずかに低い値を示したものの , その差は非常に小さい結果となった . しかし , オーバーシュートとプレパレーション特性の変更は , ほとんど知覚できない程度に小さく , AMA から PRO-ALL への自然性向上は , ヴィブラート特性の変更が大きく寄与している結果となった .

以上の結果から , 歌唱技術の違いによるヴィブラート特性の差異が歌声自然性の違いとなって明確に現れることが明らかとなり , ヴィブラート特性が歌唱技術の違いを規定する要因の 1 つである可能性が高いことが確認された . 一方で , オーバーシュートやプレパレーションは , 歌唱技術の違いによる特性の変化は軽微であり , その変化も歌声知覚に殆んど影響を与えないことが明らかとなった .

## 4.6 F0 動的変動成分の特性と歌声知覚の関係

前節の実験では , 歌唱技術の異なる歌声データを対象に抽出した制御パラメータ値を用いた歌声合成を行うことで , 各 F0 動的変動成分の特性と歌声知覚の関係について検討を行った . しかし , 両者の関係をより詳細に検討するためには , 各

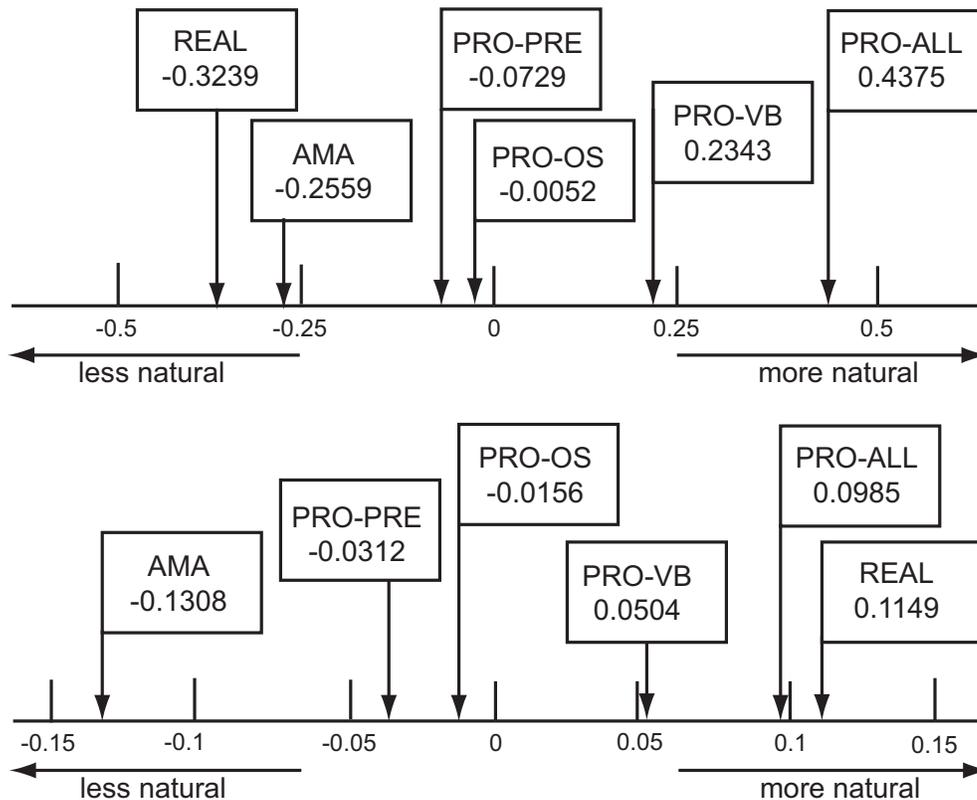


図 4.14: 歌声合成音の自然性の関係。上図：Klatt formant synthesizer を用いた場合（但し REAL-AMA , AMA-PRO-PRE , AMA-PRO-OS , PRO-PRE-PRO-OS の刺激対間に有意差無し）, 下図：STRAIGHT を用いた場合（但し AMA-PRO-PRE , PRO-PRE-PRO-OS の刺激対間に有意差無し）。

成分の特性の変動に対する歌声知覚の反応を詳細に調べる必要がある。

また、これまでの評価実験では、予め導出された制御パラメータ値を用いた場合の歌声合成音の自然性評価を行ってきた。しかし、これらの実験方法では、「フィッティング処理によって抽出された制御パラメータ値が歌声知覚的に最適なのか？」という問題に対して検証したことはない。

以上から、本節では、抽出された制御パラメータ値を変動させた場合の歌声合成音の自然性の変化を調査し、各成分の特性と歌声知覚の関係を明らかにする。

### 4.6.1 歌声合成音の作成

歌声合成音は，4.2 節で提案した Klatt formant synthesizer による手法で作成した．また，合成対象にした歌声は，歌声データセット A に収録された女性 1 名による歌声で，日本童謡「七つの子」の“からすなぜなくの”と，“こがあるからよ”の小節を日本語母音/a/のみで発声したものである．以下に，各 F0 動的変動成分の制御パラメータ値を変動させた歌声合成の作成手順について記す．

#### オーバーシュート，プレパレーションについて

オーバーシュートとプレパレーションの特性を変化させた歌声合成音は，制御パラメータ  $\Omega$  と  $\zeta$  を変動させることで作成した． $\Omega$  を変動させる場合は，前章の表 3.2 に示した制御パラメータ値（オーバーシュート：0.0383 rad/ms，プレパレーション：0.0292 rad/ms）を中心として，0.09 ~ 0.15 rad/ms（0.005 rad/ms 刻み）の幅で変化させた．尚，この際の  $\zeta$  は，表 3.2 に示した値（オーバーシュート：0.542 rad/ms，プレパレーション：0.668 rad/ms）で固定とした．また  $\zeta$  を変化させる場合は，先ほどの固定値を中心に 0.99 ~ 0.25（0.05 刻み）の幅で変化させ，その際の  $\Omega$  の値は， $\Omega$  を変動させる場合の中心値で固定した．尚，パラメータ  $K$  の値は， $\Omega$  と同様に变化させた．

#### ヴィブラートについて

ヴィブラート特性を変化させた歌声合成音は，制御パラメータ  $\Omega$  の調整で vibrato rate を， $K$  の調整で vibrato extent を変化させることで作成した．vibrato rate に関しては，前章の F0 制御モデルの検証時に行った分析で得られた平均値 5.6 Hz を中心に， $\pm 0.2$  Hz 刻みで  $\pm 7$  Hz（途中  $\pm 1$  Hz からは  $\pm 4$  Hz 刻み）まで独立に変化させたヴィブラートを付与した歌声合成音を作成した．尚，この際の vibrato extent は，ヴィブラート区間の音高値の 5.2 % の振動幅で固定した．また，vibrato extent に関しては，ヴィブラート区間の音高値の 5.2 % の振動値を中心とし， $\pm 1$  % 刻みで 0.5 ~ 15.0 %（途中 4 % からは 2 % 刻み）まで変化させたヴィブラートを合成音に付与した（vibrato rate は 5.6 Hz で固定）．

## 4.6.2 聴取実験

前述の歌声合成音を聴覚刺激とし、シェッフェの対比較法による聴取実験によって自然性の評価を行った。実験条件は、前節と同様である。

## 4.6.3 実験結果・考察

聴取実験で得られた評価値を浦の変法によって処理し、制御パラメータ値の変動に対する歌声合成音の自然性の変化を表したものを図4.15, 4.16に示す。図の横軸は各制御パラメータ値（ヴィブラートに関しては vibrato rate と vibrato extent）を示す。縦軸はそのパラメータ値を用いたの歌声合成音の自然性を表し、値が大きければその合成音が自然性の高い歌声として知覚されたことを表す。また、図中に示した AMA と PRO は、表4.4に示したアマチュア歌唱とプロ歌唱の歌声データセットそれぞれを対象に抽出したパラメータ値を示す。

すべての実験結果において、制御パラメータ値の変動に伴って歌声の自然性が変化する結果となった。これにより、各 F0 動的変動成分は、メロディ変化に存在さえすれば歌声の自然性向上に寄与できるのではなく、適切な特性を持つことでじめて歌声知覚に必要な成分になることが明らかとなった。

次に、ここで得られた F0 動的変動成分の特性と歌声知覚の関係を基に、歌唱技術と各成分の関係について考察する。

オーバーシュートとプレパレーションに関しては、図4.15のいずれの結果においても、PRO と AMA の違いによる歌声の自然性の差は非常に小さいものとなった。これは、前節の実験で、両成分の特性をアマチュア歌唱からプロ歌唱のものへ変更させても歌声の自然性があまり向上しなかった結果と対応がとれる。これらの結果より、オーバーシュートとプレパレーションは、あらゆる歌声に共通して含まれ、且つ歌声知覚に大きな影響を与える成分であるが、歌唱技術や歌唱法の違いを規定する要因では無いことが明らかとなった。

一方ヴィブラートに関しては、図4.16のどちらの結果においても、PRO と AMA の違いによる歌声の自然性の差が大きい結果となった。これは、前節の実験でヴィブラート特性の変更によって歌声の自然性が大きく向上した結果と対応が取れる。これらの結果より、ヴィブラートは、あらゆる歌声に共通して含まれる一方で、歌

唱技術の違いによって大きく特性が異なり，且つその特性の違いが歌唱技術の違いを規定する要因の1つであることが明らかとなった．また，これにより，前章において F0 制御モデルで記述されたヴィブラートと原音声に含まれるヴィブラートの二乗誤差が大きかったことが，歌唱技術の違いによってヴィブラート特性が大きく異なっていたために生じていた可能性が強いことが明らかとなった．

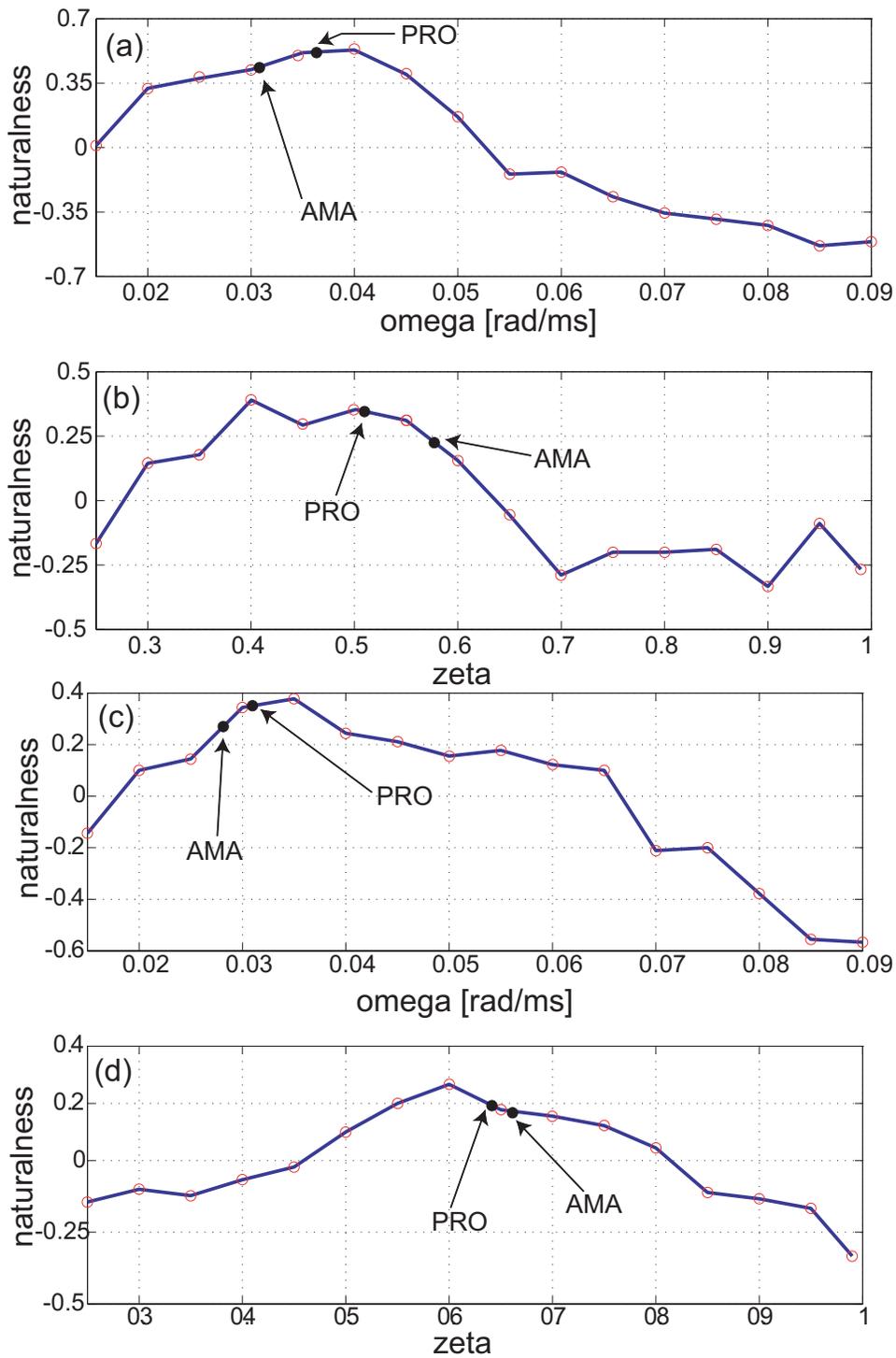


図 4.15: 制御パラメータ値変動と歌声の自然性の関係 ( (a):オーバーシュートの制御パラメータ  $\Omega$  を変動させた場合, (b):同じく  $\zeta$  を変動させた場合, (c):プレパレーションの  $\Omega$  を変動させた場合, (d):同じく  $\zeta$  を変動させた場合) .

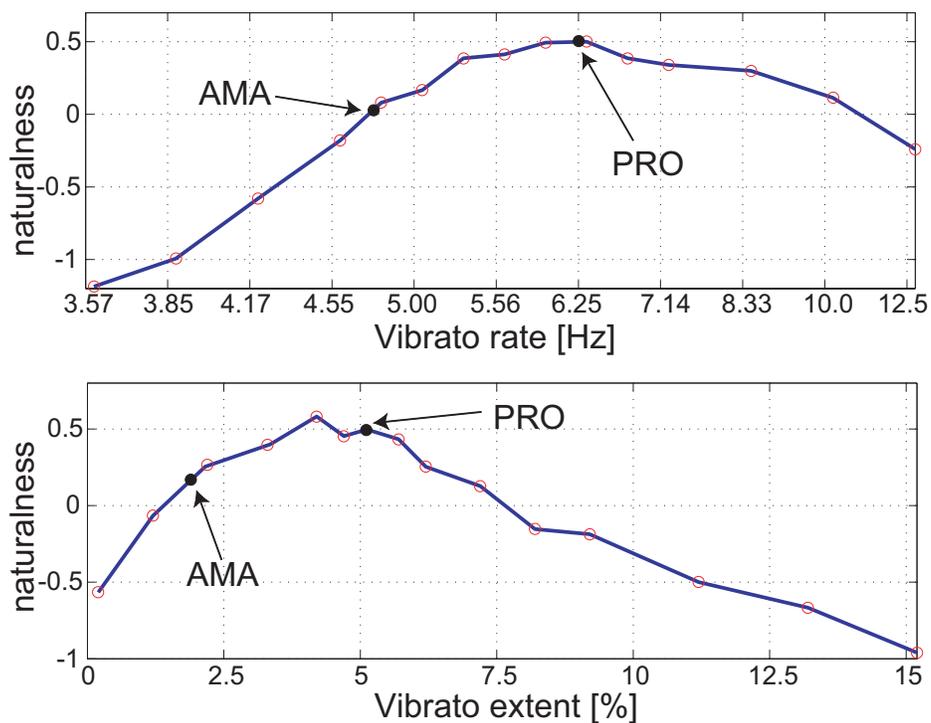


図 4.16: ヴィブラート特性の変動と歌声の自然性の関係（上：vibrato rate を変動させた場合，下：vibrato extent を変動させた場合）。

## 4.7 Vibrato rate の時間変動と歌声知覚の関係の調査

ヴィブラートは、4～7 Hz 程度の vibrato rate が一般的な特性とされている [71]。その為、歌声 F0 制御モデルにおいても vibrato rate が 5.6 Hz で定常な振動としてヴィブラートを記述してきた。しかし、前章 4.5.2 節で記したように、Prame の報告によって、vibrato rate が時間的に上昇する特性が示されており、その特性が歌声知覚に影響を与えている可能性が示唆されている [62]。

そこで、本節では、vibrato rate の時間特性に関する分析を行う。そして、その特性が歌声知覚に与える影響を調査する。これにより、これまで明確にされて来なかった vibrato rate の時間特性について検証する。

### 4.7.1 Vibrato rate の分析

Prame は、オペラ歌唱を対象に行ったヴィブラート分析の結果、その特性として以下の2つを報告している [62] .

a: Vibrato rate が時間と共に上昇する傾向が強い .

b: ヴィブラート区間終盤の 4~6 サイクルにおいて vibrato rate が顕著に上昇し、その変化率は 15 % 程度 .

そこで、プロのオペラ歌唱だけでなく、アマチュア歌唱や邦楽的歌唱データも対象に分析を行う事で、歌声一般の vibrato rate の時間変化特性を調査した。分析対象にした歌声データは、歌声データセット A に収録されたすべての歌声、及びデータセット B に収録されたすべての洋楽と長唄・民謡の歌声である。TEMPO によって各データの基本周波数を抽出し、そこに含まれる 413 区間のヴィブラートの瞬時周波数の時間変化を分析することで、vibrato rate の時間変動特性の調査を行った。

分析の結果、vibrato rate の時間変化パターンは、図 4.17 に示した 3 つの Type (Type1: 上昇, Type2: 下降, Type3: その他) に大別できることが分かり、Type1 から順に歌声データ全体の 54, 23, 23 % を占める結果となった。これにより、Prame の知見 a の信憑性を確認できたと同時に、vibrato rate の上昇特性がオペラ歌唱だけでなく邦楽歌唱やアマチュア歌唱においても共通して観測されることが明らかとなった。また、その上昇率に関しては、プロ歌唱の場合で約 13.5 % であり、やはり Prame の知見 b に記した特性に近いことが確認された。しかしその一方で、アマチュア歌唱における上昇率は、3.8 % とプロ歌唱に比べ小さい値となった。

また、邦楽歌唱のヴィブラートでは Type2 の特性も多く観測され、演歌を除いた邦楽データ全体では、38 % が vibrato rate が下降する特性を示した。これらの傾向は、洋楽的歌唱法のみを対象とした Prame の実験からは観測出来なかった邦楽独特のヴィブラート特性の存在を示唆する結果となった。

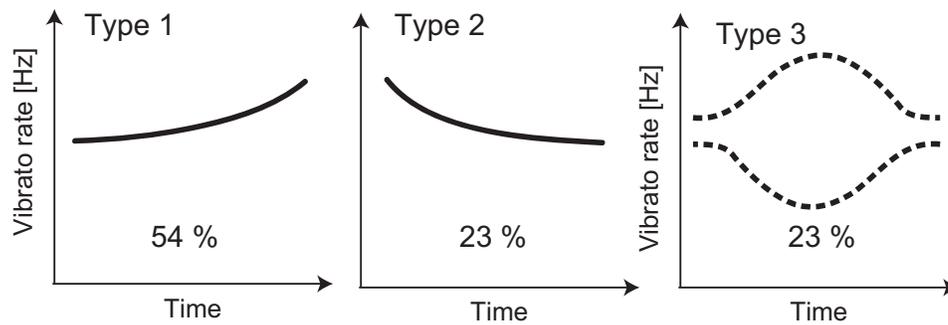


図 4.17: 3 種類の vibrato rate 時間変化パターン .

#### 4.7.2 Vibrato rate の時間変動が歌声知覚に与える影響の調査

先の分析で明らかとなった vibrato rate の時間的上昇及び下降する特性が歌声知覚に与える影響を調査するために、歌声 F0 制御モデルにおけるヴィブラート制御に改良を加え、vibrato rate が時間的に変化するヴィブラートを制御した歌声合成音を作成した .

##### 歌声合成音の作成

前述の通り、歌声 F0 制御モデルでは、ヴィブラートは制動 2 次系の定常振動モデル、

$$h(t) = (K/\Omega) \sin(\Omega t) , \quad (4.1)$$

で表現され、vibrato rate は時間的に一定としている . そこで、上式を以下の周波数変調の式に変更することで、vibrato rate の時間的な変化を制御可能にした .

$$h(t) = (K/\Omega) \sin\{\Omega t + \exp(M_f t)\} , \quad (4.2)$$

ここで、 $\Omega$  はヴィブラート区間における vibrato rate の初期値である . また、ヴィブラート区間終盤での vibrato rate の急激な上昇・下降を指数関数で表現し、パラメータ  $M_f$  によって vibrato rate の時間変化率を制御することで任意のヴィブラートを生成した . 尚、この際のヴィブラート振動の初期位相も 0 としている .

式 (4.2) を用いて、時間方向に伸長した日本語母音/a/(テノール歌手の話声を時間伸長したもの)と、日本童謡「七つの子」の歌声データに対して、様々な vibrato

rate の時間変化パターンを持つヴィブラートを制御した合成音を作成した。歌声合成の際、vibrato rate は、5.6 Hz を中心に変動幅を  $\pm 0.05$  Hz の刻みで広げていき（変動幅 1 Hz 以上では  $\pm 0.1$  Hz の刻み）、最大で 4 ~ 7 Hz の範囲で上昇・下降する特性を与えた。尚、初期値 5.6 Hz は、第 3 章の表 3.2 に示した制御パラメータ値で制御される歌声データ A, B から抽出された平均的な vibrato rate である。

## 聴取実験

作成した歌声合成音を聴取刺激とし、絶対評価実験によって歌声の自然性（5 段階尺度：不自然 ~ 非常に自然）の評価を行った。被験者は正常な聴力を有した大学院生 8 名（男性 6 名，女性 2 名）で、実験条件はこれまでのものと同様である。

## 実験結果・考察

図 4.18 に実験結果を示す。図の横軸は vibrato rate の変化率を、縦軸は歌声合成音の自然性をそれぞれ示す。

どちらの刺激を用いた場合も、vibrato rate の時間的な上昇を制御したヴィブラートの方が、5.6 Hz の定常振動に比べて、歌声の自然性が上昇する傾向が見られた。また、最も高い自然性を示した時の vibrato rate の変動幅が、歌声刺激では 15.3 % となり、分析結果で得られたプロ歌唱における平均的な特性値、及び Prame が報告している値と非常に近い値であった。一方で、アマチュア歌唱における上昇率 3.8 % を制御した場合には、自然性はほとんど上昇しない結果となった。vibrato rate を減少させた場合は、どちらの音声刺激の場合でも、上昇させた場合に比べ、大きな自然性の上昇は見られなかった。

以上から、一般的なヴィブラート特性として、vibrato rate が時間的に上昇し、且つその特性が歌声の自然性を向上させる要因であることが明らかになった。この結果は、Prame の知見を支持すると共に、これまで明確にされて来なかった vibrato rate の時間変動と歌声知覚の関係を定量的に示すものである。また、vibrato rate の上昇率がプロとアマチュア歌唱の間で大きな差があり、プロ歌唱の特性を制御した場合の歌声合成音の自然性が非常に高かったことから、vibrato rate の時間特性が、多くの歌声において共通のヴィブラート特性であると同時に、歌唱技術の

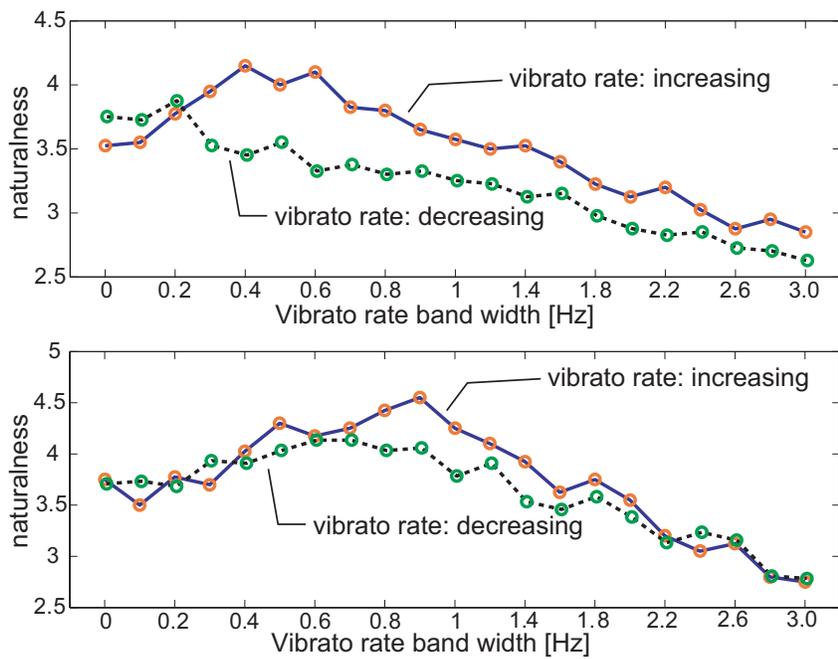


図 4.18: Vibrato rate の時間変動と歌声の自然性の関係 (上: 伸長母音刺激の場合, 下: 歌唱刺激の場合),

違いを規定する要因である可能性も示唆する結果となった。本節及び前節までの実験結果から、ヴィブラートは歌声共通の F0 動的変動成分であると同時に、歌声がより良い歌声として知覚される為に重要な役割を果していることが明らかとなり、顕著な vibrato extent, 更には vibrato rate の時間上昇特性が、卓越した歌唱技術によって制御されている可能性が高いことが示された。

## 4.8 音高遷移方向の違いによるオーバーシュート特性の差異に関する検証

de Krom らは、オーバーシュートエクステントが音高上昇・下降時どちらにおいても存在することを示している [31]。更には、「プロの歌唱者は定められた音高遷移を誇張に表現する」として、オーバーシュートエクステントが歌唱者の意図的な制御によって生成される成分であると報告している。その為、本論文では、この知見を基にして、これまでのすべての実験において、音高上昇・下降時を区別

せず同じ制御パラメータ値によってオーバーシュートを記述してきた。

一方で Mori らは、音高上昇時にはオーバーシュートエクステントは殆んど存在せず、その存在は音高下降時に特化したものであると報告している [54]。更に、これら特性の違いから、オーバーシュートは歌唱者の意図的制御ではなく、急峻な基本周波数変化の制御に伴った物理現象の可能性が高いとし、de Krom らの報告とは異なる見解を示している [55]。また、音高遷移方向の違いによるオーバーシュート特性の差異に関しては、Sundberg によって、音高下降時の方が上昇時に比べ基本周波数遷移の速度が速い（ポルタメントの傾きが急峻）ことが報告されている。

この様に、音高遷移方向の違いとオーバーシュート特性の差異に関しては、議論が分かれた状況にある。そこで、歌声  $F_0$  制御モデルを用いることで、様々な歌声を対象に音高上昇・下降時におけるオーバーシュート特性の差異を調査する。そして、その特性の違いが歌声知覚に与える影響を調査する。これらの取り組みを通じて、これまで明確にされてこなかったオーバーシュートの特性について検証すると共に、オーバーシュートエクステントの存在が歌唱者の意図的制御によるものか否かを検証する。

#### 4.8.1 音高遷移方向の違いによるオーバーシュート特性の差異に関する検証

音高上昇・下降時それぞれにおいて、第3章で記した最小二乗誤差に基づくフィッティング処理 [64] によってオーバーシュートの制御パラメータ値  $\Omega$ ,  $\zeta$  を求め、両者の特性の違いについて調査した。また、プレパレーションに関しても同様の調査を行った。対象とした歌声データは、歌声データセット A に収録されたすべての歌声と、データセット B に収録されたすべての洋楽歌唱である。抽出された制御パラメータ値を表 4.6 に示す。

オーバーシュートに関しては、 $\Omega$  の値が上昇・下降時でわずかな差があり、 $\zeta$  の値は大きく異なることが分かった。この結果から、音高変化の速度（ポルタメント傾き）は、音高下降時の方が速いことが明らかとなり、Sundberg の報告と一致する結果となった。また、オーバーシュートエクステントも音高下降直後の方が大きい事が明らかになった。しかしこの特性は、音高上昇時では過渡振動がほとんど

表 4.6: F0 制御モデルにおけるオーバーシュートとプレパレーションの制御パラメータ値（音高遷移方向の違いを考慮した場合）。

音高変化	オーバーシュート		プレパレーション	
	$\Omega$ [rad/ms]	$\zeta$	$\Omega$ [rad/ms]	$\zeta$
上昇	0.0339	0.704	0.0287	0.553
下降	0.0401	0.461	0.0291	0.577

存在しないとする Mori らの報告とは異なり，上昇においても過渡振動が存在する結果となった．更に，これらの特性は，歌声データセット A, B どちらにおいても共通して観測された．以上から，これらオーバーシュートの特性が歌唱技術や歌唱法の違いに関係なく，歌声共通の特性であることが明らかとなり，de Krom の知見とは異なる結果が得られた．

また，プレパレーションに関しては，どちらのパラメータ値においても大きな違いは無く，音高遷移方向に関わらずほぼ同じ特性であることを示す結果となった．

#### 4.8.2 オーバーシュート特性の差異が歌声知覚に与える影響

先の分析で明らかとなった音高上昇と下降時におけるオーバーシュート特性の差異が歌声知覚に与える影響を聴取実験によって調査した．

##### 歌声合成音の作成

聴覚刺激として用いた歌声合成音は，図 4.1 に示した F0 制御モデルと Klatt formant synthesizer を用いた手法（図中の [b] の過程）で作成した．合成に使用したメロディは，日本童謡「七つ子」である．

- **BASE**

メロディ成分のみを用いた合成音．

- **REF1**

音高上昇・下降時の違いを考慮せず，同一の制御パラメータ値（表 3.2）に

よってメロディ成分にオーバーシュートのみ付与した合成音。

- REF2

Moriらの研究を参考に、音高下降時のみ目的音高を越える振動を制御した合成音（その際の制御パラメータ値は、上昇時が $\Omega : 0.0348 \text{ rad/ms}$  ,  $\zeta : 0.99$  , 下降時は表 3.2 の値を使用）。

- OPT

表 4.6 に示す制御パラメータ値を用いて、音高上昇・下降時それぞれにおいてオーバーシュートのみを制御した合成音。

## 聴取実験

上記の歌声合成音をシェッフェの対比較法で評価した。実験方法はこれまでと同様であり、被験者は正常な聴力を有した大学院生 6 名（男性 4 名，女性 2 名）である。

## 実験結果・考察

実験結果を浦の変法によって処理した結果を図 4.19 に示す。尚，F 検定により，どの刺激間においても有意差水準 5 % で有意な差があることを確認している。

図 4.19 から明らかのように，今回新たに抽出した制御パラメータ値を用いて，音高上昇と下降時において異なる特性のオーバーシュートを付与した合成音の自然性が最も高い結果となった。また，Moriらの報告による特性を付与した合成音 REF2 は，BASE の次に低い自然性を示したものの，REF1 の自然性との差は非常に小さい結果となった。以上から，音高上昇時においてもオーバーシュートエクステントの存在が歌声知覚に重要であることが確認された。一方で，その特性は音高下降時とは大きく異なり，且つ上昇・下降それぞれにおける特性の差異が歌唱技術の違いに関係なくあらゆる歌声において共通して観測された。以上のことから，オーバーシュートは歌唱者の意図によるものでなく，メロディに伴った急峻な基本周波数の上昇・下降を制御する上で生じる物理現象である可能性が高いことが考えられる。そして，オーバーシュートエクステントが物理現象である

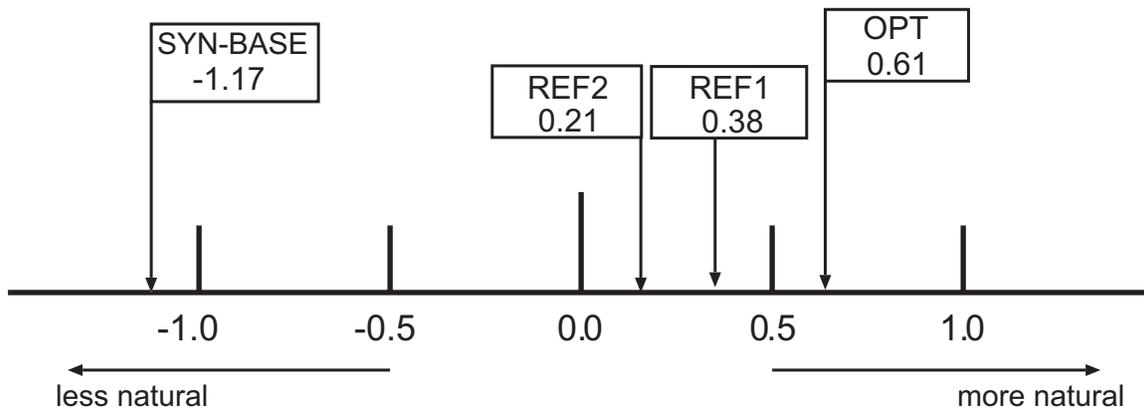


図 4.19: 歌声合成音の自然性の関係。

が為に、本実験における REF2 や、4.3 節の NO-OS の歌声合成音のようなオーバーシュートエクステン트가無い歌声が、人間の歌声として不自然に知覚されたと考えられる。

## 4.9 まとめ

本章では、歌声 F0 制御モデルを用いた歌声合成システムを提案することで、F0 動的変動成分と歌声知覚との関係を詳細の調査した。

はじめに、歌声 F0 制御モデルを用いた歌声合成システムを提案した。ここでは、F0 制御モデルを Klatt formant synthesizer に適用したシステムと、STRAIGHT に適用したシステムの 2 種類を提案した。

次に、F0 動的変動成分の歌声知覚に与える影響について調査した。ここでは、原音声の基本周波数変化から各 F0 動的変動成分を除去した歌声合成音を作成し、心理物理実験によって評価を行った。その結果、各 F0 動的変動成分を除去することで、歌声の自然性が劣化することが確認された。中でも、オーバーシュートを除去した場合の自然性劣化が顕著であった。

次に、最初に提案した 2 種類の歌声合成システムを用いてメロディ変化に個々の F0 動的変動成分を付与した歌声合成音を作成し、心理物理実験によって評価した。その結果、どちらの合成システムを用いた場合も、各 F0 動的変動成分をメロディ変化に付与することで、歌声の自然性が向上することが確認された。中でも、オー

バーシュートを付与した場合の自然性向上が顕著であった。また、STRAIGHTを用いた実験では、F0 制御モデルによってすべての F0 動的変動成分を制御することで、原音声に遜色無い極めて自然性の高い歌声合成が可能なが確認された。これらの実験により、各 F0 動的変動成分の歌声知覚における重要性を定量的に示すことができた。

次に、歌唱技術と F0 動的変動成分の特性の関係について調査した。ここでは、アマチュアとプロのそれぞれ歌声データを対象に抽出した制御パラメータ値から、歌唱技術の違いによる各成分の特性の差異について調査した。そして、各歌声データから抽出した制御パラメータ値を用いた歌声合成を行い、心理物理実験によって評価した。この結果、オーバーシュートとプレパレーションは、歌唱技術の違いで特性が僅かに異なるものの、その違いは歌声の自然性に影響を与えないほど小さいものであった。また、ヴィブラートに関しては、歌唱技術の違いによって大きく特性が異なり、その特性の違いが歌声知覚に大きな影響を与えていることが明らかとなった。

次に、個々の F0 動的変動成分の特性を変化させた場合の歌声知覚の反応を調査した。ここでは、オーバーシュート、プレパレーション、ヴィブラートの制御パラメータ値を様々に変化させた歌声合成音を作成し、心理物理実験によって評価した。この結果、いずれの成分の特性の変化に対しても、歌声の自然性は大きく変化することが明らかとなった。また、この結果と、前節の歌唱技術の差と F0 動的変動成分の特性の差異に関する実験結果から、ヴィブラートは歌声共通の音響的特徴であると同時に、歌唱技術の違いを規定する大きな要因であることが確認された。

次に、vibrato rate の時間変動特性と歌声知覚の関係について調査した。ここでは、vibrato rate の時間変動特性に関する先行研究の知見を基に、様々な歌声を対象に vibrato rate の時間変動特性を分析し、vibrato rate の時間的変動に対する歌声知覚の反応を調査した。この結果、特にプロ歌唱において vibrato rate が時間的に上昇する傾向が強いことが明らかとなり、且つこの特性を制御することで、定常振動のヴィブラートに比べ歌声合成音の自然性が向上することが明らかとなった。更には、歌声に自然性を与える vibrato rate の時間変化特性は、卓越した歌唱技術を持つプロ歌唱においてのみ観測されることが明らかとなった。これにより、

vibrato rate の特性も，歌唱技術の違いを規定する要因の 1 つであることが明らかとなった．

最後に，先行研究においてオーバーシュート特性が音高上昇時と下降時で特性が異なるか否かで議論が分かれている問題を検証した．ここでは，様々な歌声を対象に音高上昇・下降時それぞれのオーバーシュート特性を調査し，それら特性の違いが歌声知覚に与える影響を調査した．この結果，音高遷移方向が異なることで，オーバーシュートの特性に明確な違いがあることが明らかとなり，且つその特性を制御することで，歌声合成音の自然性が向上することが明らかとなった．その中でも，オーバーシュートエクステントは，歌唱時特有の急峻な基本周波数遷移に伴って必然的に生成される物理現象である可能性が高いことが確認された．更に，プレパレーションの特性が音高上昇直前，下降直前それぞれにおいて，ほとんど違いが無いことも明らかとなった．これら一連の結果から，オーバーシュートとプレパレーションは，歌唱技術や歌唱法の違いに関係なく様々な歌声に共通して存在し，その特性も大きな違いがないことから，あらゆる歌声を知覚する上で基本的な役割を果たしている成分である可能性が高いことが確認された．

## 第 5 章

# 歌声特有のスペクトル特性と歌声知覚 の関係の調査

## 5.1 はじめに

本章では、前章で提案した歌声 F0 制御モデルを STRAIGHT に適用した歌声合成システムを用いて、歌声特有のスペクトル特性と歌声知覚の関係を調査する。これにより、歌声のスペクトルに含まれる歌声知覚に重要な音響的特徴を明らかにする。

はじめに、singers' formant と歌声知覚の関係を調査する。ここでは、前章で提案した歌声合成システムの処理過程において singers' formant を制御する方法を提案する。そして、singers' formant を操作した歌声合成を行い、心理物理実験を行うことで、singers' formant と歌声知覚の関係を定量的に明らかにする。

次に、ヴィブラートに同期したホルマントの振幅変調と歌声知覚の関係を調査する。ここでは、前章で提案した歌声合成システムの処理過程においてヴィブラートに同期したホルマントの振幅変調を制御する方法を提案する。そして、ホルマントの振幅変調を制御した歌声合成を行い、心理物理実験を行うことで、ヴィブラートに同期したホルマント振幅変調と歌声知覚の関係の調査を行う。

## 5.2 Singers' formant が歌声知覚に与える影響の調査

Sundberg は、男性オペラ歌唱における singers' formant の特性として、母音スペクトルの 2.8 kHz 付近の成分が話声に比べ平均で 20 dB 高いことを示している [81]。また他の研究 [27, 75] においても、Sundberg の知見を支持する結果が報告されている。また、中山らは、邦楽歌唱におけるスペクトル包絡の 3 ~ 4 kHz 付近にも同様のピークがあることを示している [59]。更に辻らはこれらの知見を踏まえて、オペラ歌唱と邦楽歌唱（長唄、民謡等）を対象に singers' formant の分析を行うことで、図 5.1 に示すようにテノール歌唱における 2.5 kHz 付近、及び民謡歌唱における 4 kHz 付近に singers' formant が存在することを示し、これまでの知見を支持している [14]。

歌声知覚における singers' formant の役割についての検証も、先行研究において行われている。Vennard は、singers' formant が存在する約 2.8 kHz 付近の周波数成分が、歌声の「響き」の知覚に影響を与えていると報告している [89, 90]。また、

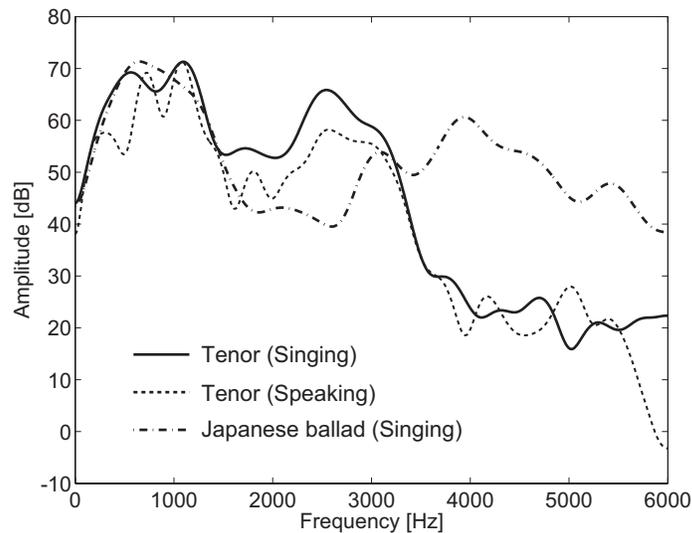


図 5.1: 洋楽・邦楽歌唱における singers' formant[14].

過らも同様の結果を示すと共に，話声の 3 kHz 付近のスペクトル包絡を 12 dB 持ち上げることによって，話声が歌声らしく知覚されることを示している [14]．更に Sundberg は，singers' formant の存在によって，オペラ歌唱がオーケストラ演奏にマスキングされず明瞭に聞こえる事を示している [77, 78, 79] ．

以上の知見から，singers' formant は，主にプロ歌唱において歌唱法の種類に関係なく存在し，更には歌声を知覚する上で重要な役割を果たしている音響的特徴であることは間違いないと考えられる．しかし，singers' formant と歌声知覚の関係を定量的に調査した例は無い．また，singers' formant の特性と歌声知覚の関係も明らかになっていない．そこで本節では，singers' formant を制御可能な歌声合成システムを構築することで，上で述べた 2 つの関係について検証する．

### 5.2.1 Singers' formant の制御法

前節で提案した STRAIGHT を用いた歌声合成システムにおいて singers' formant を扱う為には，スペクトル包絡と非周期成分割合の音響パラメータに対する制御法を構築する必要がある．非周期成分割合とは，音声の非周期成分の比率を表すパラメータであり，非周期成分に対応するスペクトルの下側包絡と周期成分に対応する上側包絡の差分によって記述される．その為，スペクトル包絡のある区間

を強調させることで singers' formant の制御を行った場合，同じ区間の非周期成分割合を弱める制御が必要となり，これらの先行研究においても非周期成分割合制御の必要性が報告されている．そこで，図 4.2 に示した歌声合成システムにおいて，STRAIGHT-core で抽出されたスペクトル包絡への singers' formant の付与と，それに連動した非周期成分割合の制御法を提案する．制御法の概要を図 5.2 に示す．

Singers' formant は，各時刻の対数振幅スペクトル包絡  $Sp_t(f)$  に対して次式で記述される hanning 窓を用いた関数  $W_{sf}(f)$  を掛けることで制御される．

$$W_{sf}(f) = \begin{cases} (1 + A_{sf}) \left( 1 - \cos \left( 2\pi \frac{f}{2000+1} \right) \right), & (F_{sf} - 1000 \leq f \leq F_{sf} + 1000), \\ 1, & otherwise, \end{cases} \quad (5.1)$$

ここで， $F_{sf}$  は入力音声の対数振幅スペクトル包絡  $Sp_t(f)$  の 3 kHz 付近に存在するホルマント周波数値で，ここを中心として 2000 Hz の窓幅でスペクトル包絡を持ち上げる．また，持ち上げる割合は， $A_{sf}$  によって調整される．非周期成分割合の制御に関しても，各時刻における非周期成分割合  $Ap_t(f)$  に対して  $W_{sf}(f)$  を掛けることで，スペクトルを強調させた区間を弱める加工を施す．

### 5.2.2 Singers' formant が歌声知覚に与える影響

本節の冒頭で述べたように，singers' formant の存在が歌声の「響き」や，オーケストラ演奏における「聞こえやすさ」に影響を与えていることは，これまでの研究で示されている．しかし，これだけの議論では，歌声知覚における singers' formant の重要性を明確にしたことにはならない．また，singers' formant の存在が，話声と歌声の違いを規定する要因となっているのかについても明らかになっていない．そこで，話声に対して singers' formant を付与することで，自然な歌声として知覚されるか調査した．

#### 歌声合成音の作成

歌声合成音は，図 5.2 に示した方法でスペクトル系列と非周期成分割合に対して singers' formant を付与する加工を加え，それらを用いて STRAIGHT によって再

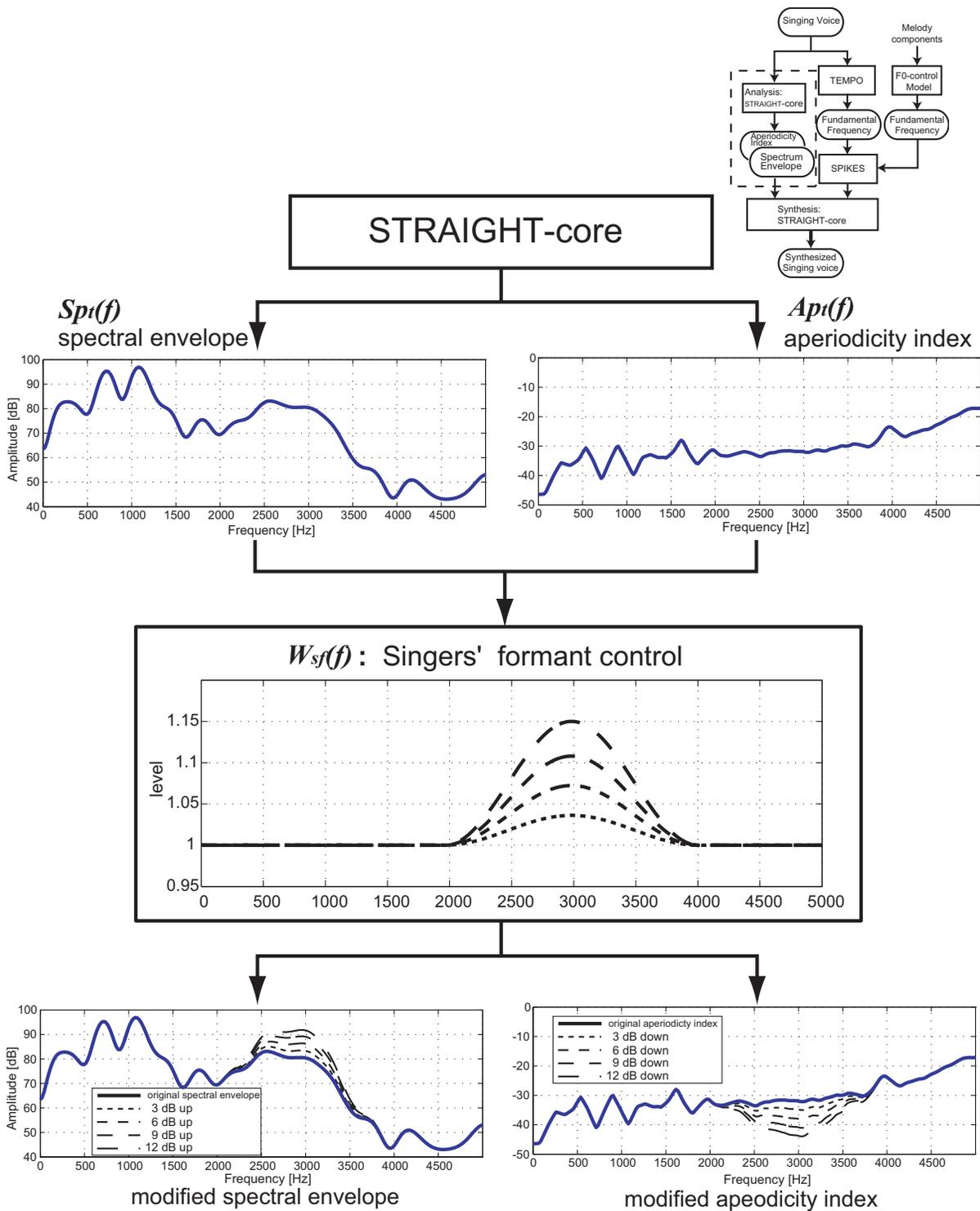


図 5.2: Singers' formant の制御法 .

合成することで作成した．以下に作成した合成音の種類を示す．

- **BASE**

歌声データセット B に収録されるテノール歌手の単独発話母音 /a/ ( 時間長 : 0.9 ms ) を STRAIGHT によって時間長 1.6 ms まで伸長した合成音 .

- **VIB**

BASE の基本周波数に対して , F0 制御モデルによってヴィブラートを制御した合成音 ( 制御パラメータ値は表 3.2 の値を使用 )

- **SIG-FM**

BASE のスペクトル包絡の 3 kHz 付近に存在するホルマント周波数を中心に , 帯域幅 2000 Hz で最大 12 dB 上昇させ , 同じ帯域の非周期成分割合に対して最大で 12 dB 減少させることで , singers' formant を付与した合成音 .

- **SF-VIB**

BASE に対してヴィブラートと singers' formant を付与した合成音 .

尚 , F0 動的変動成分のヴィブラートのみを制御した合成音 VIB , 及び両者を制御した合成音 SF-VIB は , singers' formant が歌声知覚に与える影響の比較対象として作成した .

## 聴取実験

前述の歌声合成音を聴覚刺激とし , 心理物理実験による評価を行った . 実験はシェッフェの対比較法で行い , 図 5.3 に示す 5 段階の評価尺度を用いて歌声の自然性について評価を行った . 被験者は正常な聴力を有した大学院生 6 名 ( 男性 4 名 , 女性 2 名 ) で , 実験環境と実験手続きに関しては , 4.3 節で行った実験と同じである .

## 実験結果・考察

上記の実験方法で得られたデータを , 浦の変法によって処理した結果を表 5.1 に示す . また , 表に示した母数の値に従って , 5 つの刺激の距離関係を直線上で示

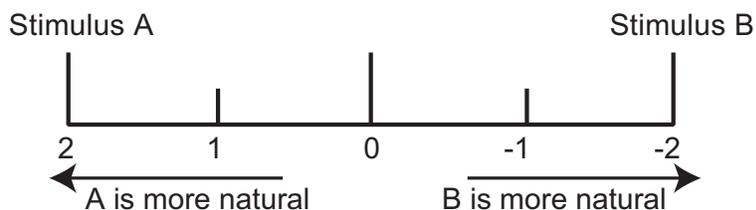


図 5.3: シェッフエの一対比較実験で用いた歌声の自然性に関する 5 段階評価尺度 .

表 5.1: 母数の推定 (自然性) .

歌声	母数
BASE	-0.73
VIB	0.42
SIG-FM	-0.46
SF-VIB	0.77

したものが、図 5.4 になる . 母数の値は、刺激がどれだけ人間の歌声として自然に聴えたかを表す値であり、正の大きな値を示せば、その刺激が自然な歌声である事を表す . 尚、F 検定法による各歌声刺激間の有意差検定を行った結果、各刺激間に 5 % の水準で有意な差があることを確認している .

音響的特徴を何も付加していない合成音 BASE に対して singers' formant を付与することで歌声の自然性が向上する結果となった . これにより、話声と歌声を聴き分ける上で singers' formant が重要な役割を果していることが確認された . 一方で、歌声知覚に与える影響の大きさは、ヴィブラートと比較すると小さい結果となった . しかし、SF-VIB の自然性が VIB に比べ顕著に大きいことから、歌声の自然性向上には、ヴィブラートだけでなく、singers' formant の存在が大きく貢献していることが示された .

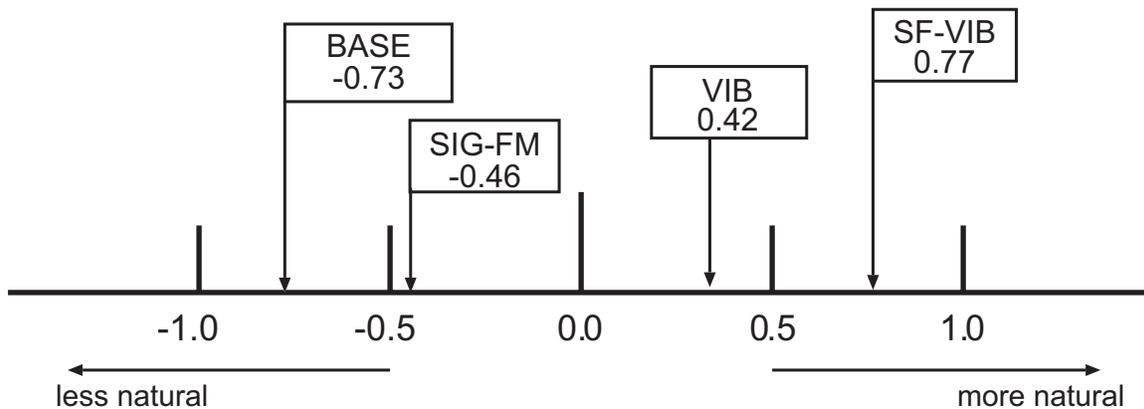


図 5.4: 歌声の自然性の関係 .

### 5.2.3 Singers' formant の特性と歌声知覚の関係の調査

前述のように，Sundberg は，singers' formant のスペクトルピーク成分が話声に比べ平均で 20 dB 高いことを示している．しかし，この特性と歌声知覚の関係の定量的な調査は行われていない．そこで，singers' formant の大きさを操作した歌声合成音を作成し，その自然性の評価を行った．

#### 歌声合成音の作成

歌声合成音は，前項の実験で作成したテノール歌手の伸長母音合成音 BASE に対して，図 5.2 に示した方法で，3 kHz 付近のホルマントを 0 ~ 25 dB まで 1 dB 刻みで持ち上げたスペクトル系列，及びそれに合わせた非周期成分割合を制御した合成音を作成した．

#### 聴取実験

歌声合成音の評価は，歌声の自然性（5段階評価：不自然～非常に自然）に関する絶対評価実験で行った．実験条件は，先の実験と同様である．

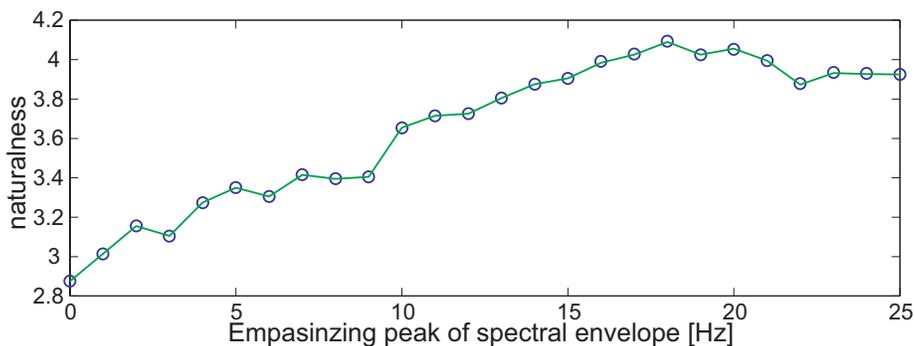


図 5.5: Singers' formant の大きさと歌声の自然性の関係 .

### 実験結果・考察

実験結果を図 5.5 に示す . 図の横軸はスペクトル包絡を強調させた大きさを , 縦軸は合成音の自然性をそれぞれ示す . singers' formant 制御に対する顕著な自然性変動は見られないものの , 10 dB 以上スペクトルを強調させることで徐々に自然性が向上し , 18 dB 強調することで最も高い自然性を示す結果となった . この特性は , Sundberg が報告する平均的な singers' formant のピークの大きさ 20 dB に近い値となった .

以上の 2 つの実験を通じて , singers' formant の歌声知覚に与える影響は , 歌声に「響き」を与える , 又はオーケストラ演奏の中でも歌声を「聞こえやすく」する為だけでなく , 話声と歌声の違いを規定し , 且つ歌声に自然性を与えていることが定量的に明らかとなった .

## 5.3 ヴィブラートに同期したスペクトル制御

Svec は , 歌唱時の声区変換に伴う基本周波数変化に連動し , スペクトルの高域成分が弱まることを報告している [84] . また , Slawson は , 合成母音を作成する際に , 母音の基本周波数を 1 オクターブ上げると共にホルマント周波数を平均 10 % 上げることで , 合成音の音質が向上することを示している [76] . 更に高野らは , 歌唱時に見られる急峻な基本周波数上昇に伴い , 声帯の位置が変化していることを , MRI 分析の結果から示している [87] . これは , 歌唱している場合に , 声道長が基

本周波数の変化に起因して変化している可能性を示唆している。

この様に，歌唱における急峻な基本周波数変化に連動した歌声特有のスペクトル特性がいくつか報告されている．その中でも，ヴィブラート区間における音声の振幅変調やホルマントの振幅変調の存在が多くの研究によって示されている [38, 39, 42, 65, 67]．だが，これら多くの研究では，音声波形全体やホルマントの振幅変調がヴィブラートに付随した成分であるという報告に留まっており，それら成分が歌声知覚に与える影響については調査されていない．しかし，歌声知覚におけるヴィブラートの重要性を考えると，ヴィブラートに伴った一連の振幅変調成分も歌声知覚に大きな影響を与えることが考えられる．実際，过らは，ヴィブラートに連動した音声振幅変調が，歌声らしさの知覚において重要なことを示している [14]．

以上から，本節では，ヴィブラートに同期したホルマント振幅変調を制御可能な歌声合成システムを構築し，その変調成分が歌声知覚に与える影響を調査する．

### 5.3.1 ホルマントの振幅変調制御

図 5.6 に，歌声データセット B に収録されたテノール歌手の歌唱音声/a/の基本周波数変化，振幅エンベロープ，サウンドスペクトログラムを示す．ここから，振幅エンベロープとホルマントが，基本周波数におけるヴィブラートに同期して振幅変調していることが確認できる．そこで，図 4.2 の歌声合成システムにおいて，STRAIGHT で合成された音声の振幅エンベロープを変調させることで，個々のホルマントも振幅変調させる制御法を提案する．制御法の概要を図 5.7 に示す．

最初に，入力音声  $X(t)$  は振幅エンベロープ  $E_x(t)$  とキャリア成分  $C_x(t)$  に分解される．振幅エンベロープ  $E_x(t)$  はヒルベルト変換を用いた次式によって抽出される．

$$E_x(t) = LPF[Hilbert(X(t))], \quad (5.2)$$

ここで， $Hilbert(\cdot)$  はヒルベルト変換である．また，ローパスフィルタ  $LPF[\cdot]$  のカットオフ周波数は 30 Hz とした．

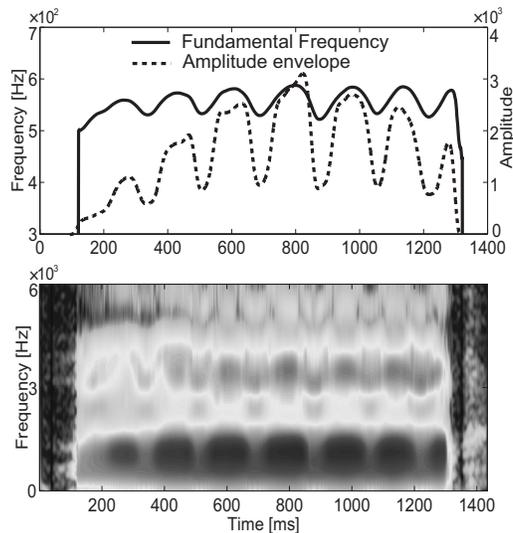


図 5.6: ヴィブラートに同期した振幅エンベロープとホルマントの振幅変調（上：基本周波数と振幅エンベロープ，下：サウンドスペクトログラム）。

キャリア成分  $C_x(t)$  は次式で計算される。

$$C_x(t) = \frac{X(t)}{E_x}, \quad (5.3)$$

その後，抽出された振幅エンベロープ  $E_x(t)$  に対し次式を適用することで振幅変調されたエンベロープ  $E_{am}(t)$  が計算される。

$$E_{am}(t) = (1 + \alpha \sin(2\pi\Omega t))E_x(t), \quad (5.4)$$

ここで， $\alpha$  は，振幅変調の偏移幅の大きさを決めるパラメータ値， $\Omega$  は，F0 制御モデルにおけるヴィブラート制御パラメータ値である。また，上式が適用されるのは，F0 制御モデルによってヴィブラートが制御される区間であり，同位相の振動が制御される。

最後に，振幅変調された音声信号  $X_{am}(t)$  は，次式によって求められる。

$$X_{am}(t) = C_x(t)E_{am}(t), \quad (5.5)$$

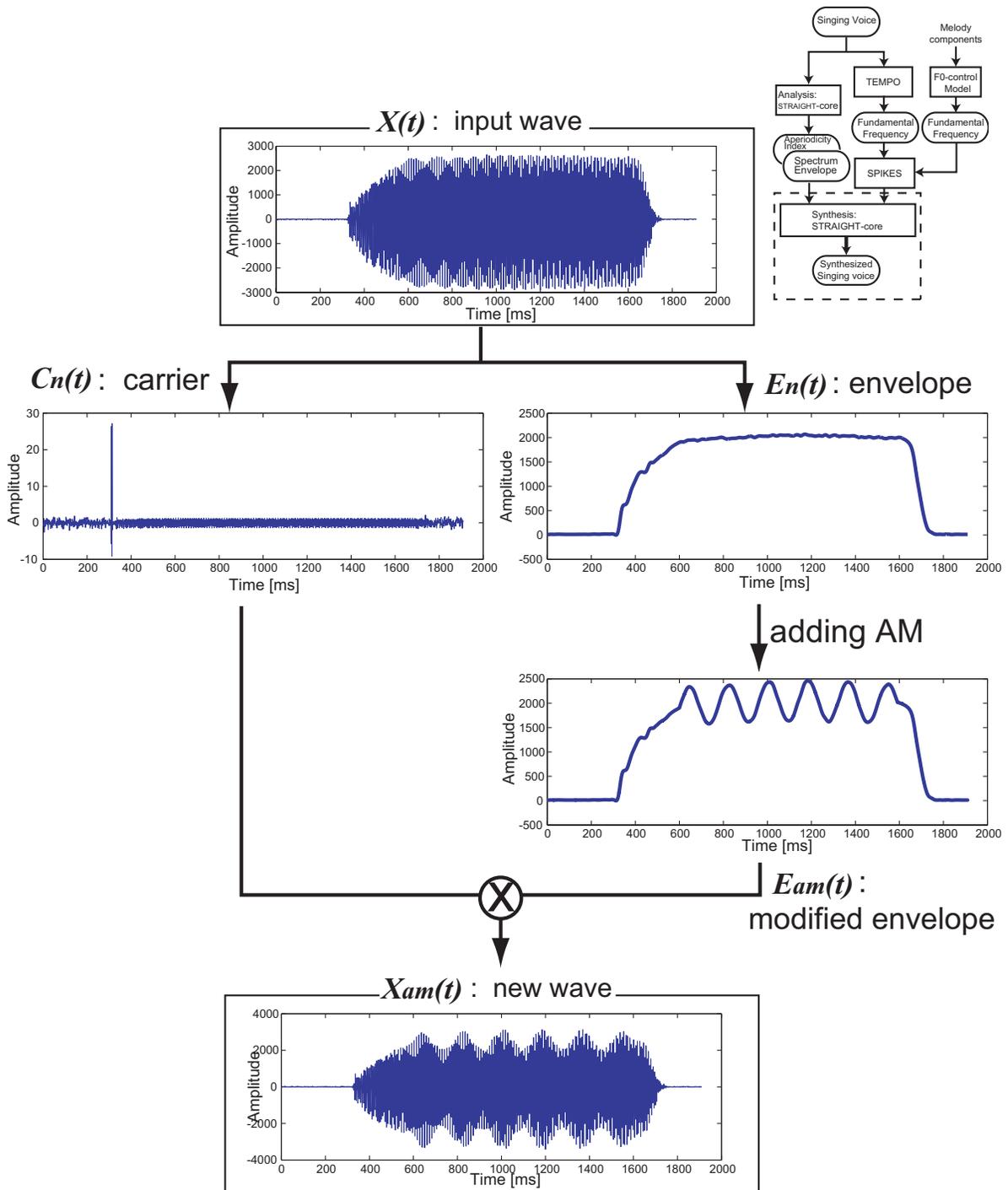


図 5.7: 振幅変調制御法の概要 .

### 5.3.2 ホルマントの振幅変調が歌声知覚に与える影響の調査

#### 歌声合成音の作成

歌声合成音は、図 5.7 の方法で作成した以下の 4 つである。

- BASE

歌声データセット B に収録されるテノール歌手の単独発話母音/a/ (時間長: 0.9 ms) を STRAIGHT によって時間長 1.6 ms まで伸長した合成音。

- VIB

BASE の基本周波数に対して、F0 制御モデルによってヴィブラートを制御した合成音 (制御パラメータ値は表 3.2 の値を使用)

- AM

BASE の音声エンベロープに対して振幅変調を制御した合成音。この際、振幅変調の偏移幅の大きさを設定する制御パラメータ値  $\alpha$  は 0.2 とした。

- VIB-AM

BASE に対してヴィブラートとエンベロープの振幅変調を制御した合成音。

#### 聴取実験

上記の歌声合成音をシェッフェの一対比較法で評価した。実験方法は 5.2.2 節の実験と同様であり、被験者は正常な聴力を有した大学院生 6 名 (男性 4 名、女性 2 名) である。

#### 実験結果・考察

上記の実験で得られたデータを、浦の変法によって処理した結果を表 5.2 及び図 5.8 に示す。図の見方は、5.2.2 節の実験結果と同様である。尚、F 検定により、どの刺激間においても有意差水準 5% で有意な差があることを確認している。

実験の結果、ヴィブラートに加えて、エンベロープに振幅変調を加えた合成音が最も自然性の高い歌声として知覚された。また、振幅変調のみ制御した場合も、その影響はヴィブラート制御に比べ小さいものの、自然性が向上することが確認

表 5.2: 母数の推定 (自然性) .

歌声	母数
BASE	-1.18
AM	0.10
VIB	0.36
VIB-AM	0.72

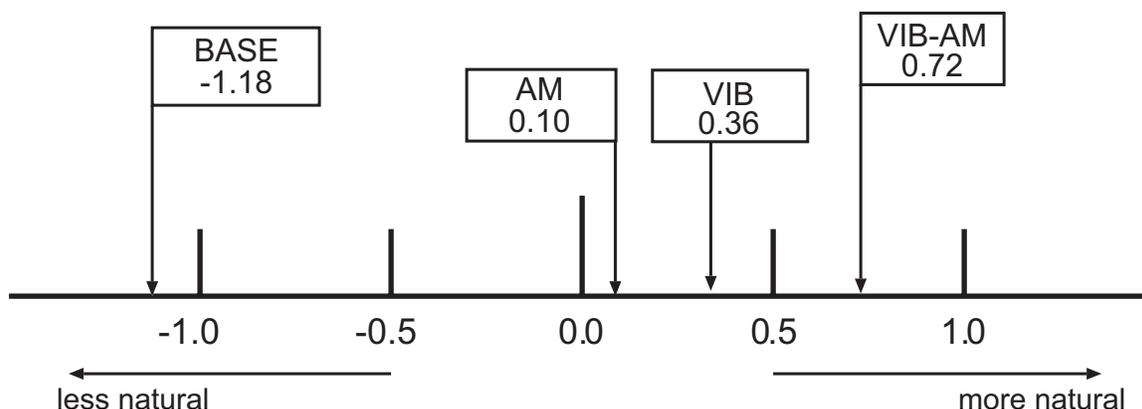


図 5.8: 歌声合成音の自然性の関係 .

された．この結果から，ヴィブラートに同期したホルマントの振幅変調（音声振幅変調）が単なるヴィブラートに付随して生成される成分ではなく，話声と歌声の違いを規定し，歌声知覚に影響を与えていることが明らかとなった．

### 5.3.3 振幅変調の偏移幅の大きさと歌声知覚の関係の調査

これらの報告では，ヴィブラートに同期した音声振幅変調の偏移幅の大きさは，20～40%の範囲にあると示している．そして，20%の偏移幅の振幅変調制御を行うことで，歌声らしさの知覚が向上することを示している．しかし，偏移幅の大きさを変えた場合の歌声知覚の反応は調査していない．そこで，振幅変調の偏移幅の大きさと歌声知覚の関係について調査を行う．

## 歌声合成音の作成

歌声合成音は、図 5.7 の方法で作成した以下の 4 つである。

- BASE

歌声データセット B に収録されるテノール歌手の単独発話母音 /a/ ( 時間長 : 0.9 ms ) を STRAIGHT によって時間長 1.6 ms まで伸長した合成音。

- VB-AM02

BASE に対してヴィブラートと音声エンベロープの振幅変調を制御した合成音。ヴィブラート制御パラメータは表 3.2 の値，振幅変調の偏移幅の大きさを設定する制御パラメータ  $\alpha$  は 0.2 の値を使用。

- VB-AM03

BASE に対してヴィブラートと音声エンベロープの振幅変調を制御した合成音。振幅変調の偏移幅の大きさを設定する制御パラメータ  $\alpha$  は 0.3 の値を使用。

- VB-AM04

BASE に対してヴィブラートと音声エンベロープの振幅変調を制御した合成音。振幅変調の偏移幅の大きさを設定する制御パラメータ  $\alpha$  は 0.4 の値を使用。

## 聴取実験

上記の歌声合成音をシェッフェの一対比較法で評価した。実験方法は 5.2.2 節の実験と同様であり，被験者は正常な聴力を有した大学院生 6 名 ( 男性 4 名，女性 2 名 ) である。

## 実験結果・考察

上記の実験で得られたデータを，浦の変法によって処理した結果を表 5.3 及び図 5.9 に示す。図の見方は，5.2.2 節の実験結果と同様である。尚，F 検定により，どの刺激間においても有意差水準 5 % で有意な差が確認された。

表 5.3: 母数の推定 (自然性) .

歌声	母数
BASE	-1.24
VB-AM02	0.61
VB-AM03	0.47
VB-AM04	0.17

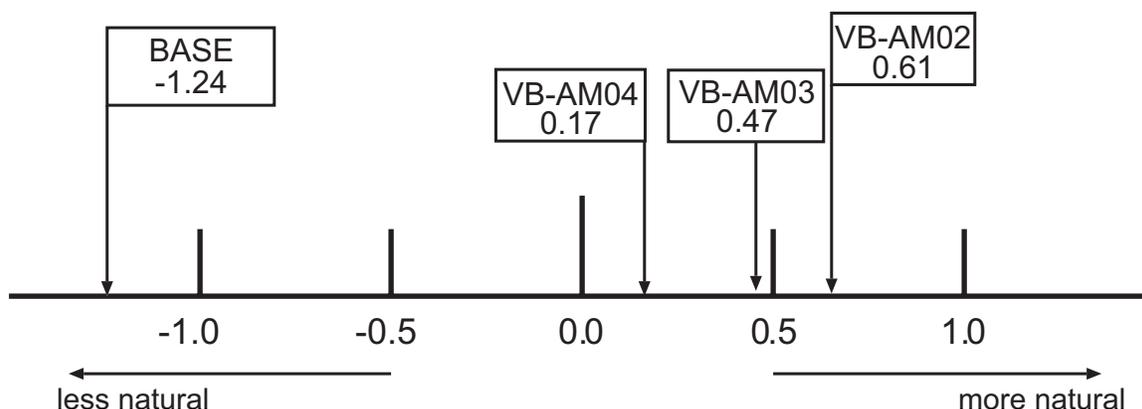


図 5.9: 歌声合成音の自然性の関係 .

実験の結果，音声振幅変調の偏移幅の大きさを変えることで，歌声の自然性が大きく変化することが確認された．そして，20 % 偏移幅の大きさを制御することで，歌声の自然性が最も高くなることが確認され，辻らの報告を支持する結果となった．しかしこの特性は，vibrato extent の値によって変わる可能性も考えられ，今後詳細に検討する必要がある．

## 5.4 まとめ

本章では，前章で提案した歌声合成システムにスペクトル制御を組み込むことで，歌声特有のスペクトル特性と歌声知覚の関係を調査した．

はじめに，singers' formant と歌声知覚の関係を調査した．ここでは，STRAIGHT を用いた歌声合成システムの処理過程において singers' formant を制御する方法を

提案した．そして，singers' formant を制御した歌声合成音を作成し，心理物理実験によって合成音の自然性を評価した．この結果，singers' formant を付与することで歌声の自然性が向上することが明らかとなった．また，singers' formant の特性として，3 kHz 付近のホルマントピークを 18 dB 強調することで歌声の自然性が最も高くなることが確認された．この特性は，先行研究で報告されている平均的な singers' formant の特性に近いものであった．これらの結果により，singers' formant が歌声知覚における「響き」や「聞こえやすさ」に影響を与えるだけの成分でなく，話声と歌声の違いを規定し，歌声の自然性に影響を与える重要な特徴であることが分かった．

次に，ヴィブラートに同期したホルマントの振幅変調と歌声知覚の関係を調査した．ここでは，STRAIGHT を用いた歌声合成システムの処理過程において音声振幅全体を変調することでホルマントの振幅変調を制御する方法を提案した．そして，ホルマントの振幅変調を制御した歌声合成音を作成し，心理物理実験によって合成音の自然性を評価した．この結果，ヴィブラートに同期させホルマントの振幅変調（音声振幅変調）を制御することで，ヴィブラートのみを付与した場合に比べて，歌声の自然性が向上することが明らかとなった．また，音声振幅変調を 20 % の偏移幅で制御することで，歌声の自然性が最も高くなることが確認された．これらの結果から，音声振幅変調とホルマント振幅変調が，ヴィブラートによって引き起こされた副次的成分というだけでなく，話声と歌声の違いを規定し，且つ歌声の自然性を向上させる上で重要な役割を担っていることが明らかとなった．

## 第 6 章

### 高品質な歌声合成システムの実装

## 6.1 はじめに

本章では，第2章で提案した歌詞の朗読音声から歌声に変換・合成するシステムを構築する．そして，これまで着目してきた音響的特徴が，話声と歌声の違いを規定する上で，又歌声を知覚する上でどれだけの役割を担っているかについて検証する．

はじめに，第2章で示した歌声合成システムを実装する為に音韻長制御法を提案する．ここでは，曲のテンポ情報に基づき歌詞の朗読音声の音韻長を歌声の音韻長に伸長する方法を提案する．

次に，第2章で提案した歌声合成システムの実装を行い，歌声知覚において最も重要な役割を担っている音響的特徴について検討する．ここでは，歌声F0制御モデル，スペクトル制御法，音韻長制御法を統合することで，歌詞の朗読音声から歌声に変換・合成可能なシステムを実装する．そして，歌声の基本周波数特性とスペクトル特性を付与した歌声合成音を作成し，聴取実験で合成音の評価を行う．これにより，話声と歌声の違いを規定し，且つ歌声知覚に影響を与える音響的特徴を明らかにする．

## 6.2 音韻長制御法の構築

前章までにおいて，基本周波数とスペクトルそれぞれの音響パラメータにおいて，歌声を知覚する上で重要な音響的特徴が明らかになった．しかし，我々人間が話声と歌声の違いを聴き分ける為にどの音響的特徴が重要であるか，また，基本周波数とスペクトルのどちらに含まれる音響的特徴が最も歌声知覚に影響を与えるのかは明確になっていない．それを検証するためには，話声と歌声の大きな違いである音韻長を制御する方法を構築し，第2章で提案した歌詞の朗読音声から歌声に変換・合成するシステムを実装する必要がある．そこで本節では，話声（歌詞の朗読音声）と歌声の音韻長の違いを検討し，音韻長制御法を提案する．

歌詞を朗読した際の各音韻長は，その音韻に割り当てられた音符の時間長で決まる．そして音符の時間長は，その音符固有の長さ（例えば，四分音符か八分音符）と曲全体のテンポによって決まる．また，それに伴い各音韻を構成する音素の

長さも話声と歌声によって変化する．そこで，1つの音符が割り当てられた音韻ごとに時間伸長処理を施す方法を提案する．図 6.1 に音韻長制御法の概要を示す．

この方法によって時間伸長を行う前処理として，子音-母音の連鎖部を子音部 + 結合部 + 母音部にセグメンテーションする．ここで，結合部の長さは子音-母音境界の  $-10 \sim 30$  ms までの計 40 ms とする．そして，各部それぞれに対して，以下に示す時間伸長処理を施す．

### 子音部の時間伸長

歌声データセット A に収録されるラベリングされた朗読音声と歌声のデータから子音の音韻長を計測し，朗読音声の子音長に対する歌声の子音長の比率を求めた結果，表 5.2 に示す伸長率が得られた．この表から，子音長の制御は調音位置によらず調音方式のみで行えることが分かる．すなわち，摩擦音 1.28 倍，破裂音 1 倍，半母音 2.37 倍，鼻音 1.43 倍， $/y/$  1.22 倍である．また，音符の種類が変わっても，子音部分の伸長率はほとんど変わらないことも確認された．よって子音部スペクトル包絡及び非周期成分割合の時間伸長は，これらの伸長率に従って線形に伸長を行う．

### 結合部の時間伸長

結合部の 40 ms 区間においては時間伸長処理を行わず，朗読音声と同じスペクトル包絡をもつものとする．

### 母音部の時間伸長

母音部では，この音韻に割り当てられた音符の長さから，伸長した子音部の長さおよび結合部の 40 ms を差し引いた長さにスペクトルを線形に伸長する．

尚，子音部と母音部のスペクトル伸長には，線形補間を用いる．また，非周期成分割合に関しても，上記と同様の手順で時間伸長を行う．

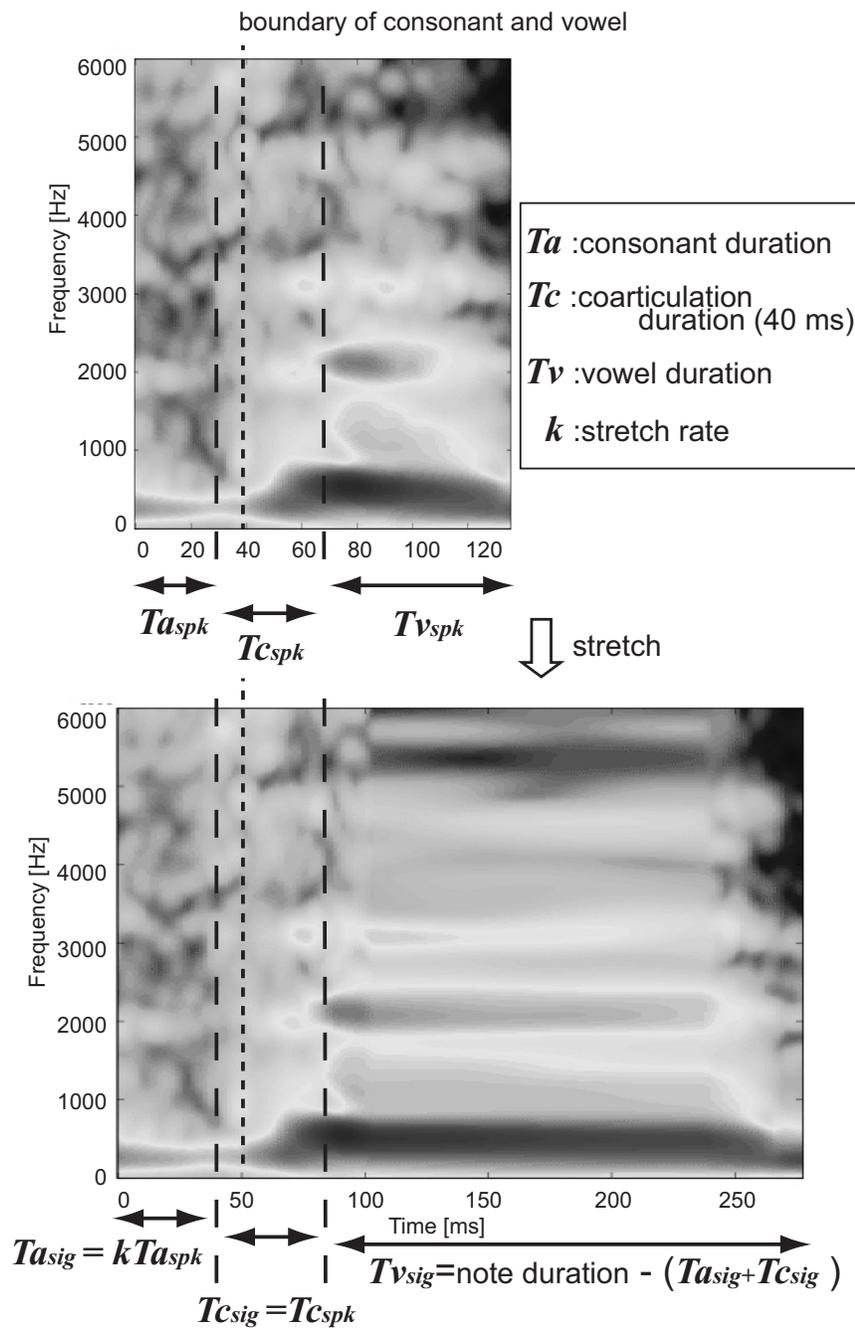


図 6.1: テンポに基づいた音韻長制御の概要 .

表 6.1: 話声の子音の時間長に対する歌声の子音の時間長 .

	口唇	歯・歯茎		口蓋		声門
	有声	有声	無声	有声	無声	
摩擦音		/z/, 1.37	/s/, 1.18			/h/ 1.28
破裂音		/d/, 1.00	/t/, 1.09	/g/, 1.14 ,	/k/, 0.97	
半母音・流音	/w/, 2.61	/r/, 2.12				
鼻音	/m/, 1.35	/n/, 1.50				

## 6.3 歌声知覚に重要な音響パラメータの検討

### 6.3.1 歌声合成システムの実装

前節までにおいて提案した基本周波数制御モデル，スペクトル制御法，そして音韻長制御を図 2.1 に示した STRAIGHT ベースの歌声合成システムに組み込むことで，歌詞の朗読音声から歌声を合成するシステムを実装した．システムの構成を図 6.2 に示す．

#### 基本周波数制御 (F0 control model)

歌声の基本周波数は，譜面情報から作成されたメロディ成分に対し，歌声 F0 制御モデルによってオーバーシュート，ヴィブラート，プレパレーション，微細変動を付与することで記述される．オーバーシュート制御に関しては，4.8 節で行った音高上昇時と下降時それぞれにおいて異なる制御パラメータ値を抽出して制御を行う方法を採用する．プレパレーションに関しては，3.4 節で歌声データセット A，B を対象に抽出した制御パラメータ値 (表 3.2) によって制御を行う．ヴィブラート制御に関しては，4.7 節で提案した vibrato rate の時間変動を制御可能な方法を採用する．そして微細変動は，F0 制御モデルにおける従来手法によって制御を行う．

## 音韻長制御 (duration control)

音韻長制御は、STRAIGHT の分析部で抽出された朗読音声のスペクトル包絡、及び非周期成分割合の時間系列に対して適応される。ここでは、予め用意した朗読音声のセグメント情報と、上記の F0 制御モデルの入力であるメロディ成分から得られるテンポ情報を利用することで、各音韻長を時間伸長する。

## スペクトル制御 (spectral control 1,2)

スペクトル制御は、2つの処理から構成される。はじめに、音韻長制御によって時間伸長された朗読音声の母音部分のスペクトル包絡及び非周期成分割合の時間系列に対して、5.2節で提案した手法によって singers' formant が制御される。ここで制御される singers' formant の特性は、3 kHz 付近に存在するホルマントのピークを 18 dB 持ち上げたものである。この制御をスペクトル制御 1 (spectral control 1) と呼ぶ。その後、singers' formant が付与されたスペクトル包絡と非周期成分割合の時間系列、及び歌声 F0 制御モデルによって記述された基本周波数変化を用いて STRAIGHT によって合成を行う。そして、この合成音に対して 5.3 節で提案した手法によってヴィブラートに同期した振幅変調制御が行われる。ここで付与される振幅変調の偏移幅は話声の振幅エンベロープの 20 % の大きさである。この制御をスペクトル制御 2 (spectral control 2) と呼ぶ。

以上の処理工程によって、歌詞の朗読音声から歌声合成音に変換される。

### 6.3.2 歌声合成音の作成

図 6.2 に示したシステムを用いて歌声合成を行った。システムへの入力は、日本童謡「七つの子」の最初の 2 小節「からすなぜなくの」の朗読音声である。以下に、作成した歌声合成音を記す。

- SPEAK

歌詞の朗読音声の分析・再合成音。

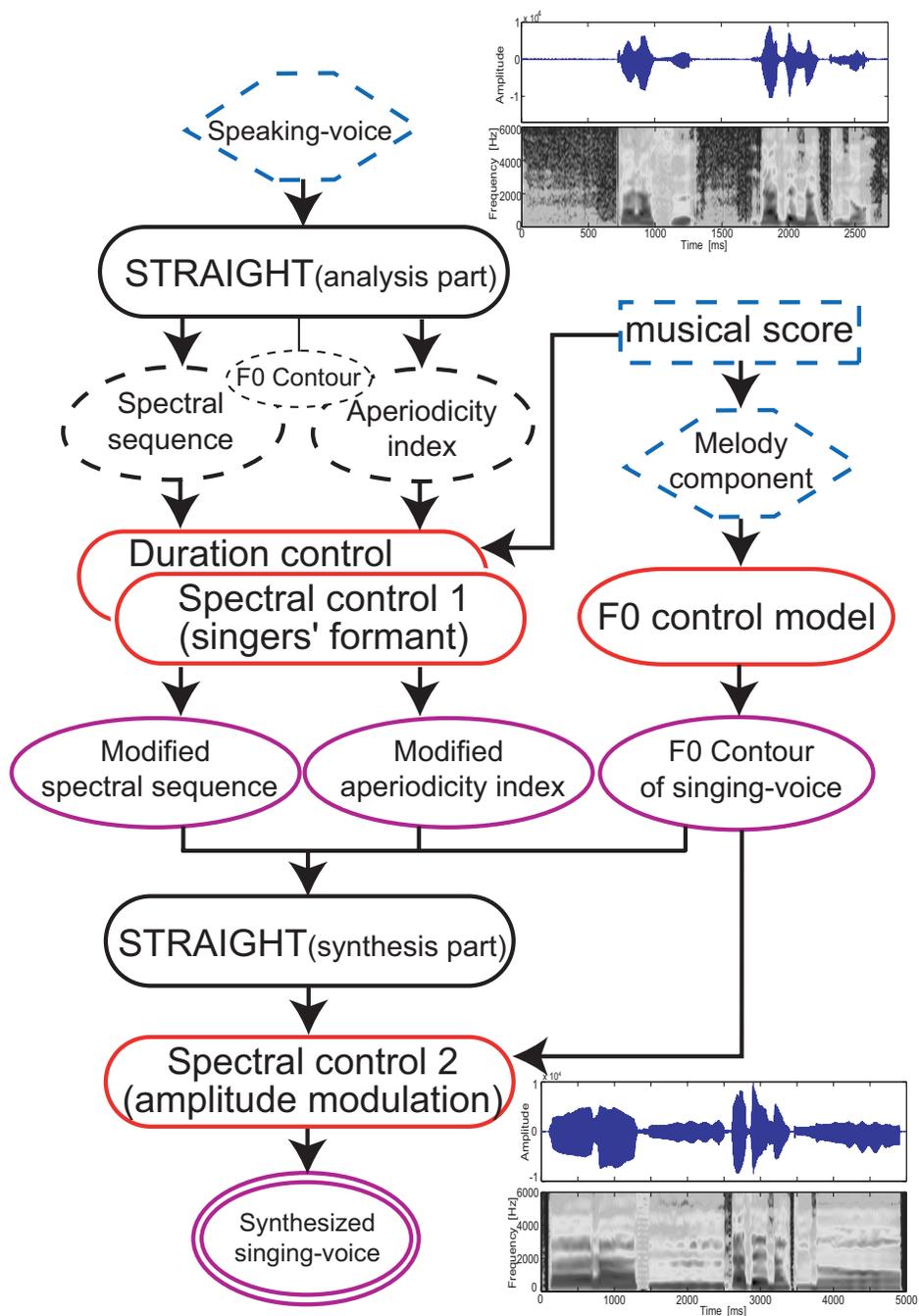


図 6.2: 話声からの歌声合成システム.

- SING-BASE  
メロディ成分に F0 動的変動成分を制御せず，スペクトル系列と非周期成分割合に対しては音韻長制御のみ行った合成音．
- SING-F0  
メロディ成分にすべての F0 動的変動成分を付与し，スペクトル系列と非周期成分割合に対して音韻長制御のみ行った合成音．
- SING-SP  
メロディ成分に F0 動的変動成分を制御せず，時間伸長したスペクトル系列と非周期成分割合に対してスペクトル制御 1，2 の処理を施した合成音．
- SING-ALL  
メロディ変化，スペクトル系列，非周期成分割合に対してすべての制御を行った合成音．
- SING-REAL  
同一発声者による歌声の分析・再合成音．

### 6.3.3 聴取実験

上記の歌声合成音を聴覚刺激とし，シェッフェの対比較法によって自然性に関する評価（7段階尺度）を行った．実験方法はこれまでと同様であり，被験者は正常な聴力を有した大学院生 5 名（男性 4 名，女性 1 名）である．

### 6.3.4 実験結果と考察

実験結果を浦の変法によって処理した結果を表 6.2 及び図 6.3 に示す．各刺激の数値は歌声の自然性の高さを表し，値が大きいほど（水直線上の右にあるほど），その歌声合成音が自然であると知覚されたことを示す．尚，F 検定により，どの刺激間においても有意差水準 5 % で有意な差があることを確認している．

実験結果から，音韻長制御，基本周波数制御，そしてスペクトル制御をそれぞれ行うことで，朗読音声に近づく結果となった．そして，すべての制御を

表 6.2: 母数の推定 (自然性) .

歌声	母数
SPEAK	-1.3
SING-BASE	-0.23
SING-F0	0.41
SING-SP	-0.11
SING-ALL	0.59
SING-REAL	0.64

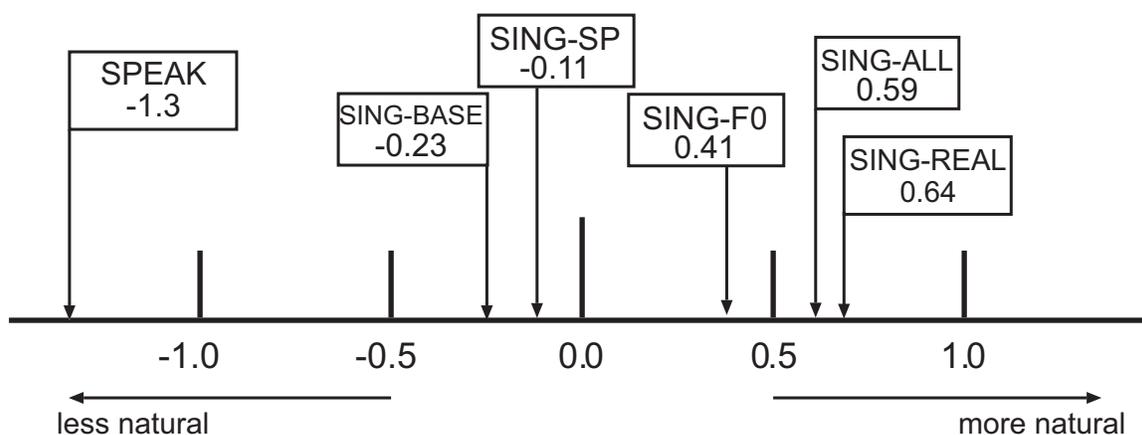


図 6.3: 歌声合成音の自然性の関係 .

行った歌声合成音は、実際の歌声と比べても遜色無い音質で、かつ発声者の声質も保たれた歌声であることが明らかとなった。

SPEAK から SING-BASE に歌声としての自然性が顕著に上昇していたことから、やはりメロディ成分（メロディに対応した矩形的な基本周波数変化とテンポに対応した音韻長）の存在が話声と歌声の違いを規定する決定的な要因であることは間違いないと考えられる。その一方で、SING-BASE から SING-SP、SING-F0 への自然性向上から分かるように、メロディ成分のみ存在しても自然な歌声として知覚されず、これまで着目してきた音響的特徴が存在してはじめて原音声の自然性に近づく結果となった。その中でも、F0 動的変動成分が歌声知覚に与えている影

響は非常に大きく、これまでの実験結果を裏付けるものとなった。また、スペクトルに含まれる音響的特徴の影響はF0 動的変動成分に比べて小さいが、SING-F0 から SING-ALL への自然性向上を見て分かるように、歌声がより自然に知覚される上で重要な役割を果たしていることが確認された。

これらの結果から、歌声知覚に最も重要な音響的特徴は、メロディの情報を持った音韻長と基本周波数変化だけでなく、その基本周波数変化中に存在するF0 動的変動成分であることが明らかとなった。また、歌声特有のスペクトル特性は、F0 動的変動成分に比べ歌声知覚への影響は小さいものの、F0 動的変動成分が存在する歌声がより良い歌声として知覚される為に重要な役割を果たしていることが明らかとなった。つまりは、歌を歌うことによって必然的に生成されるオーバーシュートやプレパレーションの存在によって、我々人間が自然と感じる歌声知覚が成立し、プロの歌唱者の卓越した歌唱技術によって生成されるようなスペクトル特性の存在によって、歌声として知覚されたものがより良い歌声として知覚されることが明らかとなった。

### 6.3.5 全体考察

以上の取り組みから、我々人間が歌声を知覚する上で、大きく2種類の役割を担っている音響的特徴が存在することが明らかとなった。

1つ目の音響的特徴としては、話声と歌声の違いを規定し、且つあらゆる歌声を「歌声」として知覚する上で必要となる成分である。これには、メロディ変化に対応した基本周波数の矩形的変動成分と楽曲のテンポ情報に対応した音韻長成分の2つから成るメロディ成分と、基本周波数変化に規則的に存在するF0 動的変動成分が挙げられる。これら成分に共通して言えることは、どちらも歌を歌うことによって必然的に生成される、つまりは歌声に共通して存在するという点である。中でも、オーバーシュートとプレパレーションは、急峻な基本周波数変化に伴って生成される物理現象である可能性が強く、その為にメロディ成分中にこれら2つの成分が存在しない歌声は不自然に知覚される結果となった。

2つ目の特徴としては、上記で歌声として知覚されたものが「より良い歌声」として知覚される上で必要となる成分である。これには、F0 動的変動成分の1つで

あるヴィブラートとこれに同期したホルマントの振幅変調や, singers' formant が挙げられる。そして, これら成分に共通することは, プロ歌唱において顕著に観測される, つまりは卓越した歌唱技術によって生成される成分ということである。ヴィブラートに関しては, プロ歌唱とアマチュア歌唱において特性が大きく異なり, 中でも vibrato rate の時間特性の違いが歌声知覚に大きな変化を与えていることが確認された。また, singers' formant に関しても, メロディ成分と F0 動的変動成分が存在する歌声に付与されることによって, 歌声知覚に大きな影響を与えていることが確認された。

これら 2 種類の音響的特徴の存在が明らかになったことにより, 我々の歌声知覚機構には 2 つ知覚過程が存在することが考えられる。つまりは, メロディ成分と F0 動的変動成分を知覚することによって話声と歌声を区別, 又はあらゆる歌声を歌声として知覚する過程と, ヴィブラートや singers' formant 等の音響的特徴を知覚することによって, 歌唱技術の違いや歌唱法の違いを知覚する過程である。そして, これら歌声知覚機構の存在は, メロディに伴って基本周波数を急峻に変化させることによってオーバーシュートやプレパレーションといった F0 動的変動成分を生成させるといったあらゆる歌唱に共通の歌声生成過程と, それに加えてヴィブラートや singers' formant といった音響的特徴を制御する卓越した歌唱技術による歌声生成過程から成る歌声生成機構の存在を示唆するものと考えられる。

## 6.4 まとめ

本章では, 歌詞の朗読音声から歌声に変換・合成するシステムを構築することで, 話声と歌声の違いを規定し, 且つ歌声知覚において重要な役割を担っている音響的特徴について検討した。

はじめに, 音韻長制御法を提案した。ここでは, 歌詞の朗読音声の各音韻を母音部, 結合部, 子音部にセグメンテーションし, それぞれの時間長を楽曲のテンポ情報に基づいて伸長する方法を提案した。

次に, 第 3 章で提案した歌声 F0 制御モデル, 第 5 章で提案したスペクトル制御法, そして本章で提案した音韻長制御法を STRAIGHT の分析・合成系に組み込むことで, 歌詞の朗読音声から歌声に変換する歌声合成システムを実装した。そし

て、歌声の基本周波数特性とスペクトル特性を付与することで、歌詞の朗読音声から歌声合成音を作成した。その結果、F0 動的変動成分を制御することで歌声の自然性が最も向上することが明らかとなった。また、スペクトル特性に関しては、F0 動的変動成分が存在する歌声に対して制御することで、歌声の自然性が向上する結果となった。これにより、話声と歌声の違いを規定し、且つ歌声を知覚する上で最も基本的な役割を担っている音響的特徴として、メロディ成分とF0 動的変動成分であることが明らかとなった。また、スペクトル特性は、F0 動的変動成分が存在した上で、歌声知覚に影響を与える成分であることが明らかとなった。

## 第 7 章

### 結論

## 7.1 本論文で明らかになったことの要約

本論文では、歌声知覚・生成機構の解明に向けた基礎的取り組みとして、歌声を知覚する上で重要な役割を担っている音響的特徴を明らかにした。その手法として、高品質な歌声合成システムの構築を行い、その過程において以下の3つの課題を遂行することで、歌声特有の音響的特徴と歌声知覚の関係を調査した。

1. 歌声の基本周波数に含まれる歌声知覚に必要な音響的特徴は何か？
2. 歌声のスペクトルに含まれる歌声知覚に必要な音響的特徴は何か？
3. 歌声知覚において重要な役割を担っている音響的特徴は何か？

### 歌声の基本周波数に含まれる歌声知覚に必要な音響的特徴

歌声の基本周波数変化は楽曲のメロディを表し、話声とは大きく異なる特性を示すことから、歌声を知覚する上で基本周波数特性は重要な役割を担っていることが考えられる。その一方で、我々人間は、馴染みの無いメロディの歌声や、母音を単独発声した歌声を聴いた場合も、それらを歌声として知覚している。そこで、本論文では、歌声の基本周波数変化に含まれるメロディ変化とは異なる成分である F0 動的変動成分に着目し、歌声知覚との関係を調査した。

まず第3章にて、歌唱技術や歌唱法の異なる様々な歌声データに共通して含まれる F0 動的変動成分を抽出した。ここでは、STRAIGHT-TEMPO によって抽出された高精度な基本周波数変化を対象に、先行研究で報告された F0 動的変動成分や、それ以外の特徴的な成分が存在するか否かを調査した。その結果、オーバーシュート、ヴィブラート、プレパレーション、そして微細変動の4つの F0 動的変動成分が多くの歌声に共通して含まれることが確認された。更に、これら成分をメロディ変化に制御可能な F0 制御モデルを提案し、モデルで記述される基本周波数変化と様々な歌声の基本周波数とのフィッティング処理を行った結果、両者の誤差が小さいことから、各 F0 動的変動成分が歌声に共通して含まれることが実証された。

第4章では、前章で提案した歌声 F0 制御モデルを用いた歌声合成システムを提案することで、F0 動的変動成分と歌声知覚の関係を詳細に調査した。

はじめに、F0 動的変動成分の歌声知覚に与える影響を定量的に明らかにする為に、原音声から個々の F0 動的変動成分を除去した場合、及びメロディ変化に各成分を付与した場合の歌声の自然性の変化を調査した。この結果、各成分を除去した場合に歌声の自然性が劣化し、各成分を付与することで歌声の自然性が上昇することが明らかとなった。これにより、4 つの F0 動的変動成分の歌声知覚に与える影響が定量的に示されたと同時に、歌声を歌声として知覚する為には、メロディ成分だけでなく F0 動的変動成分の存在が必要であることが確認された。その中でも、オーバーシュートが歌声知覚に与える影響が非常に大きいことが明らかとなった。

次に、歌唱技術の違いによる F0 動的変動成分の特性の変化、及びその変化が歌声知覚に与える影響を調査した。この結果、オーバーシュートとプレパレーションは、歌唱技術の違いで特性が僅かに異なるものの、その違いは歌声の自然性に影響を与えないほど小さいことが明らかとなった。ヴィブラートに関しては、歌唱技術の違いによって大きく特性が異なり、その特性の違いが歌声知覚に大きな影響を与えていることが明らかとなった。

次に、F0 動的変動成分の変動による歌声知覚の反応を調査した結果、いずれの成分の特性の変化に対しても、歌声の自然性は大きく変化することが明らかとなった。また、この結果と、先の歌唱技術の差と F0 動的変動成分の特性の差異に関する実験結果から、ヴィブラートは歌声共通の音響的特徴であると同時に、歌唱技術の違いを規定する大きな要因であることが確認された。

次に、ヴィブラート振動速度 (vibrato rate) の時間変動特性、及びその特性が歌声知覚に与える影響について調査した。この結果、特にプロ歌唱において vibrato rate が時間的に上昇する傾向が強いことが明らかとなった。そして、この特性を制御することで、定常振動のヴィブラートに比べ歌声合成音の自然性が向上することが明らかとなった。これら一連の結果により、これまで明らかにされてこなかった vibrato rate の時間変化特性と歌声知覚の関係が定量的に示されたと同時に、プロの歌唱者が vibrato rate 特性を巧みに制御することによって、より良い歌声を発声していることが明らかとなった。

最後に、音高上昇・下降時におけるオーバーシュート特性の差異、及びその差異が歌声知覚に与える影響について調査した。この結果、オーバーシュートの特性は、音高上昇時・下降時で異なる特性を示し、更にこの特性が歌唱法や歌唱技術

の違いに関係なくあらゆる歌声に共通したものであることが明らかとなった。また、この特性を制御した歌声合成音の自然性が、特性を同じにした場合に比べて高いことが示された。これら一連の結果により、先行研究において議論の余地が残っていたオーバーシュートの特性、及びその特性が歌声知覚に与える影響が定量的に明らかとなり、歌声知覚・生成過程におけるオーバーシュートが担う役割の解明に有益な知見を齎すものとなった。更にこの実験では、プレパレーションの特性が、音高上昇直前、下降直前それぞれにおいて、ほとんど違いが無いことも明らかとなった。

以上の実験結果から、歌声知覚における F0 動的変動成分の役割として、以下のことが明らかとなった。

- 歌声を歌声として知覚する為には、メロディ変化だけでは十分でなく、その変化中に F0 動的変動成分 -オーバーシュート、ヴィブラート、プレパレーション、微細変動- が存在することが必要である。
- オーバーシュートは、様々な歌声に共通して含まれ、且つ歌唱技術や歌唱法の違いに依存せず同じ特性を持つ傾向が強い。そして、4つの F0 動的変動成分の中で最も歌声知覚に影響を与え、歌声が歌声として知覚される為に最も基本的な役割を果たしている成分である。更に、オーバーシュートは、発声者によって意図的に制御されているのではなく、いかなる歌唱においても、メロディ変化に対応した急峻な基本周波数制御を行う過程において必然的に生成する物理現象である可能性が高く、その為に歌声の自然性に大きな影響を与えていると考えられる。
- ヴィブラートは、様々な歌声に共通して含まれると同時に、歌唱法や歌唱技術の違いを規定している大きな要因となっている。そして、歌声がより良い歌声として知覚される為に最も重要な役割を果たしている。更に、プロの歌唱者はヴィブラート特性（特に vibrato rate）を巧みに制御することによって、アマチュア歌唱では表現できないような豊かな歌声生成を実現させている可能性が高い。
- プレパレーションは、あらゆる歌声に共通して含まれ、且つ歌唱技術や歌唱法の違いに依存せず同じ特性を持つ傾向が強い。歌声知覚への影響は、オー

バーシュートに比べ小さいが、あらゆる歌声を歌声として知覚する為に必要な成分である。更に、オーバーシュートと同様に、メロディ変化に対応した急峻な基本周波数制御を行う過程において必然的に生成する物理現象である可能性が高い。

#### 歌声のスペクトルに含まれる歌声知覚に必要な音響的特徴

スペクトルに含まれる歌声特有の音響的特徴として、本論文では 3 kHz 付近に存在する顕著なスペクトルピーク成分である *singers' formant* と、ヴィブラートに同期したホルマントの振幅変調成分に着目し、これら成分と歌声知覚の関係について調査した。

第 5 章では、はじめに *singers' formant* が歌声の自然性に与える影響について調査した。この結果、*singers' formant* の存在が、歌声の自然性向上の要因であることが確認された。これにより、これまでオペラ歌唱における重要性しか議論されて来なかった *singers' formant* が、一般的な歌声を知覚する上でも重要な役割を担っていることが明らかとなった。また、*singers' formant* のピーク成分の大きさの変化に対する歌声知覚の反応を調査した結果、3 kHz 付近に存在するホルマントピークを 18 dB 強調させた時が、歌声の自然性が最も高い結果となった。これは、Sundberg が示している平均的な *singers' formant* 特性の 20 dB に近い値となり、オペラ歌唱における *singers' formant* の特性が、一般的な歌声知覚においても最も影響を与える特性であることが確認された。

次に、ヴィブラートに同期したホルマント振幅変調成分が歌声の自然性に与える影響を調査した。この結果、ヴィブラートに同期したホルマントの振幅変調（音声振幅変調）を付与することで、ヴィブラートのみを付与した場合に比べて、歌声の自然性が向上することが明らかとなった。また、音声振幅変調を 20 % の偏移幅で制御することで、歌声の自然性が最も高くなることが確認された。この結果は、辻らが報告しているヴィブラートに同期した音声振幅変調の偏移幅の特性と対応が取れる結果であった。

以上の実験結果から、歌声知覚におけるスペクトルに含まれる音響的特徴の役割として、以下のことが明らかとなった。

- singers' formant は、歌唱法の違いに関係なく、プロの歌唱者による歌声全般に存在する成分である。また、先行研究で報告されているような歌声の「響き」や「聞こえやすさ」に影響を与えるだけの成分でなく、話声と歌声の違いを聞き分ける上でも利用され、且つ歌声の自然性に影響を与える重要な音響的特徴である。
- 音声振幅、及びホルマント振幅の変調成分は、ヴィブラート振動に連動して生成される成分で、ヴィブラートが存在する歌声において必然的に存在する。しかし、ヴィブラートによって引き起こされた副次的成分というだけでなく、ヴィブラートと同様に歌声の自然性を向上させる上で重要な役割を担っている。

#### 歌声知覚において重要な役割を担っている音響的特徴

第6章では、前章までで構築した歌声F0制御モデルとスペクトル制御法に加え、楽曲のテンポ情報を基に話声の音韻長を歌声の音韻長に伸張する音韻長制御法を構築することで、歌詞の朗読音声から歌声に変換・合成可能な歌声合成システムの実装を行った。そして、基本周波数とスペクトルそれぞれ音響的特徴の制御を考慮した歌声合成及び聴取実験を行うことで、話声と歌声の違いを規定し、且つ歌声知覚に最も影響を与えている音響的特徴を明らかにした。この結果、話声と歌声の違いを規定する最も基本的且つ重要な音響的特徴はメロディ成分であるが、歌声を歌声として知覚する為にはF0動的変動成分がメロディ変化中に存在する必要があることが明らかとなった。また、スペクトルに含まれる音響的特徴の自然性向上への寄与度は、F0動的変動に比べ小さいものの、歌声がより良い歌声として知覚される上で重要な役割を担っていることが明らかとなった。

以上の実験結果から、我々人間が話声と歌声を聞き分ける、更にはあらゆる歌声を知覚する上で最も重要な音響的特徴は、いかなる歌声においても共通して含まれるメロディ成分とオーバーシュート、プレパレーションに代表されるF0動的変動成分であることが明らかとなった。つまりは、あらゆる歌声に共通した音響的特徴が存在することによってはじめて歌声が歌声として知覚され、そしてその音響的特徴を生成する行為こそが、歌唱技術や歌唱法の違いに依存しない「歌を

歌う」という人間共通のコミュニケーション手段に他ならないと考えられる。そして、優れた歌唱技術を持つ歌唱者は、「歌を歌う」という行為の中で、上記の音響的特徴の生成はもちろん、ヴィブラートや、それに同期したホルマント振幅変調、更には singers' formant といった音響的特徴を巧みに制御・生成することによって、話声からは得られないほどの大きな感動を与える歌声を生み出しているのである。

以上の取り組みによって、我々人間がどのような音響的特徴を利用することで歌声を知覚しているのか、その一端が明らかとなった。ここで得られた知見は、歌声知覚の解明に大きく貢献できると考えられる。更に、それら特徴を明らかにしていくことで高品質な歌声を合成可能なシステムを構築することができた。これにより、序論で述べた話声を対象とした音声合成と同様の枠組みで研究を遂行できる環境が整ったと考えられる。今後は、これまでに得られてきている「歌を歌う」ことに関する様々な知見を今回提案した歌声合成システムに適用することで、それら知見の信憑性が検証可能となると同時に、そこから新たな知見が生み出されることが期待できる。そして、これら一連の取り組みは、歌声合成の研究の発展はもちろん、歌声知覚・生成機構の解明に大きく貢献できると考えられる。歌声合成の発展は、計算機が卓越した歌唱技術を習得することと同値であり、新たな音楽創造の可能性を提供できることが期待できる。それに加え、歌声生成機構の解明が進められれば、優れた歌唱技術の習得に役立つようなボーカル育成支援システムといった音楽教育の分野に大きく貢献できることも考えられる。そして、これらの研究を遂行していくことで、最初に挙げた問題である“我々人間は如何にして歌を聴き・発声しているのか？”という大きな問題に対して、近い将来明確な解答を得ることができると考えられる。

## 7.2 今後の課題

以下に、今後の課題を列挙する。

### 1. F0 動的変動成分の特性に関する詳細な検討

本論文で対象とした F0 動的変動成分は、オーバーシュート、ヴィブラート、プレパレーション、微細変動の4つである。これ以外に、歌声共通して存在する F0 動的変動成分が存在することは考えにくいだが、各成分の特性に関してより詳細に検討する必要がある。例えばヴィブラートに関しては、本論文を通して vibrato rate の時間変動特性が明らかになったものの、vibrato extent の時間特性については未だ明らかになっていない。その為、vibrato extent の時間変動を考慮した歌声合成を行うことで、その特性が歌声知覚に与える影響について調査する必要がある。

### 2. 歌声スペクトルに含まれる音響的特徴に関する検討

本論文では、singers' formant のピークの大きさと歌声知覚の関係は検討したが、ピークの位置と帯域幅の違いが歌声知覚に与える影響についても検討する必要がある。何故なら、中山らの報告では、邦楽歌唱における singers' formant が 4 kHz 付近にも存在することが示されており [59]、singers' formant が歌唱法の違いを規定する役割を担っている可能性が考えられるからである。また、ヴィブラートに同期したホルマント振幅変調に関しても、vibrato extent と振幅変調の偏移幅の関係についても詳細に検証する必要がある。更には、上記以外のスペクトルにおける音響的特徴についても検証して行く必要がある。例えば、基本周波数の急峻な変化に連動したスペクトル構造の変化や、ホルマント周波数の移動に関する検討は、歌声知覚機構はもちろん、歌声生成機構を解明する上で重要な取り組みと考えられる。

### 3. 音響的特徴と歌声生成機構の対応関係に関する検討

歌声生成機構に関しては、特に基本周波数制御を司る声帯制御機構において未解明なことが数多くある。例えば、音高変化に伴う急峻な基本周波数変化、とりわけ基本周波数を急峻に下降させる声帯制御機構は明らかになっていない。また、ヴィブラートの制御機構に関しても、今だ明確な解答が得られていない [88]。これらの問題に取り組む1つの方法として、本論文で提案した歌声 F0 制御モデルに対して生理的な意味付けを行うことが有効と考えられる。

実際に藤崎は、話声の基本周波数を2次系のモデルで記述する方法を提案し、そのモデルと声帯制御を司る喉頭内筋の運動を模擬した力学モデルとの対応関係を見出すことで、発話における声帯制御機構の解明に有益な知見を齎している [36]。同様の枠組みで歌声 F0 制御モデルを発展させていけば、歌声特有の声帯制御機構の解明に大きく貢献できると同時に、生成機構を考慮した高品質な歌声合成システムが構築可能になると考えられる。

#### 4. 多様な歌唱法を考慮した歌声分析・合成に関する検討

本論文によって、あらゆる歌声を知覚する上で重要な音響的特徴が明らかとなり、更にはそれらを制御することで高品質な歌声を合成可能なシステムが構築された。今後は、この歌声合成システムに、先行研究で報告されている様々な歌唱法に関する知見を適用することで、それら知見の検証、更には多様な歌唱法を考慮した歌声合成システムの構築が可能になると考えられる。これらの取り組みによって、歌唱法の違いを規定する音響的特徴の存在や、その特徴の生成機構の解明に対し有益な知見を得ることができると考えられる。

以上一連の課題を遂行することは、高品質な歌声合成システムの実現、更には歌声知覚・生成機構の解明に繋がるものである。そして本論文における歌声知覚に関する新たな知見の発見や歌声合成システムの構築は、今後の歌声研究の発展に大きく貢献できると考えられる。

# 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の意を表します。

本論文を執筆するにあたり、草稿の段階から貴重な御助言を御指導を賜りました北陸先端科学技術大学院大学 情報科学研究科 党建武 教授，情報科学研究科 鶴木祐史 助教授，知識科学研究科 西本一志 助教授，和歌山大学 システム工学部 河原英紀 教授に心より感謝致します。

本研究の遂行，ならびに筆者が NTT コミュニケーション科学基礎技術研究所において副テーマを遂行するにあたり熱心な御指導と御協力を賜りました NTT コミュニケーション科学基礎技術研究所 榊原健一氏に心より感謝致します。

筆者が金沢大学在学中から今日に至るまで，多大なる御指導と御助言を賜りました金沢大学 工学部 船田哲男 教授に心より感謝致します。

本研究を進めるにあたり，日頃から熱心な議論と多面に渡る御協力を賜りました熊本大学 工学部 加藤浩介氏，並びに北陸先端科学技術大学院大学 音情報処理学講座の皆様，及び諸先輩方に厚くお礼申し上げます。

最後に，日頃から暖かく見守り研究生活を支えてくれた妻，大学院での貴重な研究生活を与えてくれた両親に心から感謝しお礼を申し上げます。

## 参考文献

- [1] 天坂 格朗, 長沢 伸也. 官能評価の基礎と応用 日本企画協会.
- [2] 今川 博, 榊原 健一, 田山 二郎, 新美成二. “喉頭管および下咽頭の形状の音声への影響,” 日本音響学会秋期講演論文集, 3-8-5, 2003.
- [3] 小田切 わか菜, 粕谷 英樹. “歌声のヴィブラートの分析、合成、知覚に関する検討” 日本音響学会秋季講演論文集, 1-7-5, 1999.
- [4] 柏野 邦夫, 村瀬 洋. ”パート譜を用いたボーカル音分離システム,” 日本音響学会春季講演論文集, 2-9-1, 1998.
- [5] 河原 英紀, 山田 玲子, 久保 理恵子. “STRAIGHT を用いた音声パラメタの操作による印象の変化について,” 日本音響学会聴覚研究会資料, H-97-65, 1997.
- [6] 河原 英樹. “聴覚の情景分析が生んだ高品質 vocoder: STRAIGHT,” 日本音響学会誌, vol. 54, no. 7, pp. 521-526, 1998.
- [7] 河原 英紀, 片寄 晴弘. “音声分析変換合成システム STRAIGHT を用いたスキット生成について” 情報処理学会論文誌, Vol. 43, No.2, pp. 208-218, 2002.
- [8] 北風 裕教, “歌声の基本周波数の微細変動成分の知覚に関する研究” 北陸先端科学技術大学院大学修士論文, 2000.
- [9] 匂坂 芳典. “コーパス音声合成,” 信号処理学会誌, vol. 2, No. 6, pp. 407-414, 1998.
- [10] 酒向 慎司, 宮島 千代美, 徳田 恵一, 北村 正. “隠れマルコフモデルに基づいた歌声合成,” 情報処理学会研究会, vol.MUS-51, no.13, pp.77-82, 2003

- [11] 酒向 慎司, 宮島 千代美, 徳田 恵一, 北村 正. “隠れマルコフモデルに基づいた歌声合成システム,” 情報処理学会論文誌, vol.45, No.3, pp.719-727, 2004.
- [12] 武田 恵一. “隠れマルコフモデルの音声合成への応用,” 信学技報, SP99-61, 1999.
- [13] 竹本 浩典, 本多 清志, 正木 信夫, 島田 育廣, 藤本一郎. “3次元MRI動画データに基づく声道下部構造のモデル化,” 日本音響学会秋期講演論文集, 3-8-7, 2003.
- [14] 辻 直也, 赤木 正人. “歌声らしさの要因とそれに関連する音響特徴料の検討” 日本音響学会聴覚研究会資料, H-2004-8, 2004.
- [15] 中山 一郎, 小林 範子. “歌の声” 日本音響学会誌 52 巻 5 号, pp. 383-388, 1996.
- [16] 中山 一郎, “日本語を歌・唄・唄う” 日本音響学会誌. vol. 59, no. 11, pp. 688-693, 2003.
- [17] 難波 精一郎・桑野 園子 共著, 音の評価のための心理学測定法 (コロナ社), 1998.
- [18] 藤崎 博也, 須藤 寛, “日本語単語アクセントの基本周波数パターンとその生成機構のモデル,” 日本音響学会誌 27 巻, pp. 445-453, 1971.
- [19] 古井 貞熙 音声情報処理 (森北出版株式会社), 1998.
- [20] 矢田部 学, 粕谷 英樹, “歌声の基本周波数の動特性” 日本音響学会秋季講演論文集, 3-8-6, 1998.
- [21] 山下 崇晴, 天白 成一他, “喉頭の下降機構を考慮した基本周波数制御モデル,” 日本音響学会秋季講演論文集, 1-4-7, 1996.
- [22] YAMAHA ボーカロイド, <http://www.vocaloid.com/en/index.html> .
- [23] NTT-AT ワンダーホルン, <http://www.ntt-at.co.jp/product/wonderhorn/> .

- [24] Akagi, M. and Kitakaze, H., “Perception of synthesized singing-voices with fine-fluctuations in their fundamental frequency fluctuations,” Proc. IC-SLP2000, Vol. 3, pp.458-461, 2000.
- [25] Akagi, M., Iwaki, M., Minakawa, T., “Fundamental frequency fluctuation in continuous vowel utterance and its perception,” Proc. ICSP98, Sydney, Vol. 4, pp.1519-1522, 1998.
- [26] Noll, A. M., “Cepstrum pitch determination,” J. Acoust. Soc. Am., vol. 41, no. 2, pp.293-309, 1966.
- [27] Bartholomew, W. T. “A Physical definition of ‘ good voice quality’ in the male voice,” J. Acoust. Soc. Am., Vol. 6, pp.25-33, 1934.
- [28] Bjoklund, A., “Analyses of soprano voices,” J. Acoust. Soc. Am., Vol. 33, pp.575-582, 1961.
- [29] d’Alessandro C. and Castellongo M., “The pitch of syortduration vibrato tones.” J. Acoust. Soc. Am., Vol. 95, no. 3, pp.1617-1630, 1994.
- [30] de Cheveigne, A., Kawahara, H., “Comparative evaluation of F0 estimation algorithms,” Proc. Eurospeech2001, pp.2451-2454, 2001.
- [31] de Krom, G., and Bloothoof, G., “Timing and Accuracy of Fundamental Frequency Changes in Singing,” Proc. ICPhS 95, Stockholm, Vol. I, pp.206-209 1995.
- [32] Dmitriev, L., and Kiselev, A., “Relationship between the formant structure of different types of singing voice and the dimension of supraglottal cavities,” FP, Vol. 31, pp.238-241, 1979.
- [33] Fujisaki, H., Tatsumi, M., and Higuchi, N., “Analysis of fundamental frequency control in singing,” Transactions of the Committee on Speech Research S79-80, The Acoustical society of Japan, 1980.

- [34] Fujisaki, H., Tatsumi, M., “Analysis control in singing” Vocal fold physiology”, UNIVERSITY OF TOKYO PRESS, pp.347-363, 1981.
- [35] Fujisaki, H., and Hirose, K., “Modeling the dynamic characteristics of voice fundamental frequency with applications to analysis and synthesis of intonation,” Proc. Working Group on Intonation, the        th International Congress of Linguists, Tokyo, pp.50-57, 1982.
- [36] Fujisaki, H., Ohno, S., Narusawa, S., “Physiological mechanisms and biomechanical modeling of fundamental frequency control for the common Japanese and the standard Chinese,” Proc. 5th Seminar on Speech Production 2000, pp.145-148, 2000.
- [37] Hakes, J., Shipp, T., Doherty, T., “Acoustic characteristics of vocal oscillations: vibrato exaggerated vibrato, trill, and tillo,” J. Voice 1, pp.326-331, 1987.
- [38] Hirano, M., Yoshida, Y., Yoshida, T., Tateishi, O., “Stroboscopic video recording of vocal fold vibration,” Ann Otol Rhinol Laryngol 94 (6, Pt 1), pp.588-590, 1985.
- [39] Horii, Y., “Acoustic analysis of vocal vibrato: a theoretical interpretation of data,” J. Voice 3, pp.36-43, 1989.
- [40] Hollien, H., Keister, F., and Hollien, P. A., “Experimental data on‘ singer’s formant,” J. Acoust. Soc. Am., Vol. 64, S171, 1978.
- [41] Iida, A., Higuchi, F., Campbell, N., and Yasumura, M., “A corpus-based speech synthesis system with emotion,” Speech Commu., vol. 40/1-2, pp. 161-187, 2003.
- [42] Imaizumi, S., Saida, H., Shimura, Y., and Hirose, H., “Harmonic analysis of the singing voice: - Acoustic characteristics of vibrato,” Proc. SMAC 93, Stockholm, pp. 197-200, 1994.

- [43] Ishizaka, K., Flanagan, J. L., “Synthesis of voiced sounds from two-mass model of the vocal cords,” *Bell System Tech. J.*, 51(6), pp.1233-1268, 1972.
- [44] Kawai, H., Toda, T., Ni, J., Tsuzaki, M., and Tokuda, K., “XIMERA: A new TTS from ATR based on corpus-based technology,” *Proc. 5th ISCA Speech Synthesis Workshop*, pp. 179-184, 2004.
- [45] Kawahara H. *et al.*, “Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity,” *Proc. Eurospeech99*, pp.2781-2784, 1999.
- [46] Kawahara, H., Masuda-Katsuse, I., de Cheveigne, A., “Restructuring speech representations using a pitch adaptive time-frequency smoothing and an instantaneous-frequency based on F0 extraction: possible role of a repetitive structure in sounds,” *Speech Comm.*, Vol. 27, pp.187-207, 1999.
- [47] Kawahara, H., Hirachi, Y., Morise, M., Banno, H “Procedure ”senza vibrato”: A key component for morphing singing,” *Proc. ICSLP2004*, Spec3701o.5, 2004.
- [48] Klatt, D., “Software for a cascade/parallel formant synthesizer,” *J. Acoust. Soc. Amer.*, Vol. 67(3), pp.971-995, 1980.
- [49] Kojima, K., Yanagida, M., and Nakayama, I., “Variability of vibrato -A comparative study between Japanese traditional singing and bel canto-,” *Proc. SpeechProsody2004(Nara)*, pp. 151-154.
- [50] Large, J. and Iwata, S., “Aerodynamic study of vibrato and voluntary“ straight tone ”pairs in singing,” *Folia Phoniatr(Basel)*, Vol. 23, pp.50-65, 1971.
- [51] Meribeth, A. B., *Dynamic of the Singing Voice* (forth edition), Springer Wien New York

- [52] Minematsu, N., Matsuoka, B., and Hirose, K., “Prosodic analysis and modeling of the NAGAUTA singing to synthesize its prosodic patterns from the standard notation,” Proc. EUROSPEECH, pp.385-388 (2003-9)
- [53] Minematsu, N., Matsuoka, B., and Hirose, K., “Prosodic Modeling of Nagauta Singing and Its Evaluation,” Proc. of SpeechProsody2004(Nara), pp. 487-490, 2004.
- [54] Mori, H., Odagiri, W., Kasuya, H., “Transition characteristics of fundamental frequency in singing,” Proc. ICA2004, Mo 5. Cl.3, pp.1499-1500, 2004.
- [55] Mori, H., Odagiri, W., Kasuya, H., ”F0 dynamics in singing: Evidence from the data of a baritone singer” IEICE Trans. Inf. and Syst., Vol.E87-D, No.5, pp.1086-1092, 2004.
- [56] Moriyama, T., Ogawa, H., Tenpaku, S., “A new control model based on rising and falling fundamental frequency,” Proc. of ASA and ASJ Third Joint Meeting, pp.1171-1176, 1996.
- [57] Murray, I. R., Edgington, M. D., Champion, D., and Lynn, J., “Rule-based emotion synthesis using concatenated speech,” Proc. ISCA Workshop on Speech and Emotion, pp.167-172, 2000.
- [58] Myers D. and Michel J., “Vibrato and pitch transitions,” J. Voice, Vol. 1, no. 2, pp.157-161, 1987.
- [59] Nakayama, I., “Comparative studies on vocal expression in Japanese traditional and western classical-style singing, using a common verse,” Proc. ICA2004, Mo 4. Cl.1, pp.1295-1296, 2004.
- [60] Ohala J. and Ewan W., “Speed of pitch change,” J. Acoust. Soc. Am. Vol. 53, p.345, 1973.
- [61] Ohtsuka T. and Kasuya H., “Robust ARX-based speech analysis method taking voice source pulse train into account,” JASJ, Vol. 58, no. 7, pp.386-397, 2002.

- [62] Prame, E., "Measurements of the vibrato rate of ten singers," J. Acoust. Soc. Am. 96, 1979-1984, 1994.
- [63] Prame, E., "Vibrato extent and intonation in professional Western lyric singing," J. Acoust. Soc. Am. Vol. 102, no. 1, pp.616-621, 1997.
- [64] Press, W. H., Flannery, B. P., Teukolsky, S., Vetterling, W. T., Numerical Recipes in C. Cambridge University Press, Cambridge, 1988.
- [65] Raming L. A. and Shipp T., "Comparative measures of vocal tremor and vocal vibrato," J. Voice, Vol. 1, no. 2, pp.162-167, 1987.
- [66] Rothman H. B. and Arroyo A. A., "Acoustic variability in vibrato and its perceptual significance." J. Voice, Vol.1, no.2, pp.123-141, 1987.
- [67] Rubin, H. J., Le Cover, M., and Vennard, W., "Vocal intensity, subglottic pressure and airflow relationship in singers," Folia Pboniatica, Vol. 19, pp.393-413, 1967.
- [68] Rzhhevkin, S. N., "Cetain results of the analysis of a singer's voice," Soviet Physics Acoustics, Vol. 2, pp.215-220, 1956.
- [69] Scherer, K. R., "Vocal Communication of emotion: A review of research paradigms," Speech Cmmu, vol. 40, pp.227-256, 2003.
- [70] Schroder, M., "Emotional speech synthesis: A review," Proc. EUROSPEECH, pp. 561-564, 2001.
- [71] Seashore, C. E., The vibrato University of Iowa Studies in the Psychology of Music, Vol. I, 1932.
- [72] Seashore, C. E., Psychology of the vibato in voice and instument University of Iowa Studies in the Psychology of Music, Vol. , 1936.
- [73] Seashore, C. E., Objective analysis of musical performance University of Iowa Studies in the Psychology of Music, Vol. , 1937.

- [74] Seashore, C. E., *Psychology of Music* McGraw-Hill, New York, 1938.
- [75] Seidner, W., Schutte, H., Wendler, J., and Rauhut, A., “Dependence of the high singing formant on pitch and vowel in different voice types,” *Proc. SMAC* 83, no. 1, pp.261-268, 1983.
- [76] Slawson, A. W. “Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency, ” *J. Acoust. Soc. Am*, vol. 43, no. 1, pp.87-101, 1968.
- [77] Sundberg, J., “Effects of the vibrato and the ‘ singing formant’ on pitch,” *Musica Slovaca* , Bratislava, 51-69, 1978.
- [78] Sundberg, J., *The Science of the Singing Voice* Northern Illinois University, 1987.
- [79] Sundberg, J., Iwarsson J., and Hagegard, H., “A Singer’s Expression of Emotions in Sung Performance,” *Vocal Fold Physiology*, 1995.
- [80] Sundberg, J., “Maximum speed of pitch changes in singers and untrained subjects,” *J. Phonetics*, Vol. 7, no. 2, pp.71-79, 1979.
- [81] Sundberg, J., “Articulatory interpretation of the‘ singing formant’,” *J. Acoust. Soc. Am.*, Vol. 55, pp.838-844, 1974.
- [82] Sundberg, J., “Singing and timbre,” *Music room acoustic*, Stockholm: Royal Swedish Academy of Music Publications, Vol. 17, pp.57-81, 1977.
- [83] Sundberg, J., “Formant structure and articulation of spoken and sung vowels,” *FP*, Vol. 22, pp.28-48, 1970.
- [84] Svec. J. G., Schutte. H. K., and Miller. D.G., “On pitch jumps between chest and falsetto registers in voice: data from living and excised human larynges,” *J. Acoust. Soc. Am.*, Vol. 106, pp. 1523-1531, 1999.

- [85] Takara, T., Chinen, S., and Ooishi, T., "Transition characteristics of fundamental frequency in singing," Proc. ICA2004, Mo 5. C1.2, pp.1495-1498, 2004.
- [86] Takagi, T., Seiyama, N., and Miyasaka, E., "A method for pitch extraction of speech signals using autocorrelation functions through multiple window-lengths," IEICE Trans. Fundamentals (Japanese Edition), Vol. J80-A, no. 9, pp.1341-1350, 1997.
- [87] Takano, S., Honda, K., Masaki, S., Shimada, Y., and Fujimoto, I. "High-resolution imaging vocal gesture using a laryngeal MRI coil and a synchronized imaging method with external triggering," Proc. Spring Meeting of Acoust. Soc. Jpn, 2-3-8, 2003.
- [88] Titze, I. R., Principles of Voice Production Prentice-Hall, Englewood Cliffs, NJ, 1994.
- [89] Vennard, W., Singing the mechanism and the technique.(2ed ed), New York: Fischer, 1967.
- [90] Vennard, W., Hirano, M., Ohala, J., and Fritzell, B., "A series of four electromyographic studies," National Association of Teachers of Singing Bulletin, October, pp.16-21, December, pp.30-37, 1970; February-March, pp.26-32, May-June, pp.22-30, 1971.
- [91] Weait, C., and Shea, J. B., "Vibrato: an audio-video-fluorographic investigation of a bassoonist," Applied Radiology, January-February, 6:1, pp. 89-90, pp.148, 1977.

# 本研究に関する研究業績

## 論文

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, pp.405-417, 2005.

## 国際会議

Takeshi Saitou, Masashi Unoki, and Masato Akagi, “ Extraction of F0 dynamic characteristics and development of F0 control model in singing voice, ” *Proc. ICAD2002, Kyoto*, 275-278, 2002.

Masashi Unoki, Takeshi Saitou, and Masato Akagi, “ Effect of F0 fluctuations and development of F0 control model in singing voice perception, ” *NATO ASI, Dynamics of Speech Production and Perception in Il Ciocco, Italy*, July 2002.

Takeshi Saitou, Masashi Unoki, and Masato Akagi, “ Development of the F0 Control Model for Singing-Voice Synthesis, ” *Proc. Speech Prosody 2004, Nara*, 2004.

Takeshi Saitou, Masashi Unoki, and Masato Akagi, “ Control Methods of Acoustic Parameters for Singing-Voice Synthesis, ” *Proc. ICA 2004, Kyoto*, 2004.

Takeshi Saitou, Naoya Tsuji, Masashi Unoki, and Masato Akagi, ”Analysis of

Acoustic Features Affecting "singing-ness" and Its Application to Singing-Voice Synthesis from Speaking-Voice," Proc. Inter Speech - ICSLP, Jeju, Korea, 2004

## 研究会

齋藤 毅, 鷓木 祐史, 赤木 正人, " 歌声における F0 動的変動成分の抽出と F0 制御モデル, " 日本音響学会 聴覚研究会, H-2001-92,2001.

齋藤 毅, 鷓木 祐史, 赤木 正人, " 歌声の F0 動的制御モデルにおけるパラメータ決定に関する考察, " 日本音響学会 聴覚研究会, H-2003-111,2003.

齋藤 毅, 鷓木 祐史, 赤木 正人, " 歌声らしさに影響を与える音響的特徴を考慮した話声からの歌声合成法, " 電子情報通信学会/日本音響学会 電気音響・聴覚研究会, EA-2004-30,2004.

齋藤 毅, 鷓木 祐史, 赤木 正人, " 自然性の高い歌声合成のためのヴィブラート変調周波数の制御法の検討, " 電子情報通信学会/日本音響学会 言語と思考・聴覚研究会, TL2005-10, 2005.

## 口頭発表

齋藤 毅, 鷓木 祐史, 赤木 正人, " 歌声における F0 動的変動成分の抽出と F0 制御モデル, " 日本音響学会平成 13 年度春季研究発表会 講演論文集, 1-9-19,2002.

齋藤 毅, 赤木 正人, " 歌声の F0 制御モデルにおける制御パラメータの改良, " 日本音響学会平成 15 年度秋季研究発表会 講演論文集, 2-Q-21, 2003.

齋藤 毅, 鷓木 祐史, 赤木 正人, " 歌声の F0 制御モデルにおける制御パラメータ値の変動が歌声の自然性に与える影響の調査, " 日本音響学会平成 15 年度春季研究発表会 講演論文集, 2-7-3,2004.

齋藤 毅, 鷓木 祐史, 赤木 正人, " 歌声知覚に影響を与える音響的特徴を考慮した話声からの歌声合成法, "

日本音響学会平成 16 年度秋季研究発表会 講演論文集, 3-6-2, 2004.

齋藤 毅, 鶴木 祐史, 赤木 正人, “ ヴィブラート変調周波数の時間変化が歌声知覚に与える影響の調査, ” 日本音響学会平成 16 年度春季研究発表会 講演論文集, 3-P-14, 2005.

齋藤 毅, 赤木 正人, 榊原 健一, “ 声区変換を伴う歌声合成を目的とした音響パラメータ制御法の検討, ” 日本音響学会平成 17 年度秋季研究発表会 講演論文集, 2-6-10, 2005.