修 士 論 文

歌声におけるF0動的変動成分の抽出とF0制御モデルに関する研究

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

2002年3月

修 士 論 文

歌声におけるF0動的変動成分の抽出とF0制御モデルに関する研究

指導教官　赤木正人 教授

審査委員主査　赤木正人 教授
審査委員　党建武 助教授
審査委員　下平博 助教授

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

010046 齋藤 毅

提出年月: 2002年2月

概 要

本稿では、歌声の基本周波数（以下 F0 と呼ぶ）に見られる動的変動成分の効果を心理物理実験にて検討することで、歌声知覚に与える影響を明らかにし、主要となる F0 動的変動成分の抽出を行う。また、歌声知覚に影響を与える F0 動的変動成分を制御可能な F0 制御モデルを提案することで、歌声合成への応用を試みる。その結果、歌声知覚に影響を与える F0 動的変動成分として、オーバーシュート・アンダーシュート、ヴィブラート、微細変動、予備的変化を抽出した。また、それら 5 つの動的変動を制御し F0 を記述できる F0 制御モデルを提案し、歌声合成に応用した。その結果、提案した F0 制御モデルの高品質な歌声合成への応用可能性を確認した。

目 次

第 1 章　序論 **1**

1.1 はじめに 1

1.2 本研究の背景 1

1.3 本研究の目的・特色 2

1.4 本論文の構成 3

第 2 章　歌声の F0 動的変動成分 **4**

2.1 目的 4

2.2 歌声の F0 動的変動成分の分析 4

2.2.1 歌声データ 5

2.2.2 TEMPO(STRAIGHT) による F0 推定 5

2.2.3 F0 動的変動成分 7

2.3 F0 動的変動成分の歌声知覚への影響 9

2.3.1 合成音の作成 9

2.3.2 心理物理実験 13

2.3.3 実験結果と考察 14

2.4 まとめ 16

第 3 章　歌声の F0 制御モデルと歌声合成 **17**

3.1 目的 17

3.2 話声の F0 制御モデル 17

3.3 歌声の F0 制御モデル 20

3.3.1 歌声の F0 制御モデルの構築 20

3.3.2 F0 制御モデルによる最適な F0 制御 23

3.4 F0 制御モデルの歌声合成への応用１ 29

3.4.1 Klatt Formant Synthesizer による歌声合成 29

3.4.2 心理物理実験 32
3.4.3 実験結果と考察 32
3.4.4 ポルタメントに関する考察 33
3.4.5 この節のまとめ 37
3.5 F0 制御モデルの歌声合成への応用 2 38
3.5.1 STRAIGHT による歌声合成 38
3.5.2 心理物理実験 40
3.5.3 実験結果と考察 41
3.6 まとめ 42

第 4 章 結論 **43**
4.1 本論文で明らかになったことの要約 43
4.2 今後の課題 44

図 目 次

2.1 “かわいい七つの” 歌唱時の F0 6
2.2 F0 中の動的変動成分１ 8
2.3 F0 中の動的変動成分２ 8
2.4 F0 動的変動成分を除去した合成音の作成手順 9
2.5 左:オーバーターゲット、アンダーターゲットを除去した F0(NO-OUS)、右: ヴィブラート・微細変動を除去した F0(NO-VB) 11
2.6 左:予備的変化を除去した F0(NO-PRE)、右:スムージングした F0(NO-PRE) 11
2.7 評価段階 14
2.8 歌声刺激の自然性の関係 15

3.1 藤崎 F0 モデルの概要 18
3.2 森山らのモデルの概要 18
3.3 本研究で提案した F0 制御モデルの概要 20
3.4 右:F0 制御モデルにおけるメロディ成分 (矩形パルスによる表現). 左: 歌声変動成分４ (白色雑音から作られる微細変動) 22
3.5 右：歌声変動成分１, ３ (減衰振動). 左：歌声変動成分２ (定常振動) 22
3.6 「カラスなぜなくの」歌唱時の、上：実音声の F0, 下：モデルで生成した F0 25
3.7 「カラスは山に」歌唱時の、上：実音声の F0, 下：モデルで生成した F0 . 26
3.8 「かわいい七つの」歌唱時の、上：実音声の F0, 下：モデルで生成した F0 27
3.9 「子があるからよ」歌唱時の、上：実音声の F0, 下：モデルで生成した F0 28
3.10 歌声合成の手順（Klatt Formant Synthesizer） 29
3.11 歌声合成に用いた F0. 上段の左から順に NORMAL, SYN-ALL. 中段 SYN-OUS, SYN-VB. 下段 SYN-PRE, SYN-BASE の F0 を示す 31
3.12 合成音の自然性の関係 33
3.13 合成音の F0. 上から POLTA1、POLTA2、OUS の F0 を表す 35

3.14 歌声刺激の自然性の関係 36
3.15 STRAIGHT の構成と歌声合成の手順 39
3.16 合成音の自然性の関係 41

表 目 次

2.1 母数の推定 (自然性) 15

3.1 各歌声変動成分の最適パラメータ 24
3.2 母数の推定 (自然性). Klatt Formant Synthesizer による歌声合成の場合 . . 32
3.3 母数の推定 (自然性). オーバーシュート・アンダーシュートに関して . . . 36
3.4 母数の推定 (自然性). STRAIGHT による歌声合成の場合 41

第1章　序論

1.1　はじめに

音声信号処理において、音声合成は非常に重要な技術の一つである。近年、話声を対象とした音声合成の研究は活発に行われ、感情や個人性を考慮した音声合成などの多様な展開を見せている。しかし、話声以外の音声、例えば歌声を対象とした音声合成の研究などは行われているが、実用の段階まで至ってないのが現状である。

歌声 (singing voice) は、話声 (speaking voice) と対比して、発声の強さの幅 (ダイナミックレンジ) が大きく、発声音高 (ピッチ) の幅 (音階) が広いといった、より複雑で動的な特性を持つことが知られている [1]。この特性は、音声波に含まれる特徴、特に基本周波数（以後 F0 と呼ぶ）に顕著に現れる。そのため、声楽家のような「自然」で「歌声らしい」歌声合成の実現を考えた場合、歌声の F0 における動的な特徴を明らかにし、これを制御できる F0 制御モデルを構築することが重要となる。

音声合成技術の話声から歌声への展開は、音声合成システムの応用範囲を広げるだけでなく、計算機音楽や歌声知覚などの分野にも大きな貢献をもたらすことが期待できる。

1.2　本研究の背景

歌声に関する研究として、歌声の F0 における動的変動成分に着目したものがいくつか見られる。矢田部・遠藤・粕谷は楽譜に記されている曲の旋律概形を構成する滑らかに変動するステップ状の成分、および音程変化時のオーバーシュート・アンダーシュートに着目し [2]、小田切・粕谷は歌声特有の 4〜7Hz で周期的に変化するヴィブラート成分に着目した研究 [3] を行っている。また、北風・赤木は F0 変化全体から、上述の成分を取り除いた後に残る微細変動成分に着目している [4]。どの研究においても、着目した動的変動成分が歌声の動特性として重要である事を示しているが、各 F0 動的変動成分が持つ情報、特に歌声知覚に与える効果を明確にしていない。また、他にも歌声特有の F0 動的変動成分が存在する可能性も考えられ、更なる歌声の F0 に関する分析が必要である。

一方で、音声合成の音源波形生成で重要になるF0制御モデルについて、話声に対応したものが提案されている。その中で、藤崎F0モデル[5]は、話し言葉の緩やかなF0変化をアクセント成分とフレーズ成分と呼ばれる２つの成分で制御・生成できるモデルである。このモデルは制御におけるパラメータの数が少なく、生理学的見地から理論が検討されており、現在の音声合成等に幅広く使用されている。また、森山・天白らは、藤崎F0モデルを基に、大阪方言固有のリズムをF0変化中に制御できるF0制御モデルを提案し[6]、山下・天白らは藤崎F0モデルのフレーズ成分を改良する事で、より細かなF0制御を試みている[7]。しかし、どのF0制御モデルも話声のF0を対象にしているので、歌声のF0のような動的変動の制御を行うのは困難であり、現在そのようなモデルが提案されていないのが現状である。よって高品質な歌声合成の実現のためにも、歌声に対応したF0制御モデルの構築が重要となっている。

1.3　本研究の目的・特色

本研究は、(1) 歌声のF0における動的変動成分に着する。そして、各変動成分の歌声知覚に与える影響を心理物理実験によって明らかにすることで、主要となるF0動的変動成分の抽出を行う。そして、(２) 歌声知覚に影響を与えるF0動的変動成分を制御可能なモデルを提案することで、高品質な歌声合成への応用を試みる。この2つを目的とし、研究を進める。

また、本研究の特色は、

- 高精度なF0推定と、独立にF0操作が可能な音声分析合成システムSTRAIGHTを用いる事で、歌声のF0変化における複数の動的変動成分（過去の研究において報告されているものも含め）を抽出し、個々の成分の歌声知覚に与える影響を相対的に明確にすること。

- 歌声知覚に影響を与えるF0動的変動成分の制御モデルを考える事で、少ないパラメータで正確なF0制御、更には自然性の高い歌声合成へ応用すること。

である。

1.4 本論文の構成

本論文は 4 章で構成される

1 章 本論文が対象としている研究分野の背景と問題点を指摘し、本論文の位置付けと目的を示す。

2 章 歌声における F0 動的変動成分の分析に関して述べる。
はじめに本論文で使用する歌声データ紹介、ならびに F0 推定法 TEMPO(STRAIGHT) [8] の概念を述べる。次に歌声特有の F0 変化における大きな特性を 3 つ示し、そのなかで動的変動成分に関して分析を行う。過去の研究で報告されているものも含めた全部で 3 つの F0 動的変動成分に着目し、歌声知覚に与える影響を、F0 を操作した合成音を用いた心理物理実験によって調べ、主要となる F0 動的変動成分の抽出を行う。

3 章 2 章で抽出した F0 動的変動成分を制御できる F0 制御モデルに関して述べる。
はじめに、過去に提案されている話声に適応した F0 制御モデルに関する説明をし、歌声に適応した F0 制御モデルの必要性を検討した後、本研究における F0 制御モデルの詳細に関して述べる。次に、提案した F0 制御モデルによる F0 を用いて Klatt Formant Synthesizer による歌声合成を行い、心理物理実験を通じて F0 制御モデル、ならびに記述した F0 の評価を行う。最後に STRAIGHT による高品質な歌声合成への応用を試みる。

4 章 本論文で得られた結果を要約し、今後の展望を述べる。

第2章　歌声のF0動的変動成分

2.1　目的

歌声のF0変化において、歌声特有の成分であるF0動的変動成分の抽出を2段階のステップで行う。まず第一段階で、歌声データから推定したF0変化において、顕著に観測される変動をF0動的変動成分として着目し、その特性に関して分析を行う。次に第2段階で、F0動的変動成分の歌声知覚に与える影響を心理物理実験によって調べることで、着目したF0動的変動成分の歌声のF0変化における重要性を検証する。そして、歌声知覚に影響を与える成分を、本研究で扱うF0動的変動成分として抽出する。

2.2　歌声のF0動的変動成分の分析

先に述べたように、歌声は話声に比べ、より複雑でダイナミックな特性を持っており、その特性はF0にも顕著に現れている事が知られている[1]。先の研究[2]において、歌声のF0における固有の特性として大きく分けて次に示す3つが観測されることが報告されている。

1. F0の大まかな変化はメロディに、F0の値は音階に対応する。
2. ある一つの音程区間におけるF0変化は、下降せずに安定している。場合によって上昇することもある。
3. 歌声のみ観測される動的な変動成分がある。

1番目に関して、歌声のF0における周波数値は音階に対応し、例えば12平均律において440 Hzの周波数を持つ音はA4（ハ長調のラ）の音を表す。よってF0変化はメロディ変化を直接表す事になる。これは、はっきりしたメロディを持たない話声には見られず、

歌声に特化した特徴であると言える。次に 2 番目に関しては、話声の場合は一つのフレーズにおける F0 は時間と共に減少するのが通常である。これは F0 の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 F0 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ので、一つの音程内で F0 は下降せずに安定している必要がある。やはりこの特徴も歌声に特化したものと言える。以上 2 つの特性は、共にメロディに対応したものであり、F0 変化全体で見ると静的成分と言うことができる。しかし、最後の 3 番目の特性に関しては、時間と共に逐次変化する動的変動成分であり、メロディに関する特性とは大きく異なる。本研究では、3 番目に示す F0 の動的変動成分に着目し、これらの歌声知覚に与える影響を調べることで、主要な F0 動的変動成分の抽出を行う。

2.2.1 歌声データ

本研究で使用した歌声データは、北風らがボーカルスクールに通う学生 (女性 3 名) から採取したものを用いる。実験の簡略化のために、対象となる歌は大衆声楽曲である日本童謡「七つの子」を日本語母音/a/のみで歌唱したものを用いる。被験者には歌唱法に関して特別な指示を与えず、また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、被験者の発声可能な音域を考慮に入れたためである。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 による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 個を本研究における歌声データとして使用する。

2.2.2 TEMPO(STRAIGHT) による F0 推定

歌声の F0 に含まれる動的変動成分を分析するためには、高精度な F0 推定が必要である。そこで、本研究では河原らが提案した音声分析合成法 STRAIGHT の基本周波数推定アルゴリズムである TEMPO を用いることで高精度な F0 推定を行う。

TEMPO の F0 推定アルゴリズムの枠組みは、帯域フィルタの中心周波数とフィルタ出力の瞬時周波数を周波数から周波数への写像とみなし、信号の主要な正弦波成分の周波数を、写像の安定な平衡点に対応する瞬時周波数として求める事であり、2 つの処理段階から構成される。まず最初の段階で、log 周波数軸で間隔が等しい同型の帯域通過フィルタが、フィルタの中心周波数からフィルタ出力の瞬時周波数へのマッピングにおいて不動点

抽出のために用いられる。この不動点はF0に対応する不動点を選択する為に、推定されたC/N比によって評価される。この最初のF0推定を次の第2段階に通すことにより、正確な値に改善される。

第2段階ではF0情報とF0の微分係数による放射上の時間軸の伸縮が、基本周波数の適応STFTを行う前に導入される。この時間歪みSTFTを基にした不動点分析が調波成分に対応する不動点を与える。そのとき、最小推定誤差をもつF0推定を行うためのC/N情報を使うことで不動点の瞬時周波数は統合される。また、調波成分推定されたC/N比は、音声再合成に適切な源信号の周期性を制御するための情報を与える。

図2.1にTEMPOによって推定したF0の一例を示す。尚、TEMPOによるF0推定の精度評価に関しては、論文[9, 4]を参照して頂きたい。

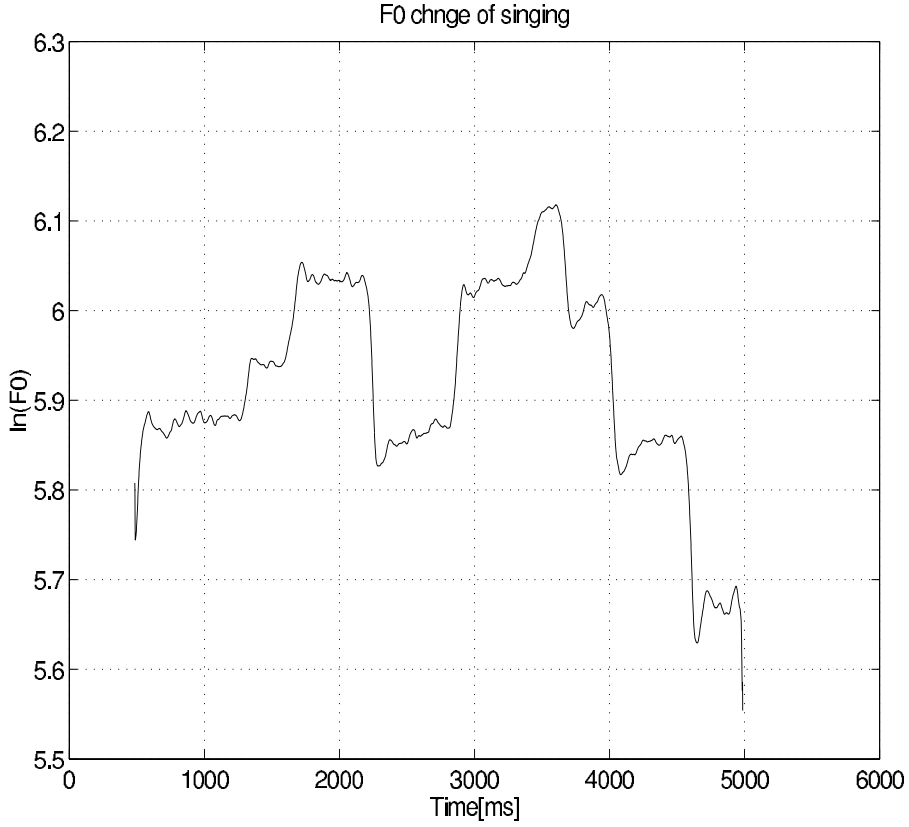

図 2.1: “かわいい七つの” 歌唱時のF0

2.2.3 F0 動的変動成分

図 2.1 に示すような F0 を、すべての歌声データから推定した結果、本研究では以下に示す 3 つの特徴を歌声の F0 動的変動成分として着目した。

- **オーバーシュート／アンダーシュート (Overshoot /Undershoot)**
  傾斜を持った滑らかな音程変化（ポルタメント）から目的音の音高より高く、または低く振れる（プロジェクション）現象。特にアンダーシュー トは顕著に見られる。
- **ヴィブラート／微細変動 (Vibrato / Fine-fluctuation)**
  一つの音高が持続した場合に観測される 4〜7 [Hz] の周期的な振動、さらにはそれを取り除いた後に残る不規則的な細かい変動。
- **予備的変動 (Preparation)**
  音程が変化する直前に観測される音程変化とは逆の方向に変化する瞬時的な変化。

最初の二つの動的変動に関しては、過去の研究において歌声特有の変動成分であると報告されている [2, 3, 4]。しかし、歌声の特有な成分として、これら二つの動的変動だけで十分なのか議論されておらず、この他にも歌声特有な動的変動の存在は十分に考えられる事である。そこで本研究では、上記の三番目の動的変動を新たに歌声の F0 動的変動として考える。これらは、用いた歌声データの F0 変化すべてにおいて観測される特性であり、歌声に特化した特性である可能性が高い。また、従来のオーバーシュート・アンダーシュート成分を、以下の二つの成分に分けて考え、より細かい分析を行う。

ポルタメント (poltament) ：音程の滑らかな変化を表す成分。

プロジェクション (projection) ：ポルタメントである傾斜を持った音程変化後に生じる、目標音高値を振り切る変動成分。

図 2.2 に F0 中におけるオーバーシュート、アンダーシュートと予備的変動を、図 2.3 にヴィブラートと微細変動を表す。

本研究では、各 F0 動的変動成分の歌声知覚に与える影響を明確にすることで、メロディ以外の特性であるこれらの F0 動的変動成分が持つ歌声としての情報を調べる。次節では聴取実験を行うことで、F0 動的変動成分の歌声知覚に与える影響を調査する。

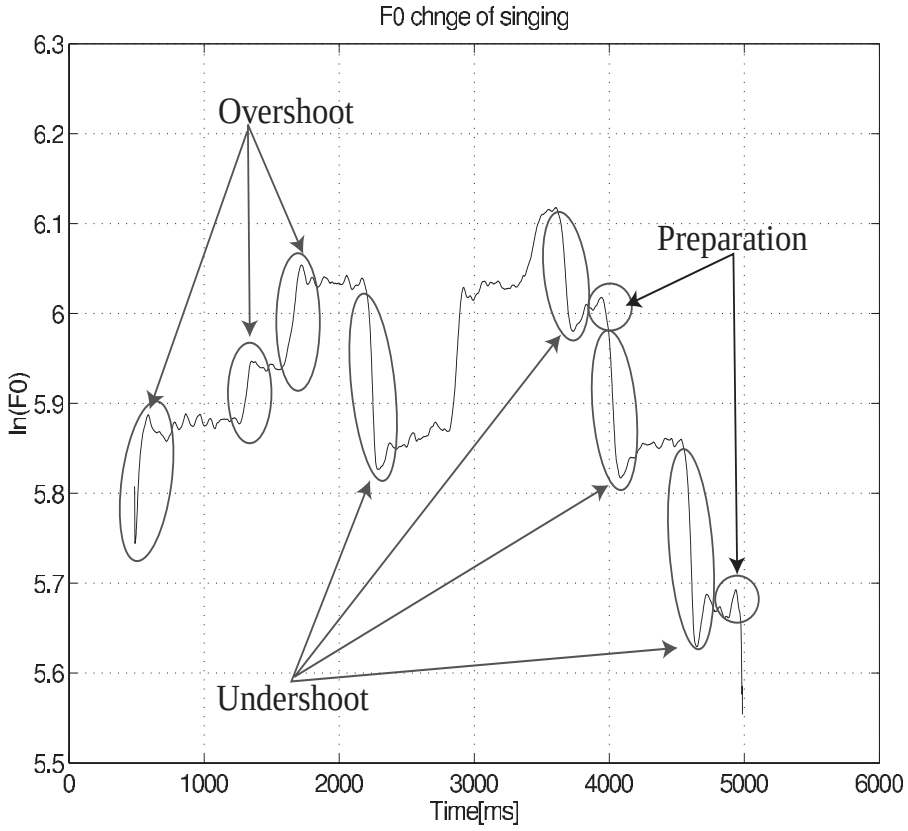

図 2.2: F0 中の動的変動成分 1

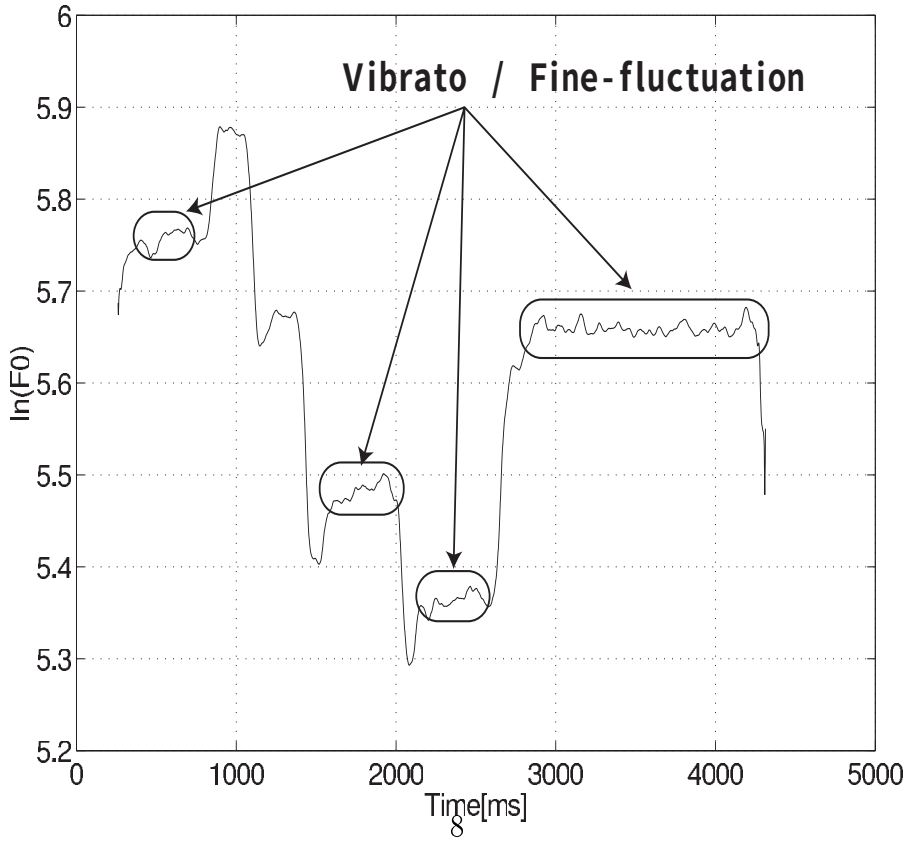

図 2.3: F0 中の動的変動成分 2

2.3 F0 動的変動成分の歌声知覚への影響

先に示した F0 動的変動成分の歌声知覚への影響を調べるために心理物理実験を行う。

2.3.1 合成音の作成

心理物理実験によって F0 動的変動成分の歌声知覚への影響を調べるには、各 F0 動的変動成分を操作した合成音が刺激として必要である。そこで、本研究では図 3.1 に示す手順で合成音を作成した。

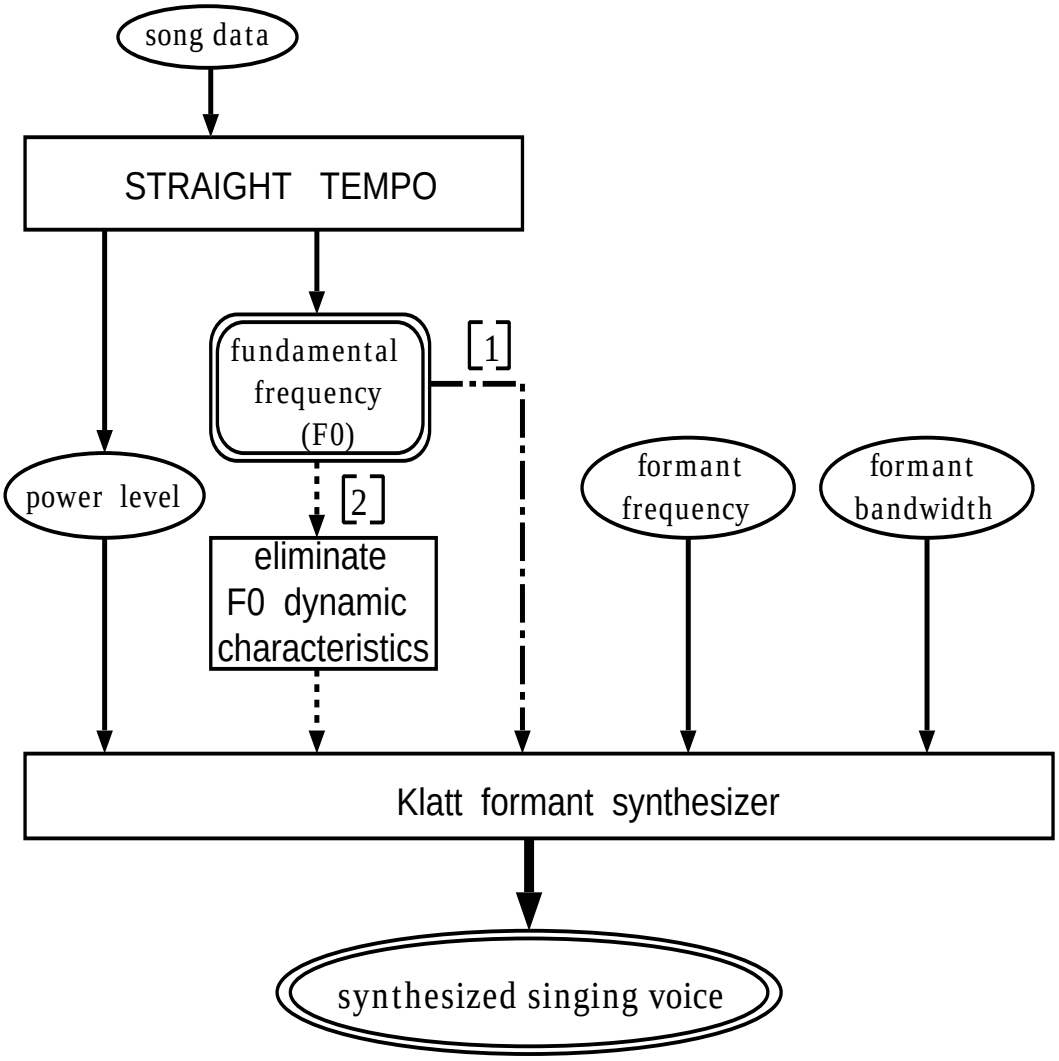

図 2.4: F0 動的変動成分を除去した合成音の作成手順

各合成音は、実音声の F0 を TEMPO によって抽出し、それを直接 Klatt Formant Synthesizer に用いて合成に行う図 3.1 中の [1] の手順と、抽出した F0 に加工を加えて合成を行う [2] の手順で作成された以下の５つである。

- **NORMAL**
  歌声データから TEMPO によって抽出した F0 を加工しないで、Klatt のホルマント合成器によって合成したもの。（図 2.4 における [1] の行程）
- **NO-OUS**
  F0 中のオーバーターゲット、アンダーターゲットを、音程区間の平均値で置き換える事によって除去したもの。（図 2.4 における [2] の行程）
- **NO-VB**
  F0 中のヴィブラート・細かい変動成分を移動平均によって平滑化し除去したもの。（図 2.4 における [2] の行程）
- **NO-PRE**
  F0 中の予備的変化を、音程区間の平均値で置き換える事によって除去したもの。（図 2.4 における [2] の行程）
- **SMS**
  F0 をカットオフ周波数 5Hz でスムージングし、動的変動を除去、及びメロディ構造を少し壊したもの。（図 2.4 における [2] の行程）

各合成音は、歌声データから TEMPO によって推定された高精度な F0 を操作することで、各 F0 動的変動成分が除去されたものとなっている。ここで、それぞれの F0 を図 2.5〜図 2.6 に示す。

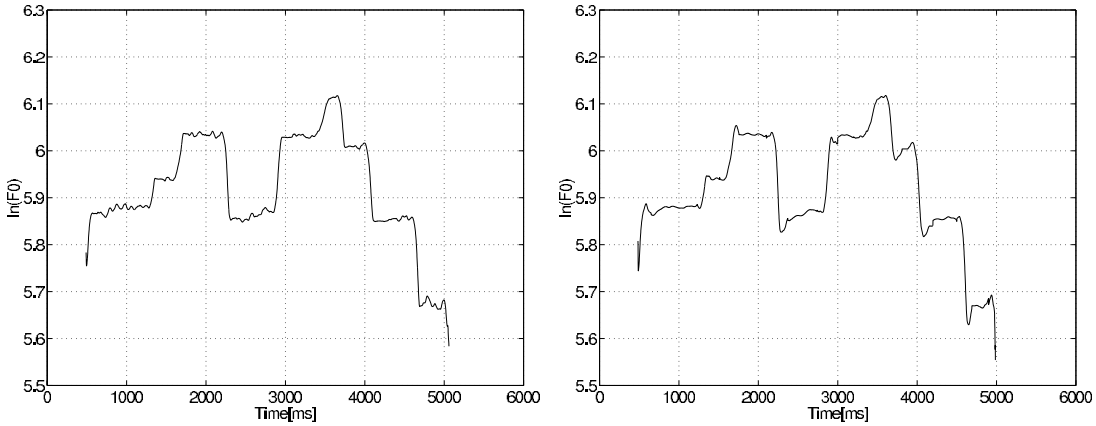

図 2.5: 左:オーバーターゲット、アンダーターゲットを除去した F0(NO-OUS)、右:ヴィブラート・微細変動を除去した F0(NO-VB)

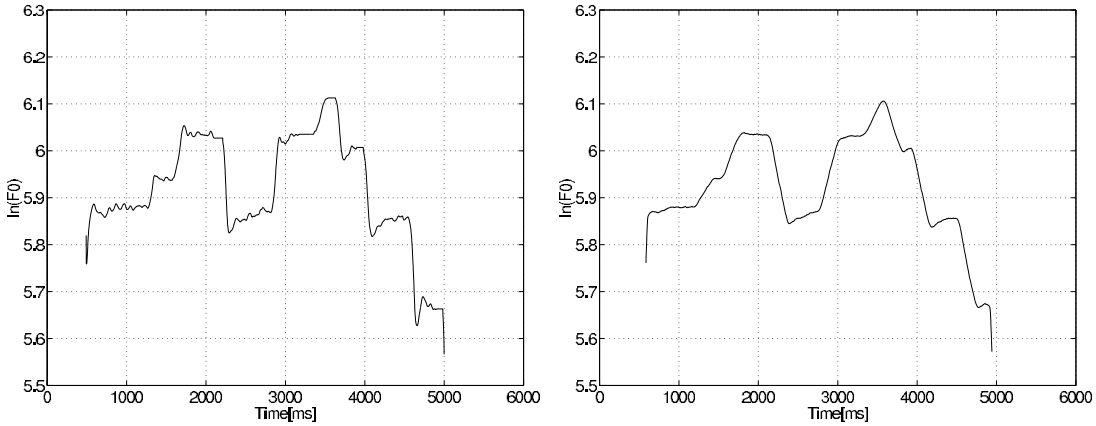

図 2.6: 左:予備的変化を除去した F0(NO-PRE)、右:スムージングした F0(NO-PRE)

今回、合成器としてKlattのホルマント合成器を使用した理由として、この合成器はホルマント周波数、帯域幅、ゲイン、そしてF0を利用して合成音を作るシンプルなものであり、F0を独立に制御できるからである。また、できた合成音のスペクトルはほぼ一定であるので、F0の変化が合成音声の性質に大きく現れる。よって、F0動的変動成分の様な細かい変動も合成音に反映できると考えた。本研究におけるKlattのホルマント合成器は以下の様に設定した。

- 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

2.3.2 心理物理実験

先に示した５つの合成音を用いて、F0 動的変動成分の歌声知覚に与える影響を調べるために心理物理実験を行った。以下に実験に関する詳細を記す。

実験方法

心理物理実験において一対比較法は、数個の刺激を２つずつ対にして判断を求める方法であり、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すみ、二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。本研究では、一対比較法にカテゴリー判断を取り入れたシェッフェの一対比較法を採用した。この方法は、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。一対比較法では、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、シェッフェの一対比較法では、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 [10]。

刺激条件

刺激として用いた歌声は、先に示した NORMAL、NO-OUS、NO-VB、NO-PRE、SMS の５つである。実験ではこれらを２つずつ対にした 20 対を、一人の歌手分の 4 つの歌声データに対して作成した合計 80 対を用いた。なお、同じ刺激同士の対は含まれていない。比較するメロディ間は約 1 秒間、対間は約 2 秒と設定した。

実験条件

実験に参加した被験者は、正常な聴力を有した大学院生 6 名（男性 5 名、女性 1 名）。実験は防音室におけるヘッドフォンを介しての両耳受聴で行った。

手続き

被験者には次のような教示を与え、判断を求めた。

聴取実験

ヘッドフォンから二つの歌を対にしてお聞かせします。前の歌と後の歌を聴き比べて、どちらがどの程度自然な歌（歌声）として聴えたかを判断してください。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。

判断は図 2.7 に示すような 7 段階評価で行った。前の歌 A の方が良ければ正の値、後の歌 B が良ければ負の値を選択してもらい、どちらも同じであるなら 0 を選んでもらった。尚、実験は 2 回に分けられており、被験者の半数は実験 1→ 実験 2 の順で、後の半数は実験 2→ 実験 1 の順で行った。

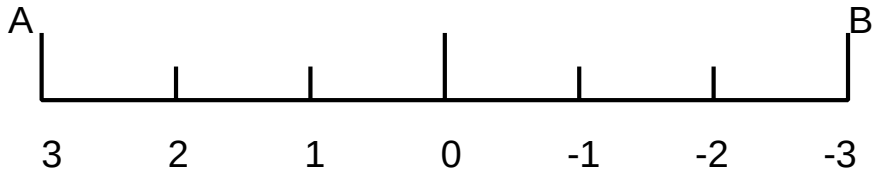

図 2.7: 評価段階

2.3.3 実験結果と考察

上記の実験方法で得られたデータを、芳賀の変法 [10] によって処理した結果を表 2.1 に示し、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 2.8 になる。ここで、母数の値は、その刺激がどれだけ自然な (良い) 歌声に聴えたかを表す値であり、正の値で大きければより自然であることを示す。

この結果を考察すると、各刺激の自然性が NORMAL の値に比べ小さくなっていることが分かる。NORMAL→NO-OUS の母数の減少はオーバーシュート・アンダーシュートにの、NORMAL→NO-VB はヴィブラートと微細変動の、そして NORMAL→NO-PRE は予備的変動の歌声知覚に与える影響をそれぞれ相対的に示している。また、NORMAL→SMS

表 2.1: 母数の推定 (自然性)

歌声	母数
NORMAL	0.72
NO-OUS	-0.03
NO-VIB	0.27
NO-TP	0.25
SMS	-0.85

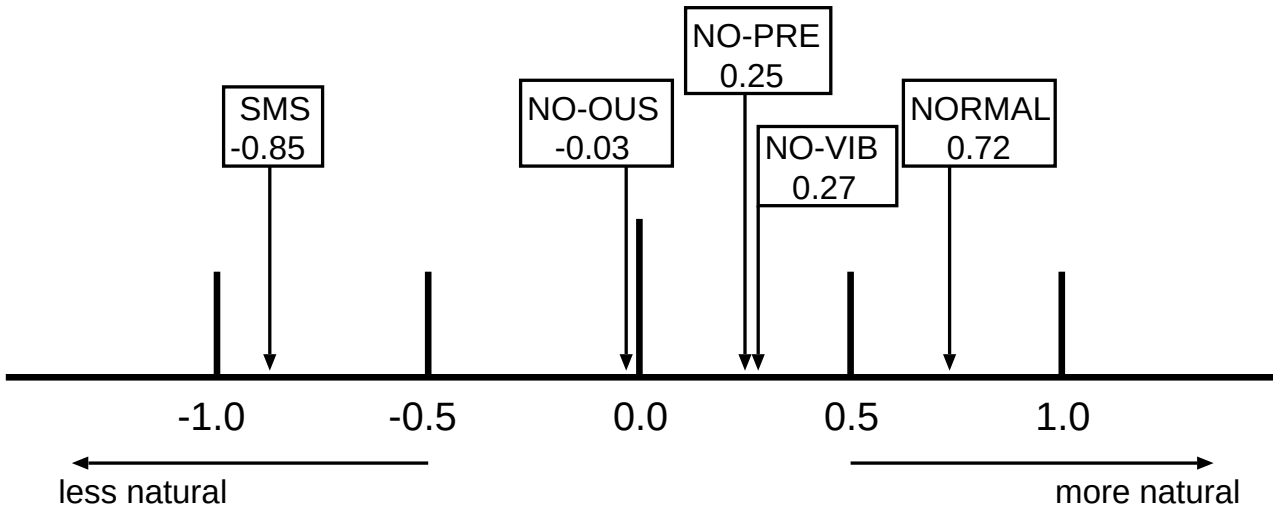

図 2.8: 歌声刺激の自然性の関係

への母数の減少を見ると、本研究で着目している歌声の F0 動的変動成分の歌声知覚への影響の大きさが分かる。過去の研究で課題として残っていたオーバーシュート・アンダーシュート、さらにはヴィブラート・微細変動 の歌声知覚に与える影響も明確になり、やはりこれらの成分が歌声の特徴的な動的変動として重要であることが言える。特にオーバーシュート・アンダーシュートに関しては、ポルタメント成分を残し、目標音高値を振り切った変動であるプロジェクション成分を除去しただけで、他の二より大きな影響を歌声知覚に与えており、歌声の自然性において必要不可欠な成分であることが分かる。また、本研究で新たに F0 動的変動成分として着目した予備的変動に関しても、ヴィブラート・微細変動と同程度の影響をもたらしていることが確認でき、この “予備的変動” も歌声の F0 動的変動成分として今後も取り扱っていく必要があると言える。

2.4　まとめ

話声には無い歌声の特性の中で、動的な特性である F0 動的変動成分を取り上げ、そのなかで３つの成分に着目して研究を進めた。まず歌声データから推定した F0 を基に、各 F0 動的変動成分を操作した歌声を作成した。次にその歌声刺激を用いた心理物理実験を行い、各 F0 動的変動成分の歌声知覚に与える影響を調べた結果、着目した成分すべてが、歌声知覚に大きな影響を与えることを確認した。よって、本研究において、最初に提示した成分である

- **オーバーシュート／アンダーシュート**
  ポルタメント
  オーバーターゲット・アンダーターゲット
- **ヴィブラート／微細変動**
- **予備的変動**

を改めて歌声に特化した F0 動的変動成分として定義し、次章から述べる歌声の F0 制御モデルの構築において扱っていく。

第3章　歌声のF0制御モデルと歌声合成

3.1　目的

高品質な歌声合成を実用的なものにするためにも、歌声に対応したF0制御モデルの構築が必要とされている。そこでまず現在提案されている話声に対応したF0制御モデルでは歌声のF0を生成できない事を示すことで、歌声に対応したF0制御モデルの必要性を示す。そして、前章で示した歌声知覚に影響を与えるF0動的変動成分を制御できるF0制御モデルを提案し、歌声合成に応用する事で、提案したF0制御モデルの検証を行う。最初にKlatt Formant Synthesizerで歌声合成を行い、心理物理実験を通じてF0制御モデル及び生成されたF0の評価を行う。最後にSTRAIGHTによる高精度な歌声合成に応用する。

3.2　話声のF0制御モデル

本研究の背景でも示したように、代表的な話声に対応したF0制御モデルとして、藤崎モデル[5]がある。このモデルの枠組みとして、話声においては、1つの単語のアクセントを表す基本周波数の標準的なパターンは、その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）に、アクセントの存在を示す強調による部分的上昇が重畳されたものと見なして、F0パターンを以下の3つの成分を対数基本周波数軸上で足し合わせたものとして考えている。

- 主に統語情報を表し、比較的ゆっくりした下降を示すフレーズ成分
- 主に辞書的情報を表し、比較的急速な上昇、下降を表すアクセント成分
- 話者にある程度固有の基底基本周波数

これらの成分のうち、上の2つはそれぞれ臨界制動2次系のステップ応答、インパルス応答によって表現され、各応答に対する入力はそれぞれ、アクセント指令、フレーズ指令と呼ばれる単位インパルス列と矩形パルス列である。このモデルの概要を図3.1に示す。

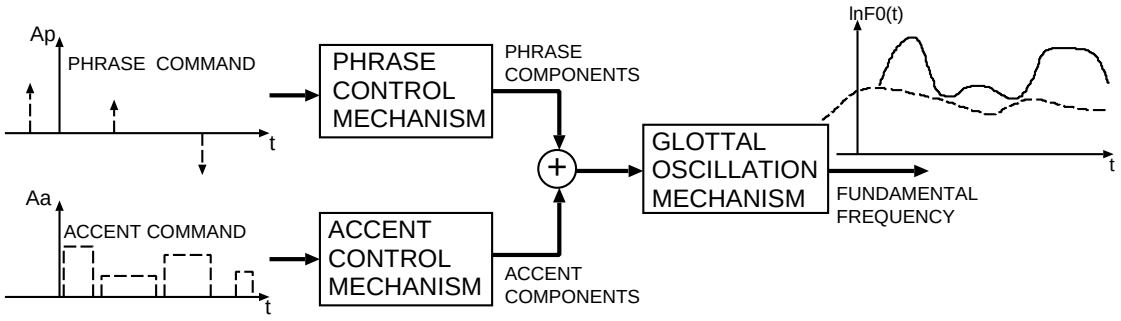

図 3.1: 藤崎 F0 モデルの概要

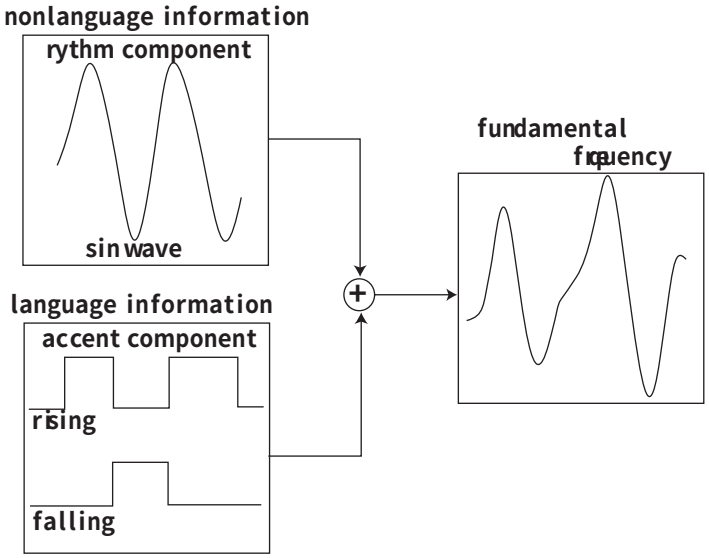

図 3.2: 森山らのモデルの概要

このモデルは、アクセント指令とフレーズ指令を与えるためのパラメータの数が比較的少ない。また、アクセント成分、フレーズ成分はそれぞれ人間の喉頭における甲状軟骨の回転と甲状軟骨の前後移動に対応するなど、生理学見地から理論と対応がとれているといった利点を持つ事から、現在の音声合成技術等において幅広く用いられている。しかし問題点として、主に F0 の上昇の制御しか考えていない事と臨界制動２次系で２つの成分を記述していることによって、急峻な F0 変化を制御・記述することができないことが挙げられる。そこで、森山、山下、天白 [6, 7] らが F0 の下降を組み込んだ F0 制御モデルや、藤崎 F0 モデルのフレーズ成分を、正弦波的変動の非言語情報成分に変える事で、大阪方言特有のリズムを含んだ F0 を制御できるモデル（図 3.2 参照）を提案している。しかし、どの F0 制御モデルも話声を対象としたものであり、前章で示した動的変動成分を

含んだ歌声の F0 を制御・生成するのは困難である。よって、歌声合成システムの実現を考えた場合、歌声に対応した F0 制御モデルが必要となる。

3.3 歌声の F0 制御モデル

前章では、歌声から F0 動的変動成分を除去することで、各成分の歌声知覚に与える影響を明らかにした。本章では、前章と反対のプロセスを考え、メロディラインに各 F0 動的変動成分を付加することで、歌声の F0 変化を制御・記述できる F0 制御モデルの構築を行った。

3.3.1 歌声の F0 制御モデルの構築

本研究では、以下の三つ条件を満たすようなシステムを考える事で、歌声の F0 制御モデルの構築を試みた。

- **歌声知覚に影響を与える F0 動的変動を、F0 変化中に組み込む制御を行える事。**
- **F0 制御モデルとして 2 次系のシステムで構築することで、少数のパラメータで正確な制御を行える事。**
- **歌声合成への応用が可能な事。**

この条件を満たすために、メロディ変化を入力として、それぞれの F0 動的変動成分を付加するシステムとして体系化することで、各変動成分を独立に制御できるモデルを提案した。モデルの概要を図 3.3 に示す。

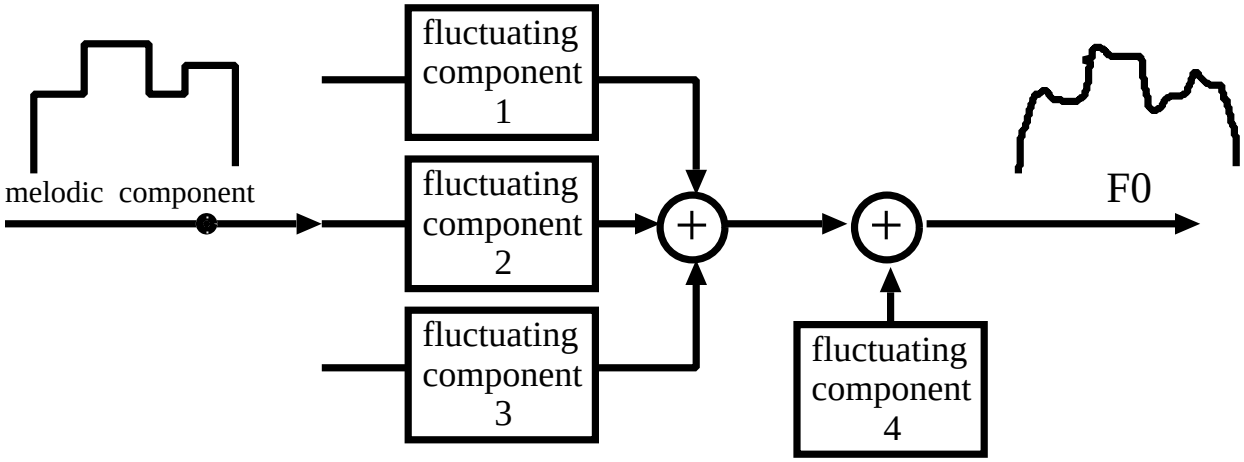

図 3.3: 本研究で提案した F0 制御モデルの概要

このモデルは、入力としてメロディ成分 (melodic component) を、これに F0 動的変動成分を付加するシステムとして歌声変動成分 (fluctuating component) 1 から 4 を組み込むことで歌声の F0 変化を制御・記述できるものである。ここで、モデルにおける各成分は、以下の様なシステムをとる。

- **メロディ成分 (melodic component)**
  曲の旋律概形成分を表す。矩形パルスで記述され、システムの入力となる。
- **歌声変動成分 1 (fluctuating component 1)**
  オーバーシュート・アンダーシュートを制御する成分。制動 2 次系のインパルス応答（減衰振動）で表され、音程変化におけるポルタメント、及びその後の目標音高を振り切る振動のプロジェクションを付加する。
- **歌声変動成分 2 (fluctuating component 2)**
  ヴィブラートを制御する成分。制動 2 次系のインパルス応答（定常振動）で表される。一定の音高が持続する際に駆動する。
- **歌声変動成分 3 (fluctuating component 3)**
  予備的変化を制御する成分。制動 2 次系のインパルス応答（減衰振動）で表される。ただし時間方向が歌声変動成分 1 に対して逆の特性。
- **歌声変動成分 4 (fluctuating component 4)**
  微細変動を制御する成分。擬似的に作成した 10 kHz 以下の不規則変動で表され、F0 全体に付加される。

各成分は、ある決まった区間において駆動するものであり、その区間は歌声変動成分 1 から 4 は順に音程変化直前から直後にかけての間、音程定常時、音程変化直前から直後にかけての間となっていて、それぞれ時変モデルの成分と言える。また歌声変動成分 4 に関しては歌唱全区間において制御を行う成分である。

このモデルの入力となるメロディ成分は、目的の歌のメロディ変化を矩形的な変化で表したものであり、その一例を図 3.4 に示す 。歌声変動成分 4 は、白色雑音をローパスフィルタに通す事で擬似的に作り出した微細変動成分であり、その一例を図 3.4 に示す。また、歌声変動成分 1 、 3 および 2 の一例を図 3.5 に示すが、これらの詳細については次項にて述べる。

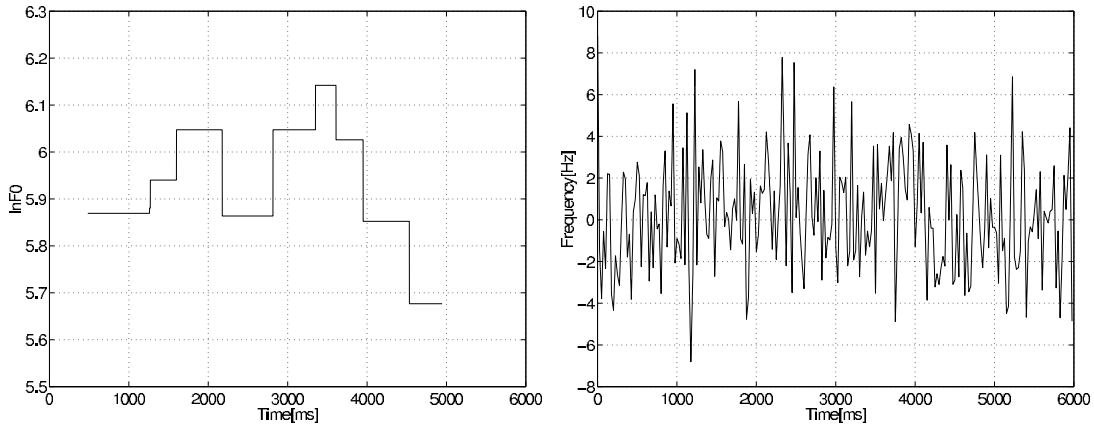

図　3.4:　　右：F0 制御モデルにおけるメロディ成分 (矩形パルスによる表現). 左：歌声変動成分４ (白色雑音から作られる微細変動)

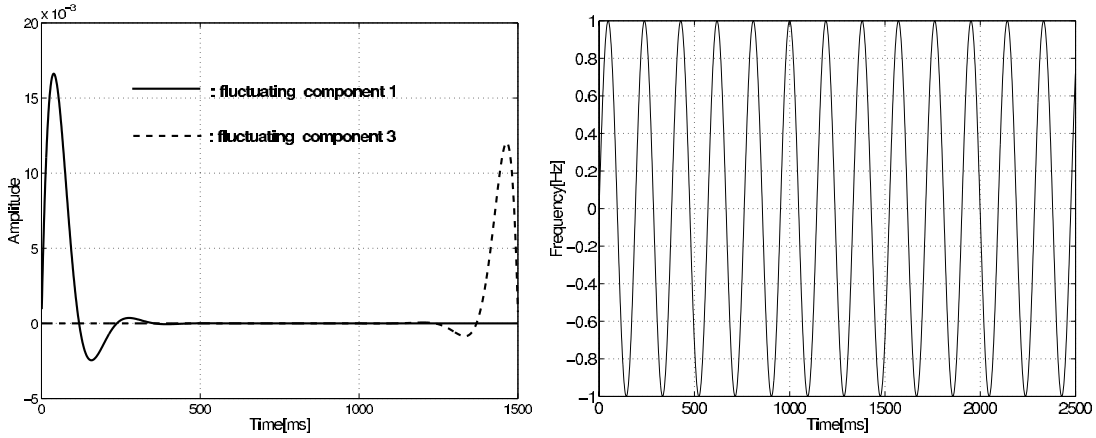

図 3.5: 右：歌声変動成分１, ３ (減衰振動). 左：歌声変動成分２ (定常振動)

3.3.2 F0 制御モデルによる最適な F0 制御

前項で示したように、歌声変動成分 1 から 3 は、それぞれ制動 2 次系のインパルス応答で表され、その時の制動 2 次伝達関数は以下の式で表される。

$$H(s) = \frac{\Omega}{s^2 + 2\zeta\Omega s + \Omega^2}$$

ここで、ζ は減衰振動、Ω は固有周波数値を表し、このシステムのインパルス応答は ζ の値によって以下の様に場合分けされる。

- $|\zeta| > 1$ のとき
  指数減衰:$h(t) = \frac{\Omega}{2\sqrt{\zeta^2-1}}\left(e^{\lambda_1\Omega t} - e^{\lambda_2\Omega t}\right)$, 但し、$\lambda_{1,2} = -\zeta \pm \sqrt{\zeta^2 - 1}$
- $|\zeta| < 1$ のとき
  減衰振動:$h(t) = \frac{\Omega}{\sqrt{1-\zeta^2}} e^{-\zeta\Omega t} \sin(\sqrt{1-\zeta^2}\Omega t)$
- $|\zeta| = 1$ のとき
  臨界制動:$h(t) = \Omega t e^{-\Omega t}$
- $|\zeta| = 0$ のとき
  定常振動:$\sin \Omega t$

ここで示す通り、各歌声変動成分はパラメータ ζ と Ω を決定する事によりそれぞれの特性を持つインパルス応答を与ることができる。よって、本 F0 制御モデルによる最適な F0 制御を考えた場合、F0 動的変動成分をインパルス応答で制御する歌声変動成分 1 から 3 に関して、それぞれ最適なパラメータ ζ と Ω を決定する事が必要となる。

パラメータの最適化

最適なパラメータを決定するために、ζ と Ω を変化させて制御した F0 の、歌声データから TEMPO によって推定した F0 に対するフィッティングを平均二乗誤差推定法 [11] を用いて行った。この方法は、あるパラメータを与えることで歌声変動成分によって制御・付加される動的変動区間における平均二乗誤差を計算し、その誤差を最小とする ζ と Ω を

最適なパラメータとして導き出すものである。その時の平均二乗誤差 E を求める式は以下の通りである。

$$\mathrm{E} = \sqrt{\frac{1}{P}\sum_{t=T}^{T+P}(x(t)-y(t))^2}$$

ここで、$x(t)$ は F0 制御モデルによって生成された F0 を、$y(t)$ は歌声データから TEMPO 推定した F0 を示す。また各歌声変動成分において、フィッティングの為の平均二乗誤差を測定する開始時刻 T[ms] およびその区間 P[ms] は以下の通りである。

歌声変動成分１　：音程変化直前から 250 ms の区間

歌声変動成分２　：音程が一定時間持続する区間

歌声変動成分３　：音程変化直後の時間から時間逆方向に 200 ms の区間

このフィッティング計算を、12 個のすべてのデータに対して ζ と Ω を 0.001 刻み変化させて行い、以下の表 3.1 に示す各歌声変動成分における最適な制御パラメータを求めた。こ

表 3.1: 各歌声変動成分の最適パラメータ

歌声変動成分	Ω[rad/ms]	ζ
歌声変動成分１	0.031	0.52
歌声変動成分２	0.033	0
歌声変動成分３	0.028	0.72

こで、歌声変動成分１及び３に関しては、図 3.5 で示したように共に減衰振動で表される。この時の Ω は主に変化の持続時間に影響し、ζ は変化するスピードと振動の大きさに影響を与えるパラメータである。また、ここで決定された最適なパラメータによって制御される変動は、歌声変動成分１に関しては音程変化における緩やかな傾き（ポルタメント）と、その直後に音程変化量の約 15% 程度の大きさで 250〜350 ms かけて減衰振動するプロジェクション成分である。また、歌声変動成分３に関しては、音程変化量の約 13% 程度の大きさを持った予備的変動が音程変化直前に制御される。

次に歌声変動成分２に関しては、図 3.5 で示した定常振動で表される。この時の Ω は定常振動の周期を決定するパラメータであり、最適なパラメータによって付加される変動は、周期 190ms 程度の定常変動であるヴィブラート成分である。

以上の最適化されたパラメータを用いて、F0 制御モデルによって生成された F0 を実音声から抽出した F0 と並べて図 3.6〜9 に示す。モデルで記述された F0 を見ると、前章で示した F0 動的変動成分が制御されていることが視覚的にも確認でき、実音声の F0 と比較しても、非常に近似した F0 変化パターンが制御・記述されていることが分かる。また、このモデルにおける必要な制御パラメータは、２次系で表される各歌声変動成分に対して与える Ω と ζ のみと非常に少ない。これより、本節の頭で述べた F0 制御モデルを構築する上での条件の内、最初の二つを満たしたモデルを構築することができたと言える。しかし、残り一つの条件である歌声合成への応用に関しては、この結果から考察することはできない。そこで次節からは F0 制御モデルを歌声合成に応用することで、モデル及びモデルで生成される F0 の評価を行い、更には高品質な歌声合成に応用できるか検討をする。

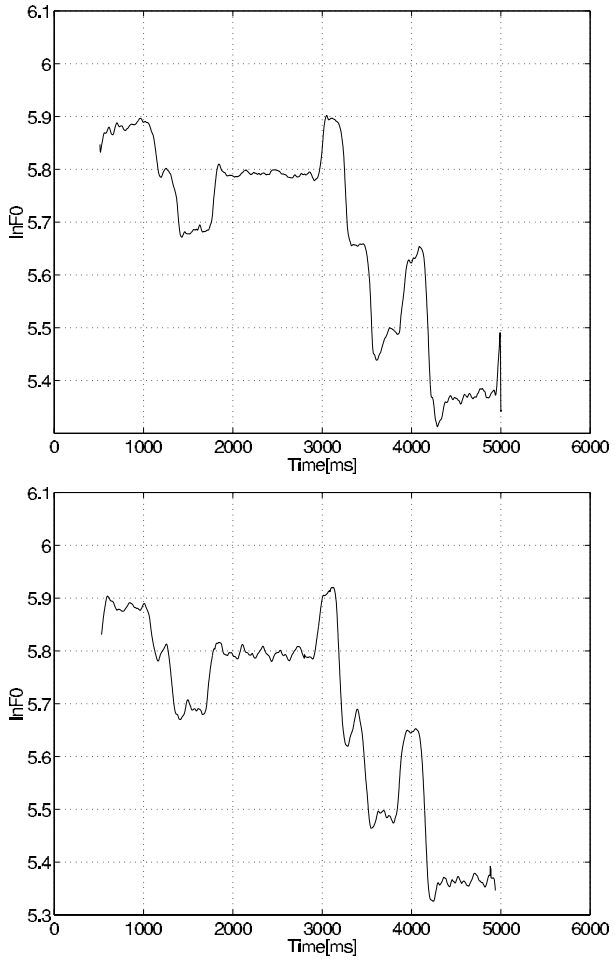

図 3.6: 「カラスなぜなくの」歌唱時の、上：実音声の F0, 下：モデルで生成した F0

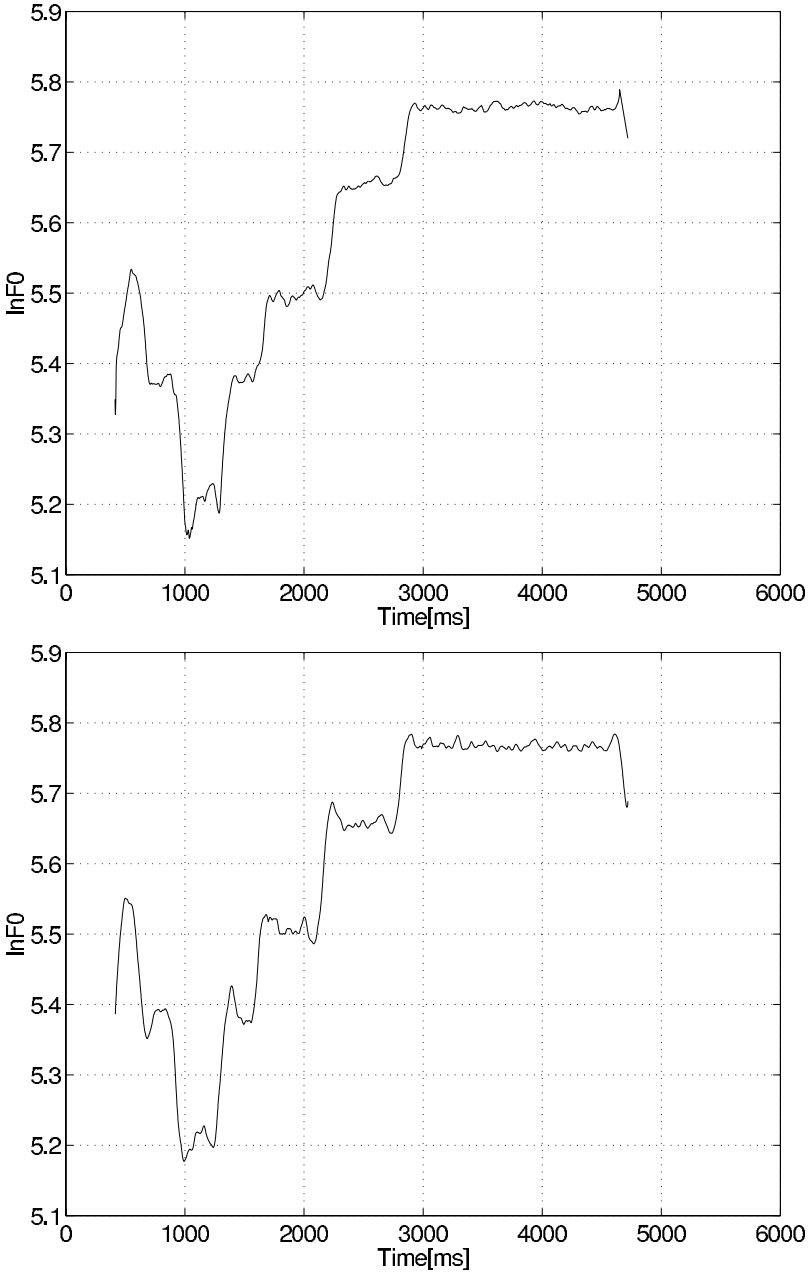

図 3.7: 「カラスは山に」歌唱時の、上：実音声の F0, 下：モデルで生成した F0

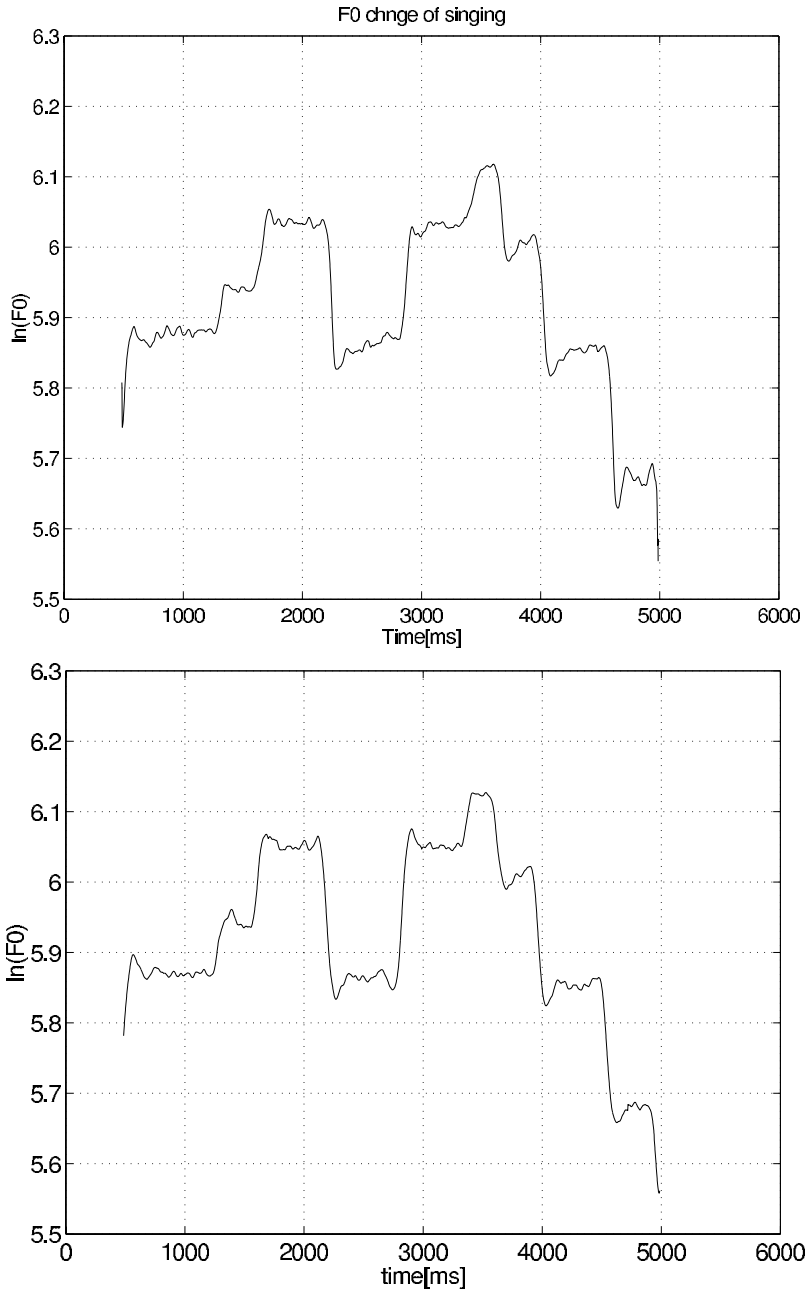

図 3.8: 「かわいい七つの」歌唱時の、上：実音声の F0, 下：モデルで生成した F0

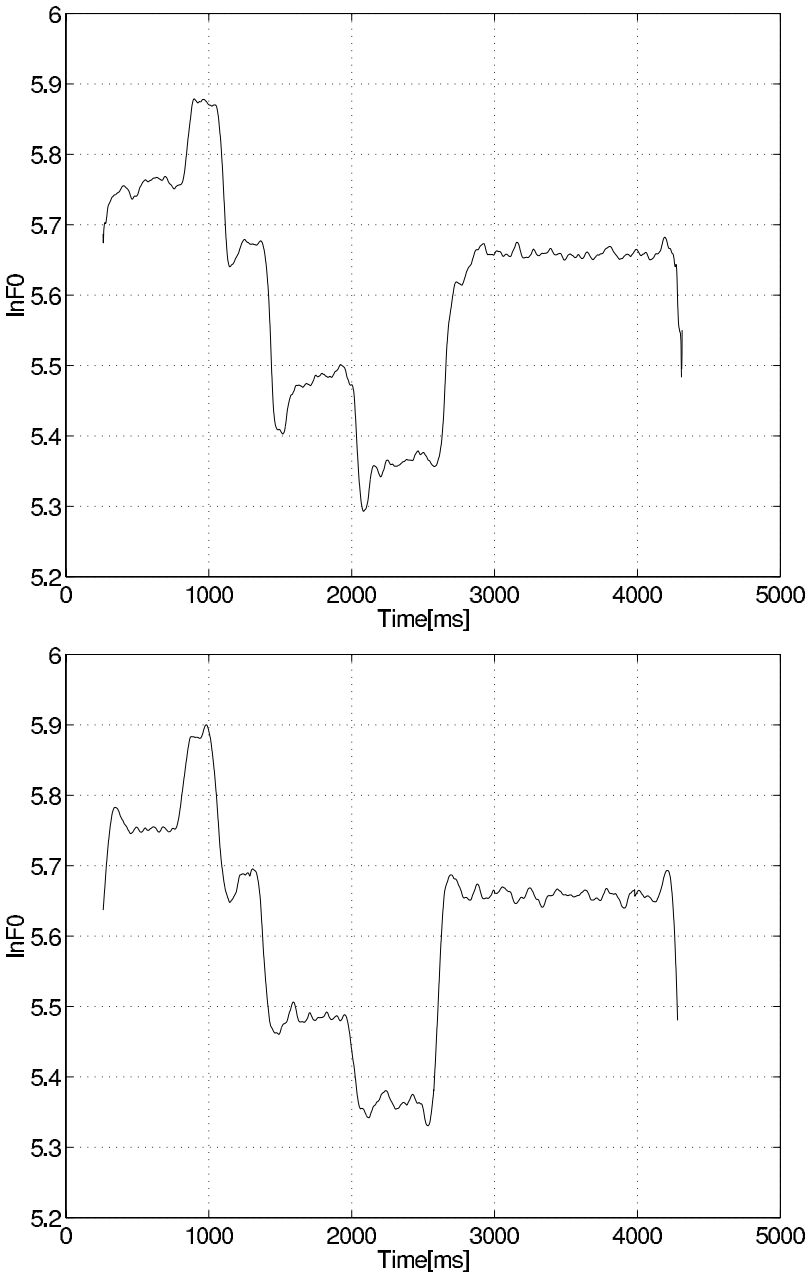

図 3.9: 「子があるからよ」歌唱時の、上：実音声の F0, 下：モデルで生成した F0

3.4 F0 制御モデルの歌声合成への応用１

前節では提案した F0 制御モデルが、F0 動的変動成分を制御することで歌声の F0 変化を生成できる事を示した。しかし、これは生成された F0 変化パターンのみを対象とし評価、及び考察した結果であり、F0 制御モデルの歌声合成への応用可能性を直接示すものではない。そこで本節では、実際に F0 制御モデルを用いた歌声合成を行い、心理物理実験によって F0 制御モデル、及びモデルで生成される F0 の評価を行う。

3.4.1 Klatt Formant Synthesizer による歌声合成

本研究で提案した歌声の F0 制御モデルを Klatt Formant Synthsizer を用いた歌声合成に適用した。合成手順を図 3.10 に示す。

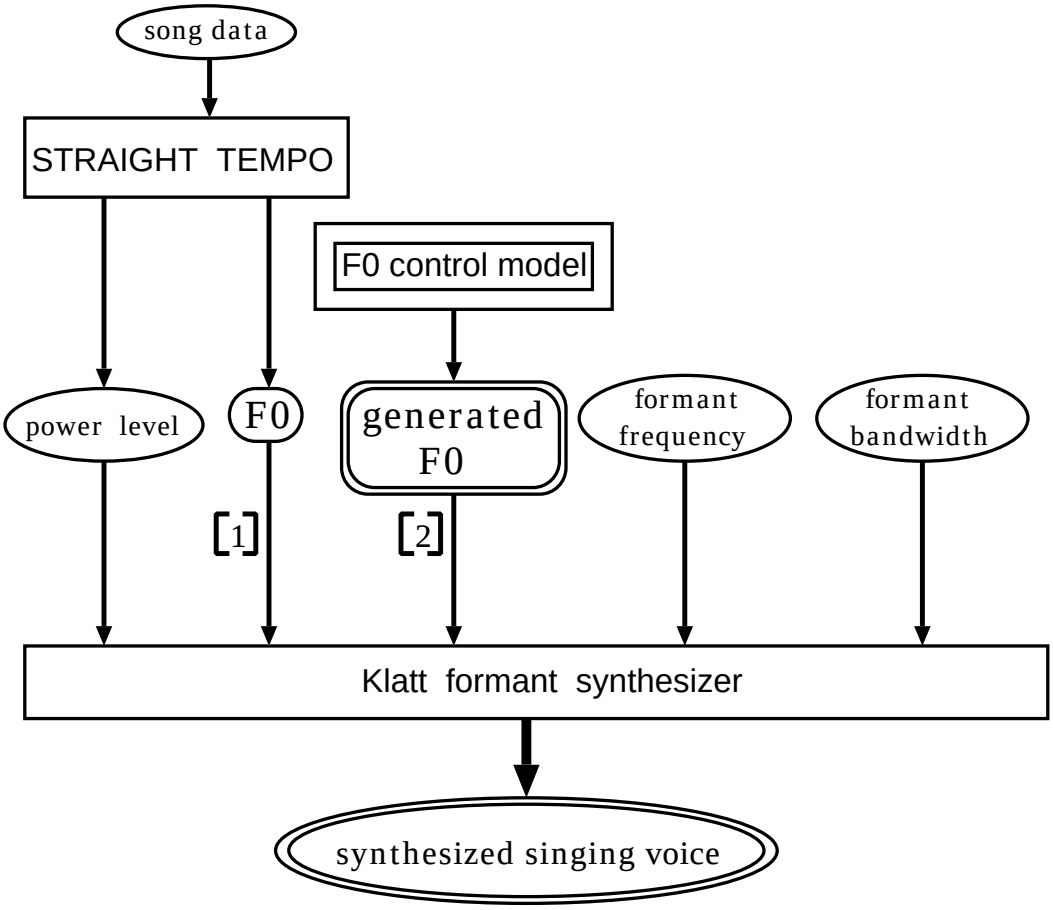

図 3.10: 歌声合成の手順（Klatt Formant Synthesizer）

実音声からTEMPOによって抽出したF0、及びF0制御モデルで制御したF0をそれぞれ用いてKlatt Formant Synthesizerによって歌声合成を行う方法であり、合成の際のパワーレベルは80 dBで一定、ホルマント周波数、及びバンド幅の設定は２章で示した値と同じである。作成した合成音は、図3.10上で[1]及び[2]の手順で作成した以下の６つである。

- **NORMAL**
  歌声データからTEMPOによって抽出したF0を用いて、Klatt Formant Synthesizerによって合成したもの。（図3.10における[1]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F0を用いた合成音。（図3.10における[２]の行程）
- **SYN-OUS**
  F0制御モデルによってオーバーシュート・アンダーシュート成分のみ制御したF0を用いた合成音。（図3.10における[2]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F0を用いた合成音。（図3.10における[2]の行程）
- **SYN-PRE**
  F0制御モデルによって予備的変化成分のみ制御したF0を用いた合成音。（図3.10における[2]の行程）
- **SYN-BASE**
  F0制御モデルの入力成分であるメロディ成分のみ用いた合成音。（図3.10における[2]の行程）

これら合成音のF0を図3.11に示す。

今回、合成器としてKlatt Formant Synthesizerを用いた理由は、２章で示した合成音作成時と同様、F0を独立に扱い合成できるからである。また、作成された合成音の音質がF0変化に大きく影響を受けるので、F0制御モデルで生成したF0の評価するのには最適であると考えた。

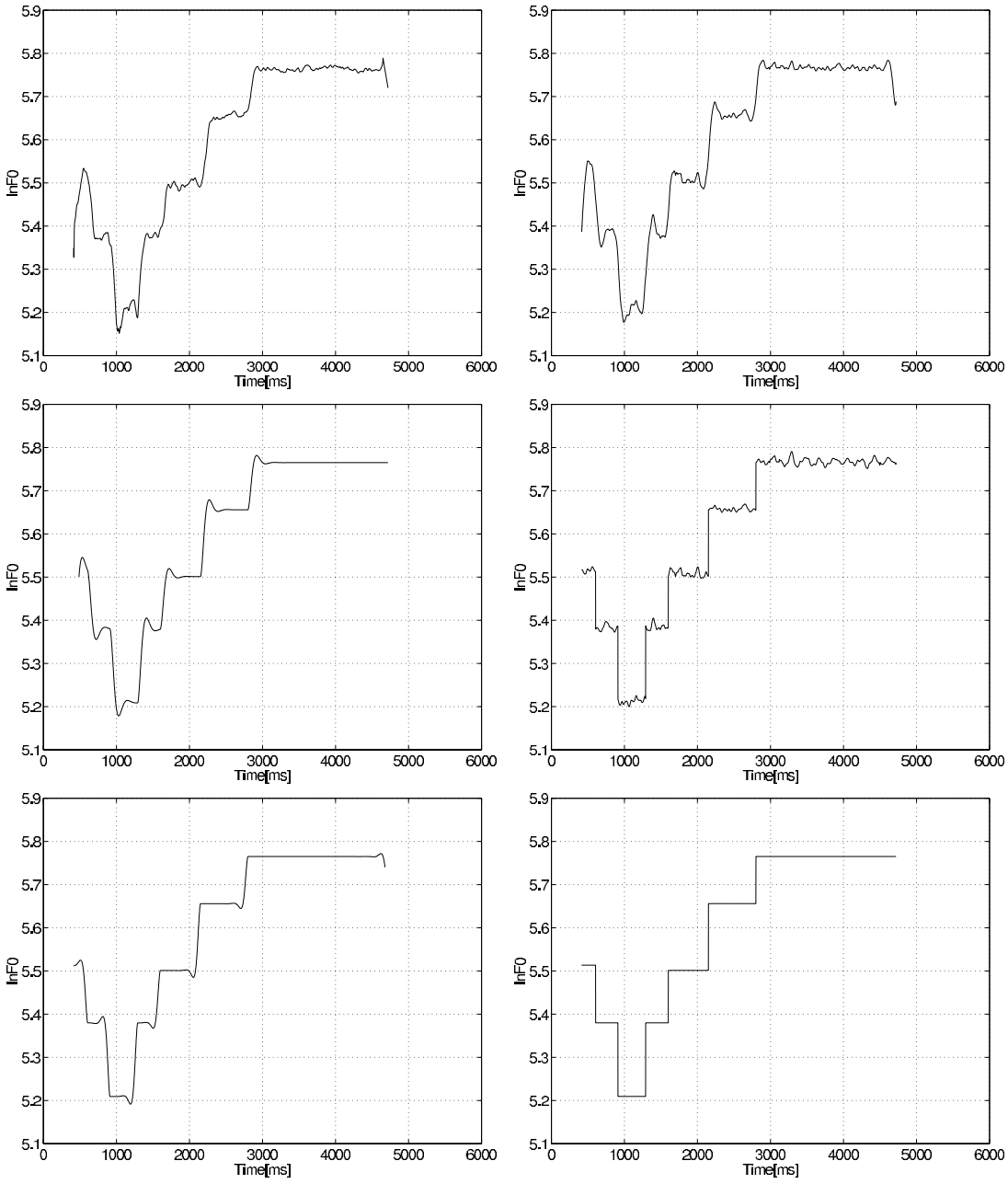

図 3.11: 歌声合成に用いた F0. 上段の左から順に NORMAL, SYN-ALL. 中段 SYN-OUS, SYN-VB. 下段 SYN-PRE, SYN-BASE の F0 を示す

3.4.2 心理物理実験

前項で示した合成音を用いて心理物理実験を行う事で、F0 制御モデルの評価を行い、歌声合成への応用可能性について検討する。

心理物理実験の方法は前回同様シェッフェの一対比較実験を採用し、評価段階も 7 段階で行った。また刺激条件として、６つの合成音から作られる 30 対を、一人の歌手分作成した計 120 対を用いた。実験条件と手続きに関しては、被験者を大学院生 7 名とした以外はすべて前回と同じである。

3.4.3 実験結果と考察

心理物理実験で得られたデータを、前回と同じ方法によって処理した結果を表 3.2 に示す。また、刺激の母数の値に従って、６つの刺激の心理距離関係を直線上で示すと、図 3.12 が得られる。

表 3.2: 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-OUS	0.56
SYN-VB	-0.73
SYN-PRE	-0.06
SYN-BASE	-1.15

この結果から、F0 制御モデルによってすべての制御を行った合成音 SYN-ALL が、実音声の F0 を用いた合成音 NORMAL より若干であるが自然性の高い歌声として被験者が判断していることが分かる。この原因は、歌声変動成分２によって制御されるヴィブラートの影響によるところが大きいと考えられる。F0 制御モデルによって付加されるヴィブラートは定常振動であり、音程安定時の声に綺麗な張りを与える。一方実音声の F0 にも若干のヴィブラートが観測されるが、モデルの F0 に比べると不規則的な振動であり、その分声に張りも生まれないため、被験者の多くはヴィブラートの違いを聴き分けて SYN-ALL の方がより自然な歌声として判断したと考えられる。

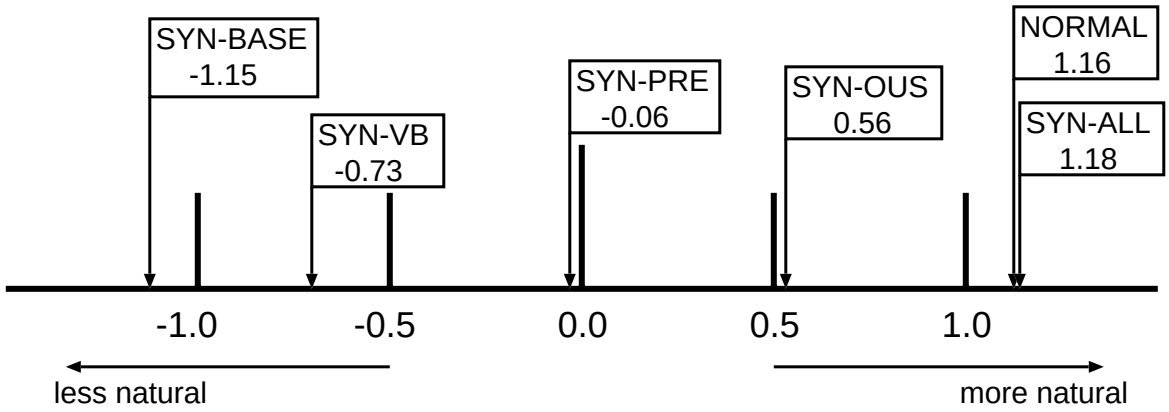

図 3.12: 合成音の自然性の関係

また、一つの歌声変動成分によってF0制御を加えた合成音SYN-OUS、SYN-VB、SYN-PRE、SYN-BASEが、メロディ成分のみの合成音SYN-BASEと比較してどれも自然性が向上しているのが確認できる。特に歌声変動成分１によってオーバーシュート・アンダーシュートのみを制御することによる自然性の向上は、他と比較して顕著である。これに関する考察は次項で詳細に述べる。

以上の事から、メロディ変化に歌声知覚に影響を与えるF0動的変動成分を付与する事で、歌声の自然性は向上することが分かり、本研究で提案したF0制御モデル、及びモデルによって生成されたF0の有効性が確認できる。また、今回はスペクトルの変化は一定にしているが、スペクトルの変化も考慮した歌声合成にF0制御モデルを適用することで、高品質な歌声を生成できる可能性も示唆された。

3.4.4 ポルタメントに関する考察

前項での考察において、F0制御モデルの歌声変動成分１によってオーバーシュート・アンダーシュートを制御することで、合成音の自然性が大きく向上することを述べた。この自然性の向上は、２章で示したプロジェクションの除去による歌声の大きな自然性劣化の結果と対応の取れるものであるが、ここで一つ疑問が生じる。それはオーバーシュート・アンダーシュートを構成する成分の一つであるポルタメント成分が、SYN-OUSの自然性向上にどの程度の影響を与えているのかということである。

ポルタメントとは、声楽・弦楽器等で音を別の高さに移動するときに非常に滑らかに移動するような歌唱、もしくは演奏することを言い、midiやシンセサイザーのようなデジ

タル楽音の演奏においては効果的で重要なものであるとされている。しかし、２章で行った実験では、ポルタメントの影響のみ歌声から除去することが困難であったため、この成分の歌声知覚に与える影響を明確にすることができなかった。本研究における F0 制御モデルでは、曲の旋律概形を表すメロディ成分に歌声変動成分１によって、オーバーシュート・アンダーシュートを付加している。その制御において、音程変化における目標音高値を振り切る変動成分のプロジェクションの特性が付加されると同時に、ステップ状の変化だった F0 変化は滑らかな傾きを持つことになる。この滑らかな傾きがポルタメント成分である。

そこで本項では、この成分のみF0 変化に組み込んだ合成音を作成し、聴取実験を行うことでの歌声知覚に与える影響を調べた。合成音作成は前項同様で Klatt Formant Synthesizer によって以下の５つの合成音を作成した。

- **POL-OUS**
  F0 制御モデルにおいて、メロディ成分に歌声変動成分１によってオーバーシュート・アンダーシュートを付加した F0 を用いた合成音。
- **POL1**
  POL-OUS からプロジェクション成分を除去することで、ポルタメント成分のみ付加された F0 を用いた合成音。
- **POL2**
  POL1 よりも急峻な変化のポルタメント成分を持つ F0 を用いた合成音。
- **OUS**
  POL-OUS からポルタメント成分を除去することで、プロジェクション成分のみ付加された F0 を用いた合成音。
- **BASE**
  メロディ成分のみを用いた合成音。

ここで、POL-OUS の F0 に関しては、前節で示した最適なパラメータを用いて歌声変動成分１によって制御されたものである。また POL1、POL2、OUS の F0 変化は図 3.13 に示す通りである。

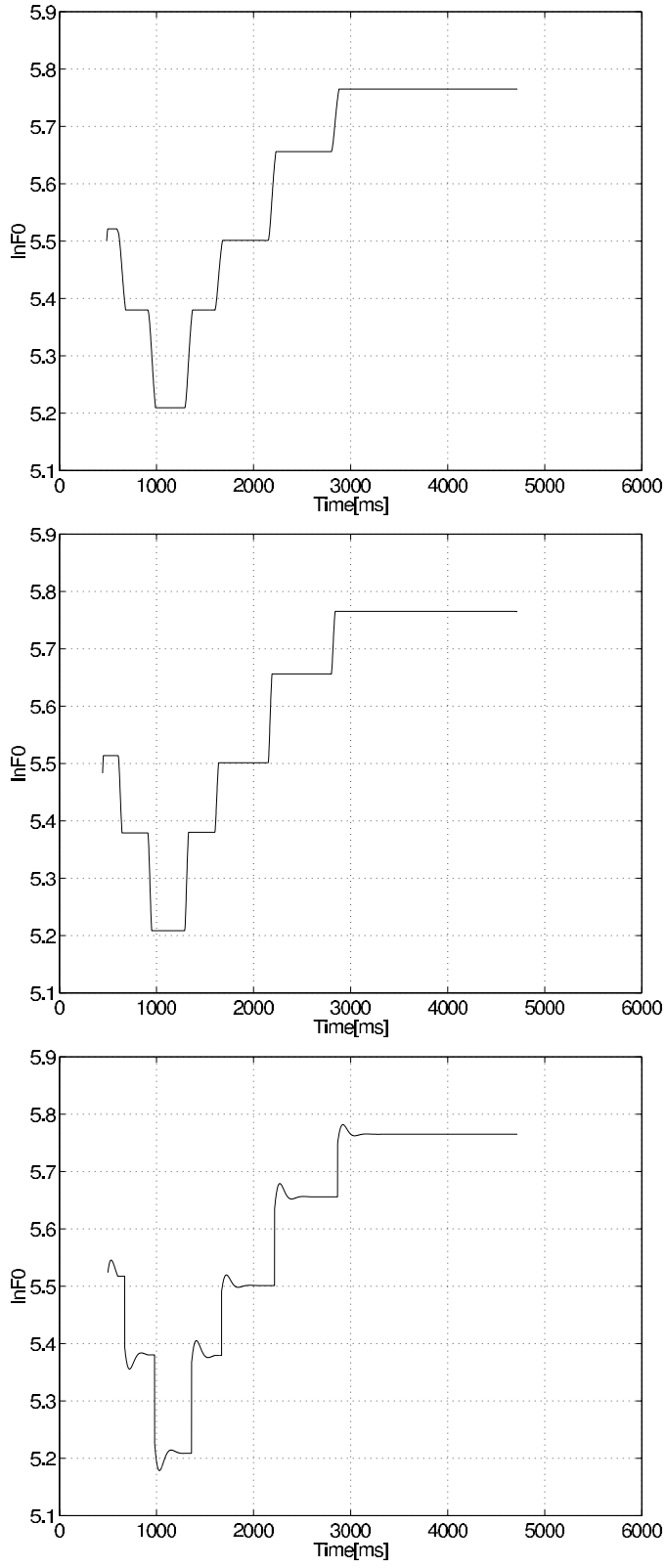

図 3.13: 合成音の F0. 上から POLTA1、POLTA2、OUS の F0 を表す

これらの合成音を刺激として、心理物理実験を行った。実験方法、条件、及び手続きは前回と同様である。実験によって得られた刺激の母数の値を表 2.3 に、その母数に従って刺激の心理距離の関係を直線上に示したものを図 3.14 に示す。

表 3.3: 母数の推定 (自然性). オーバーシュート・アンダーシュートに関して

歌声	母数
POL-OUS	1.36
POLTA1	0.65
POLTA2	0.27
OUS	-0.49
BASE	-1.28

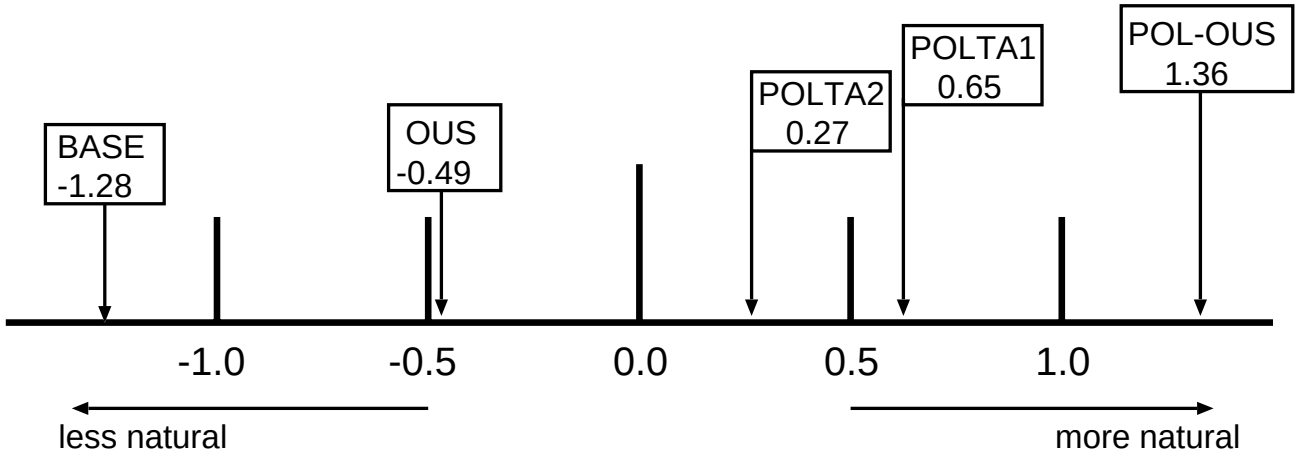

図 3.14: 歌声刺激の自然性の関係

この結果から、ポルタメント成分のみを含んだ合成音 POLTA1 と POLTA2 がメロディ成分のみの BASE と比較して高い自然性を持つことが分かる。また、同じポルタメント成分を含んだ合成音同士を比較すると、F0 制御において最適なパラメータによって付加された比較的緩やかな変化のポルタメントの方が、歌声により自然性を与えている事が確認できる。

次にプロジェクション成分に関しては、OUS の母数を見るとそれほど歌声に自然性を与えていないが、POLTA1 と POL-OUS を比較した場合には大きな自然性を与える要素

となっている。これはプロジェクション成分の生成過程に原因がある。この両成分は音程変化時の傾きをもった上昇、あるいは下降（ポルタメント）の後に生じる目標音高値を振り切る変動である。その為、ポルタメントの無い急峻なF0変化に付加されてもそれほど影響を与えず、逆にポルタメントのあるF0変化に付加されることで歌声に大きな自然性をもたらすと考えられ、この事は２章で示したプロジェクション成分の歌声知覚に与える影響の結果と対応がとれる。

以上の考察から、歌声のF0動的変動であるオーバーシュート・アンダーシュートにおけるポルタメント成分の重要性が分かり、歌声知覚に影響を与える事を確認できた。また、ポルタメント後の変動成分のプロジェクションの影響も改めて確認することができた。

3.4.5　この節のまとめ

本節では、前節で示したF0制御モデルをKlatt Formant Synthesizerを用いた歌声合成に適用し、心理物理実験によってF0制御モデル、及び生成されるF0の評価を行った。その結果、モデルによって生成されたF0を用いた合成音が非常に高い自然性を有していることが分かり、F0制御モデルの歌声合成への応用可能性を確かめることができた。しかし、問題点として、歌声合成におけるスペクトル情報を一意に決めていたこと、そして合成音と実音声の比較が行えなかった事が挙げられる。そこで次節では、F0制御モデルを高品質なVOCODERであるSTRAIGHTに適用し、スペクトル情報も考慮した高品質な歌声合成を行う事で、実音声に近い歌声を作成する。

3.5　F0 制御モデルの歌声合成への応用２

前節で示した Klatt Formant Synthesizer による歌声合成は、スペクトル情報を一定に固定したものであり、実際の歌声の F0 を用いた合成音でも、実音声と比較するとかなり音質は劣化し、歌い手の個人性情報も失ったものになってしまう。この方法は、前説のような F0 を評価する上では問題は無いが、F0 制御モデルの高品質な歌声合成への応用を考えた場合、スペクトル情報も考慮に入れた合成法が必要となる。そこで、本節では高品質な VOCODER である STRAIGHT を用いる事で、F0 制御モデルの実音声に近い高品質な歌声合成への応用を図る。

3.5.1　STRAIGHT による歌声合成

STRAIGHT は、図 3.15 の点線内に示すように３つ要素から構成され、その構造は標準的な VOCODER そのものである。STRAIGHT-core は音声のスペクトル情報を抽出し、TEMPO は２章で示した通り F0 の抽出を行う分析系である。また SPIKES は合成に用いる駆動音源の群遅延特性を操作することにより、いわゆる VOCODER らしさの音色を軽減させる合成系である。

STRAIGHT を用いた歌声合成の手順は図 3.15 に示す通り、TEMPO で抽出される F0 をモデルで生成した F0 に置き換える事で歌声合成が行われる。また、合成は以下の３つの条件を考慮して行った。

- スペクトルは STRAIGHT-core によって抽出されたものをそのまま使用する。
- F0 を入れ替える際に、TEMPO で抽出した F0 とモデルで生成した F0 との時間に関するズレを極力無くし合成する。
- SPIKES における群遅延操作のパラメータを操作する。

ここで２番目の条件の理由として、合成の際にモデルで生成した F0 に時間的なズレがあると、スペクトル情報とのマッチングが取れない区間が生じ、生成される歌声に大きな歪みを与えてしまうからである。

次に３番目に関してだが、これは STRAIGHT で合成音の品質を支配する要因の、駆動音源の特性に関する事である。STRAIGHT 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、VOCODER で問題になっていた合成音声特有のバズ音の軽減を SPIKES によって行っている。しかし、群遅延操作によ

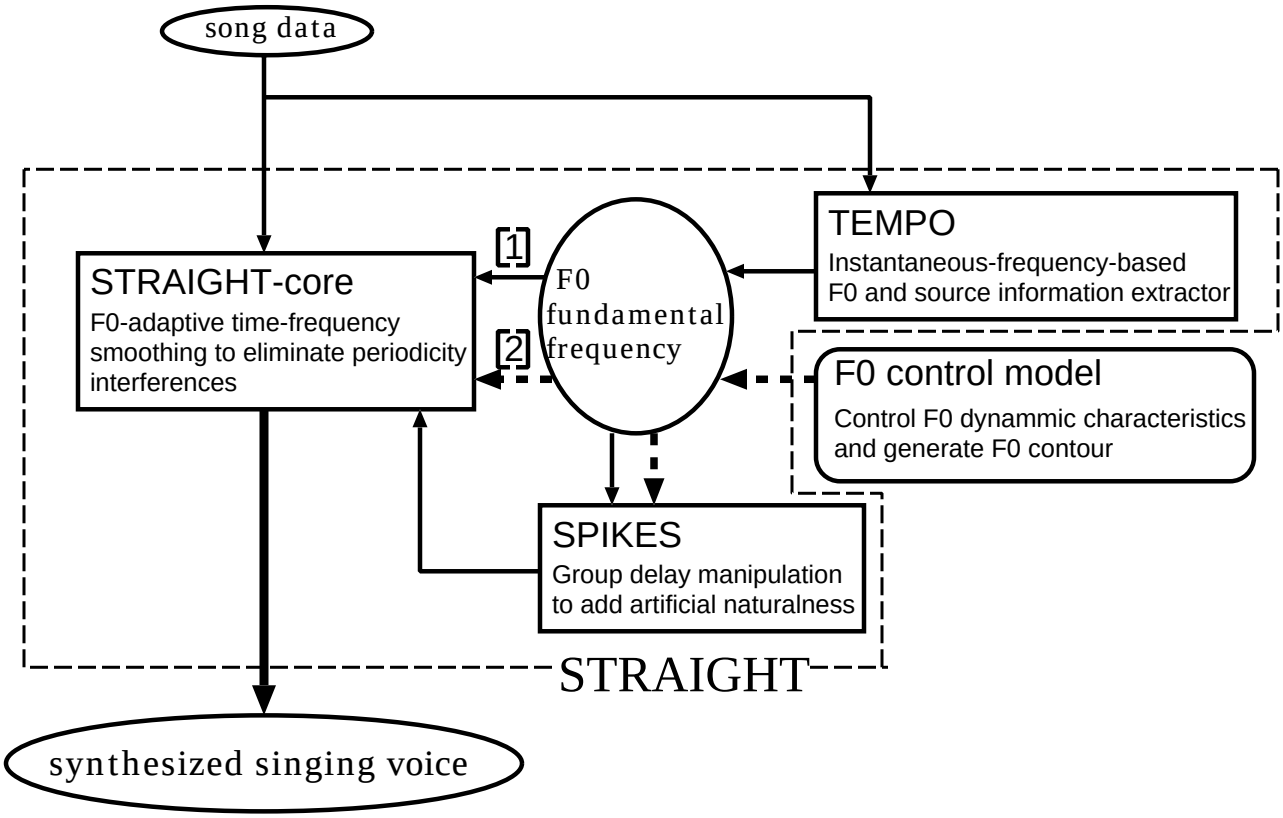

図 3.15: STRAIGHT の構成と歌声合成の手順

るオールパスフィルタの設定は、対象とする音声が例えば女性と男性の違いによって変更する必要がある [12]。これは歌声にも言える事であり、本研究では歌声合成を行う前に、歌声に適応する群遅延操作のパラメータを決定した。設定の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、一つの歌声データに対して、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 3 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り、本研究においては標準偏差 1 ms、境界周波数を 5.5 kHz に設定して歌声合成を行った。以上の条件で作成した歌声は、前回の歌声合成の時と同様の次の６つである。

- **NORMAL**
  STRAIGHT によって分析再合成した合成音。群遅延操作により実音声とほぼ同じ品質を持った歌声。（図 2.15 における [1] 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 F0 を用いた合成音。（図 2.15 における [2] の行程）

- **SYN-OUS**
  F0 制御モデルによってオーバーシュート・アンダーシュート成分のみ制御した F0 を用いた合成音。（図 2.15 における [2] 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 F0 を用いた合成音。（図 2.15 における [2] の行程）

- **SYN-PRE**
  F0 制御モデルによって予備的変化成分のみ制御した F0 を用いた合成音。（図 2.15 における [2] の行程）

- **SYN-BASE**
  F0 制御モデルの入力成分であるメロディ成分のみ用いた合成音。（図 2.15 における [2] の行程）

尚、ここで示す NORMAL は、実音声とほぼ同程度の歌声である。次項では、これらの歌声を評価するために心理物理実験を行った。

3.5.2 心理物理実験

前項での Klatt Formant Synthesizer による歌声合成を行った結果、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 F0 による合成音は、実音声の F0 を用いた合成音と同程度の自然性を持つ事が分かった。しかし、問題点として、実音声と比較することができない事があった。そこで、本項で行った STRAIGHT による歌声合成においても心理物理実験を行い、F0 制御モデルによる F0 を用いた合成音と実音声（NORMAL）を比較することで、再度 F0 制御モデルの評価を行うと共に、歌声合成における F0、スペクトル両情報の重要性に関して考察を行う。尚、今回の実験においても、実験方法、及び条件は前回と同様である。

表 3.4: 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-OUS	0.45
SYN-VB	-0.71
SYN-PRE	-0.23
SYN-BASE	-1.04

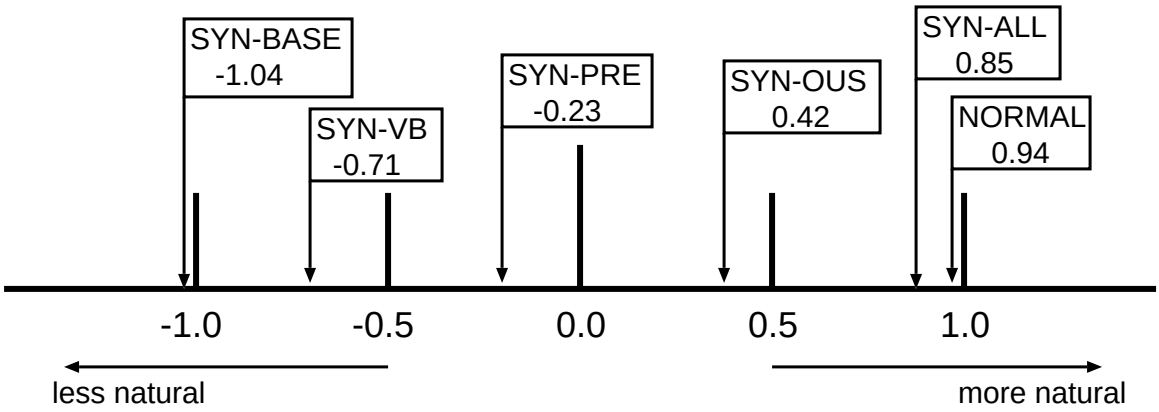

図 3.16: 合成音の自然性の関係

3.5.3 実験結果と考察

実験から得られた各刺激の母数値を表 3.4 に示し、及び図 3.16 にその値にしたがって刺激間の心理距離関係を示す。

結果が示すように、Klatt Formant Synthesizer で歌声合成を行った時とほぼ同じ結果が得られた。しかし今回の実験で用いた合成音は、スペクトル変化も考慮されているので非常に自然性が高いので、歌声 NORMAL は実音声と考えて他の歌声と比較することができる。その結果、F0 制御モデルですべて制御された F0 を用いた合成音 SYN-ALL も、NORMAL に比べ若干自然性が低いと判断された。しかし、その差はごくわずかであり、本 F0 制御モデルは、スペクトル情報も考慮に入れた歌声合成に適用することで非常に高品質な歌声を作り出す事ができると言える。また一般に歌声合成を行う上で、高品質な歌

声を実現するためにはF0の制御以外にスペクトル情報の操作も重要であることが示せた。

3.6 まとめ

本章では、２章で明らかとなった歌声知覚に影響を与えるF0動的変動成分を制御し、F0を生成できるF0制御モデルを構築した。メロディ変化にF0動的変動を付与する２次系システムを考えることで、少ないパラメータで正確なF0制御・生成できることを示した。また、このF0制御モデルを歌声合成に応用した。まずklatt Formant Synthesizerによって歌声合成を行い、心理物理実験によってF0制御モデルの歌声合成への応用可能性を確認した。次にSTRAIGHTを用いた高品質な歌声合成に応用し、実音声と分別しにくい程度の自然性の高い歌声合成が実現きることを示した。

第4章　結論

4.1　本論文で明らかになったことの要約

本研究では歌声のF0に関して、（1）歌声特有のF0動的変動成分に着目したF0分析、（2）歌声のF0制御モデルの構築と歌声合成への応用、について検討してきた。

歌声のF0動的変動成分の分析

話声には無い歌声に特化したF0変動成分の一つであるF0動的変動成分に着目し、それらの歌声知覚に与える影響を調べることで、歌声のF0において重要な動的変動成分の抽出を行った。そのために、STRAIGHTのTEMPOによって得られた高精度なF0から大きく分けて以下の3つのF0動的変動成分に着目した。

- **オーバーシュート／アンダーシュート**
- **ヴィブラート／微細変動**
- **予備的変動**

F0を操作することで各F0動的変動成分を除去した合成音を作成し、心理物理実験を行った結果、3つの動的変動成分すべてが歌声知覚に影響を与える成分であることが分かった。中でも、音程の滑かな変化であるポルタメント成分と、その後に生成される目的音高値を振り切る変動のプロジェクションから構成されるオーバーシュート・アンダーシュートの歌声知覚への影響が顕著であることが確認できた。また、本研究で新たに歌声のF0動的変動成分として着目した予備的変動も歌声のF0において重要な成分であることが分かり、着目したF0動的変動成分すべてが、歌声のF0に特化した重要な成分として定義することができた。

歌声の F0 制御モデルと歌声合成

F0 制御モデルを構築し、高品質な歌声合成への応用を試みた。F0 制御モデルの構築においては、メロディ変化に F0 動的変動成分を制御し付与できるシステムを考えることで、６つという非常に少ないパラメータによって F0 動的変動成分を制御し、F0 を生成する２次系モデルを提案した。この F0 制御モデルは、最適な制御パラメータを設定することによって、実音声の F0 変化パターンに近似した F0 を制御・記述できる有用なモデルであることを確認した。また、この F0 制御モデルの歌声合成への応用に関しては、まず Klatt Formant Synthesizer による歌声合成を行う事で、F0 制御モデル、及びモデルで生成される F0 の評価を行った。その結果、実音声の F0 を用いた合成音と同程度の歌声合成が可能であることが分かり、F0 制御モデルの高品質な歌声合成への応用可能性を確認できた。そして、最後に STRAIGHT によるスペクトル情報も考慮した歌声合成に応用することで、非常に自然性の高い高品質な歌声合成が実現できることを明らかにした。

4.2　今後の課題

以下に今後の課題を列挙する。

- 多くの歌声データに対しての本研究の検証
  今回用いてきた歌声データは、歌唱内容、被験者数共に少ないものである。よって、より多様な歌声データに対して、本研究で明らかとなった分析結果及び F0 制御モデルの検証を行うことが必要である。特に F0 制御モデルに関しては、多くの歌声データに対して適応することで、より有用性が高く頑健なシステムへの発展が重要であり、中でも歌詞のある歌声を対象にした F0 制御モデルの改善は歌声合成システムを多様化させる上でも必要である。

- F0 制御モデルの生理学的モデルへの発展
  今回提案した F0 制御モデルを、歌唱における生成機構のメカニズムに対応した生理学的モデルに発展させることで、歌声合成への応用だけでなく、歌唱学・音楽学、更には歌声生成機構の解明に大きな影響を与える事ができる。

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。

また、日頃から熱心に御討論頂き、有益な御助言を賜りました党建武助教授、ならびに心強いサポートで本研究の遂行を支えてくださった鵜木祐史助手に心より感謝致します。

また、日頃から多大なる議論と激励を頂きました赤木研究室の諸先輩方に厚くお礼申し上げると共に、本研究の遂行の多面に渡りご協力頂いた赤木研究室の皆様に感謝致します。

最後に、大学院での貴重な研究生活を与えて頂き、暖かく見守ってくれた両親、祖母、兄、姉、そして友人に心から感謝しお礼を申し上げます。

関連図書

[1] 中山 一朗, 小林 範子, “歌の声” 日本音響学会誌 52 巻 5 号, pp. 383-388, (1996).

[2] 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.

[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] 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.

[5] H. Fujisaki, M. Tatsumi, “Analysis control in singing” “Vocal fold physiology”, UNIVERSITY OF TOKYO PRESS, pp.347-363, 1981.

[6] T. Moriyama, H.Ogawa, S.Tenpaku, “A new control model based on rising and falling fundamental frequency” Proc. of ASA and ASJ Third Joint Meeting, pp.1171-1176, 1996.

[7] 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.

[8] H. Kawahara *et al.*, “Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity” Proc. Eurospeech99, pp.2781-2784, 1999.

[9] 石本 祐一, “雑音環境における基本周波数推定法とこれを用いた雑音抑圧に関する研究” 北陸先端科学技術大学院大学修士論文, 2000.

[10] 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.

[11] Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. “Numerical Recipes in C,” Cambridge University Press, Cambridge.

[12] 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.

学会発表リスト

齋藤 毅, 鵜木 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,” 音響学会聴覚研究会資料, Vol. 31, No. 10, H2001-92, 日本音楽音響学会資料, MA2001-54.

齋藤 毅, 鵜木 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,” 音響学会春季講演論文集, 1-9-19, March 2002.