

Title	基本周波数とスペクトル包絡を制御した歌声合成に関する研究
Author(s)	清水, 一郎
Citation	
Issue Date	2003-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1663
Rights	
Description	Supervisor:赤木 正人, 情報科学研究科, 修士

修 士 論 文

**基本周波数とスペクトル包絡を制御した
歌声合成に関する研究**

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

清水 一郎

2003 年 3 月

修 士 論 文

基本周波数とスペクトル包絡を制御した 歌声合成に関する研究

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 小谷 一孔 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

110057 清水 一郎

提出年月: 2003 年 2 月

概要

本稿では、話声とメロディから歌声への高品質な歌声合成を実現することを目的とする。ソースフィルタモデルの STRAIGHT を用いることにより基本周波数、スペクトルを制御しながら、話声から歌声を合成する手法を提案する。

スペクトルの時間方向の制御が必要かどうか調べるために、歌声での音節の長さはメロディにより決定されるので、歌声の子音の時間の長さがわかれば母音の時間の長さが決定されると考慮し、話声の子音の時間の長さに対する歌声の子音の時間の長さを分析する。スペクトルの周波数方向の制御が必要かどうか調べるために、話声と歌声の F0 に対するフォルマントを分析した。そして、話声のスペクトルと話声の F0 に対する歌声の F0 から歌声のスペクトルを算出する式を最小 2 乗法により導出した。また、歌声のヴィブラート部の F0 と F1 の微細変動に時間のラグなしに相関があることがわかった。これらの分析結果を用い、話声とメロディから歌声の F0 の制御、スペクトルの時間方向、周波数方向の制御を行い、歌声を合成した。これらの研究の手法で作成された歌声合成音を聴取実験によりに評価した。そして、スペクトルの時間方向、周波数方向の制御の有効性が示された。

目次

第1章	はじめに	1
1.1	本研究の背景	1
1.2	本研究の目的	2
1.3	本研究の特色	2
第2章	本研究の手法	3
2.1	本研究で提案する手法の概要	3
2.2	STRAIGHT と F0 制御モデル	3
2.2.1	STRAIGHT の特徴	3
2.2.2	STRAIGHT	4
2.2.3	F0 の制御モデル	5
2.3	話声のスペクトルを歌声のスペクトルに変形する制御	7
第3章	歌詞のついた歌声と話声の分析結果	11
3.1	歌声資料	11
3.2	歌声と話声の子音の時間の長さの比較	11
3.3	歌声と話声の F0 に対するスペクトル	13
3.4	フォルマント制御法	16
3.4.1	話声のスペクトルと歌声のスペクトルの関係	16
3.4.2	2 変数を用いた歌声のスペクトルの制御	17
3.5	F0 のヴィブラート成分と F1 の微細変動の相関	17
第4章	歌声の合成	27
4.1	歌声の合成条件	27
4.2	話声のスペクトルを用いた時間方向へのスペクトル変形手法	27
4.3	話声のスペクトルを用いた周波数方向へのスペクトル変形手法	28
4.3.1	話声のスペクトルの周波数方向の写像	28
4.3.2	ヴィブラート部の F0 に対応した F1 の微細変動	29
4.4	F0 とスペクトルの時間における関係の調整	29
4.5	歌声合成音	29
4.6	歌声合成の評価	29

4.6.1	結果の処理	31
4.7	歌声合成の結果	31
4.8	まとめ	31
第 5 章	結論	44
5.1	本研究の要約	44
5.2	今後の課題	45

目 次

2.1	歌声の制御手法	4
2.2	STRAIGHT の全体図	5
2.3	齋藤らの F0 制御モデル	6
2.4	提案する F0 制御モデル	7
2.5	実際の歌声の F0	9
2.6	F0 制御モデルにより作成された歌声の F0	9
2.7	スペクトル制御モデル	10
3.1	歌声の分析結果	19
3.2	話声の分析結果	20
3.3	話声のサウンドスペクトルグラム	21
3.4	歌声のサウンドスペクトルグラム	21
3.5	話者 A の歌声における基本周波数に対する /a/ のフォルマント	22
3.6	話者 A の話声における基本周波数に対する /a/ のフォルマント	22
3.7	話者 B の歌声における基本周波数に対する /a/ のフォルマント	23
3.8	話者 B の話声における基本周波数に対する /a/ のフォルマント	23
3.9	最小 2 乗直線による話声のスペクトルと歌声のスペクトルの関係	24
3.10	最小 2 乗 2 次曲線による話声のスペクトルと歌声のスペクトルの関係	24
3.11	話声のスペクトルと話声の基本周波数に対する歌声の基本周波数の比に対する歌声のスペクトル	24
3.12	/karasu/ の /u/ の部分での基本周波数	25
3.13	/karasu/ の /u/ の部分での F1	25
3.14	/karasu/ の /u/ の部分での ヴィブラート部の F0	25
3.15	/karasu/ の /u/ の部分での F1 の微細変動	26
3.16	/karayo/ の /o/ の部分での ヴィブラート部の F0	26
3.17	/karayo/ の /o/ の部分での F1 の微細変動	26
4.1	/ze/ の話声のスペクトル	33
4.2	/ze/ の実際の歌声のスペクトル	33
4.3	8192 点の FFT で分析された話声における /karasu/ の /ra/ の /a/ の周波数方向における 100 チャンネル目の時間に対する振幅	34

4.4	8192 点の FFT で分析された話声における /karasu/ の /ra/ の /a/ の周波数方向における 100 チャンネル目を時間方向へスプライン補間したものの時間に対する振幅	34
4.5	/karasunazenakuno/ の話声の非周期成分	35
4.6	/karasunazenakuno/ の話声の非周期成分を時間方向へ制御を行ったもの	35
4.7	/karasunazenakuno/ の話声のスペクトル	36
4.8	/karasunazenakuno/ の話声のスペクトルを時間制御したもの	36
4.9	周波数方向の制御の流れ	37
4.10	/karasu/ の /u/ における時間制御を行った歌声スペクトルと時間方向と周波数方向の制御を行った歌声のスペクトル	37
4.11	スペクトル制御を行った歌声のスペクトル	38
4.12	/karasu/ の /u/ の F0	38
4.13	/karasu/ の /u/ ヴィブラート部の F0	39
4.14	ヴィブラート部の F0 に対応した揺れを付け加えた F1	39
4.15	F0 とスペクトルの時間関係	40
4.16	モデルに作成された歌声の時間波形、F0、スペクトル包絡	41
4.17	歌声の分析合成音の時間波形、F0、スペクトル包絡	42
4.18	評価尺度	43
4.19	実験結果	43

表 目 次

2.1	歌声動的成分のパラメータ	7
3.1	話声の子音の時間に対する歌声の子音の時間	12
3.2	摩擦音の話声の時間に対する歌声の時間とその標準偏差	12
3.3	破裂音の話声の時間に対する歌声の時間とその標準偏差	12
3.4	半母音の話声の時間に対する歌声の時間とその標準偏差	13
3.5	鼻音の話声の時間に対する歌声の時間とその標準偏差	13
3.6	ヴィブラート部分の F0 と F1 の微細変動との相関	18
4.1	実験における使用器具	30

第1章 はじめに

話声とメロディから歌声への高品質な歌声合成を実現することを目的とする。歌声合成においてソースフィルタモデルにおいて 歌声の F_0 を制御したモデルが提案されている [1]。その研究は、母音/a/のみで行っておりまたスペクトルは考慮されていなかった。そこで、高品質な歌声合成の実現のためにスペクトル制御モデルを提案する。また、ソースフィルタモデルの STRAIGHT を用いることにより齋藤らの F_0 制御モデルにより基本周波数、スペクトルを制御しながら、話声から歌声を合成する手法を提案する。

人が歌を歌うことは感情や想いを伝える重要な感情表現の一つであり、歌がどのようなものかを研究することは大事である。人間は歌を歌うとき自分の音色を保ちつつ、メロディにあわせ音程を激しく変化させている [2]。つまり、人間は歌を歌うとき基本周波数を大きく変化させながら音色をたもっている。

歌声の音色に関係しているといわれるスペクトルについて述べる。話声の母音において、第 1 フォルマント、第 2 フォルマントにより音韻性が決まり、第 3 フォルマント以上のフォルマントはあまり変化せず、個人的特徴に関連していると言われている [3]。歌声において、必ずしも 2 つのフォルマントが各母音らしさを決定しているのではないことがいわれている。たとえば、男性のオペラ歌手が歌った [i:] という母音の第 1 と第 2 フォルマントの周波数は話声における母音図では [y:] に相当する。また、女性のソプラノ歌手の高い声は 1000Hz ないし 1400Hz もの高さの基本周波数で歌うので、話声の第 1 フォルマントより高いという現象が起る。そこで、第 1 フォルマントを基本周波数の近傍へ動かす操作を行う。声種による音色の違いは各声種の歌声について母音のフォルマントの違いに着目して研究されており、一般的に高い声種の場合フォルマント周波数が高いほうに偏移することが述べられている [4]。音色を保つためには基本周波数の変化にともないスペクトルが変化しているのではないかと考えられる。そこで、基本周波数の変化に対応してスペクトルを制御することを考える。

1.1 本研究の背景

歌声の F_0 において次のような研究がなされている。齋藤らは母音 /a/ において、オーバーシュート、ヴィブラート、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変化である予備的変動の F_0 動的成分を明らかにし、 F_0 動的成分を加えた F_0 を制御した歌声合成の研究を行っている [1]。

歌声のスペクトルと基本周波数に関係について次のようなことが言われている。高良ら

は各音韻ごとにそれぞれの音の長さに対応させて、接続させていくという規則合成法を用いていて、原音声で得られたピッチパターンを使用することにより次のような実験を行っている [5]。実験データは歌声の高さで発声した音韻データ 1 音 (中間音) で構成した音、歌声の音韻データ 2 音 (最低音、最高音) で構成した規則合成音、歌声の音韻データ 3 音で構成した規則合成音、歌声の音韻データ 3 音 (最低音、中間音、最高音) で構成した規則合成音、使用するすべての音高の歌声データから構成した規則合成音である。聴取実験により歌声の音韻データ 2 音 (最低音、最高音) で構成した規則合成音の音の変化がはっきりした 2 音が好まれた。これにより、歌声のピッチとスペクトルに関連性があることがわかる。

Slawson は歌声において基本周波数が 270Hz 以下の音声類似の合成音において母音の基本周波数を 1 オクターブ上げる時、ホルマント周波数を平均 10 % あげるともとの音韻性とよく似た母音になると述べている [6]。

また、話声のスペクトルと基本周波数の関係についても次のようなことが言われている。河原らは STRAIGHT を用いて音声の基本周波数を変換したときに、再合成音声の自然性を保つためにはスペクトルの周波数方向への伸長率を基本周波数の $1/2$ から $1/3$ 乗に設定すればよいと述べている [7]。西澤らは基本周波数を考慮したフォルマント音声合成を行っており、一部のフォルマントを話声に応じて制御する必要があると述べている [8]。

歌声の F_0 の制御は齋藤らの研究で行われている、しかしもう 1 つの重要なパラメータであるスペクトルの制御が行われていない。そこで、スペクトルを制御することにより、高品質な歌声合成を行う必要がある。音声において基本周波数とスペクトル包絡には関係性が存在し、歌声において基本周波数によりスペクトルの制御を行う必要があると考えられる。故に、基本周波数に対応した歌声のスペクトル包絡を行えるモデルを提案することにより、高品質な歌声合成を目指す。

1.2 本研究の目的

本研究の目的は、話声とメロディから高品質な歌声を合成することである。

1.3 本研究の特色

本研究の特色は次に述べる 2 つである。

- 基本周波数に対応して歌声の知覚に影響を与えるスペクトルを制御するモデルの提案である。
- 話声から歌声の F_0 、スペクトル包絡を制御することにより高品質な歌声を合成することである。

第2章 本研究の手法

話声とメロディを入力とし、STRAIGHT により歌声を合成するモデルである。STRAIGHT はソースフィルタモデルなので、歌声合成する際、歌声の F0 と歌声のスペクトルが必要である。話声とメロディから歌声の F0 を制御するモデル。歌声のスペクトルを制御するモデルが必要である。そこで、歌声の F0 制御は齋藤らの F0 制御モデルを用い、歌声のスペクトル制御は提案を行う。そして、話声から歌声を合成する手法を提案する。

2.1 本研究で提案する手法の概要

本研究で提案する手法を図 2.1 に示す。話声とメロディを入力とし、STRAIGHT により歌声を合成する手法である。歌声の F0 はメロディを入力とする齋藤らが提案した F0 制御モデルにより作成する。歌声のスペクトル包絡は STRAIGHT により分析された話声の F0 スペクトル包絡、メロディ、作成された歌声の F0 を入力とする歌声のスペクトル包絡制御モデルにより作成される。作成された歌声の F0 と 作成された歌声のスペクトル包絡を入力とし STRAIGHT により歌声を合成する。

2.2 STRAIGHT と F0 制御モデル

本研究で用いる STRAIGHT と F0 制御モデルについて紹介する。

2.2.1 STRAIGHT の特徴

STRAIGHT は次のような特徴がある。

- バズ臭さのない高品質な音声合成音を作成できる。
- 人間の知覚に関係のあるパラメータにより分析し、パラメータの自由な制御を行った後再合成を行える分析合成型のシステムである。
- パラメータと人間の生理学的、心理学的な要素が対応しているので、パラメータの制御が生理学的、心理学的においてどのようなことをしてい対応しているのかがわかりやすい。

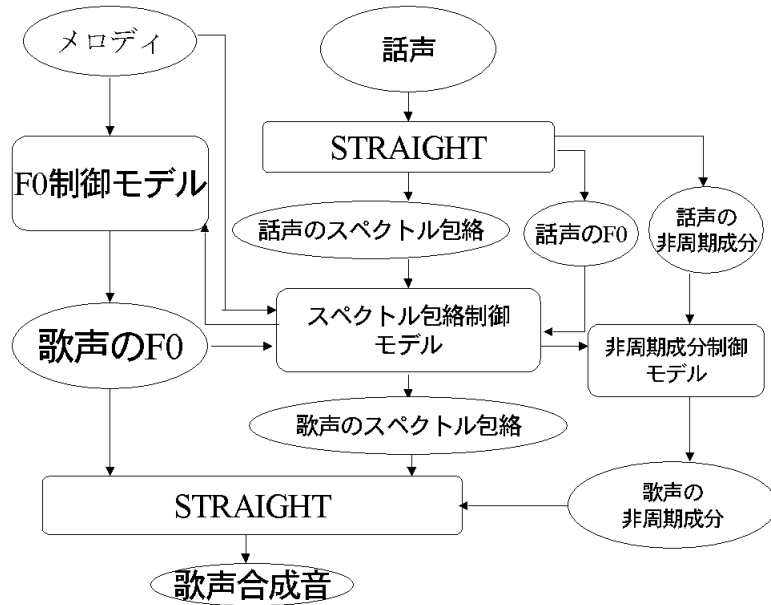


図 2.1: 歌声の制御手法

2.2.2 STRAIGHT

STRAIGHT は図 2.2 のような分析合成を行っている [9]。

STRAIGHT は音韻性に係わる成分であるスペクトル包絡から構成される時間周波数表現と韻律性に係わる成分である基本周波数と有声/無声などの情報、合成を行うときに合成音に自然性を持たせるための音源信号の生成機構である群遅延操作音源から成りたっている。

STRAIGHT は音声のような周期的な信号のスペクトログラムにみられる周期性による干渉と包絡成分の分離を行っている。スペクトル包絡は時間周波数領域での相補的な時間窓による時間平滑化と spline 理論による補償を用いて計算される。F0 抽出は二段階で行われており、始めに帯域フィルタの出力の瞬時周波数とフィルタの中心周波数との写像における不動点近傍での形状と時間変化の情報からそれぞれの調波成分の C/N 比の推定法を与えることにより行われ、次にその初期推定値を複数の調波成分からの情報を利用して改良することで求められる。音源の周期成分と非周期成分の推定は基本周波数の情報に基づき時間軸において非線型伸縮を行うことにより求められている。STRAIGHT は図 2.2 のような分析合成を行っている。

STRAIGHT-core は周期信号から時間周波数領域から滑らかな時間周波数表現を求めている。求め方は時間周波数両面を周期信号によって 2 次元で標本化された情報として復元するための 2 次元の平滑化関数 $h_t(\lambda, \tau)$ を用いることにより計算される。平滑化関数 $h_t(\lambda, \tau)$ は 2 つの 2 次カーディナル B-spline の基底関数を用いて構成される。

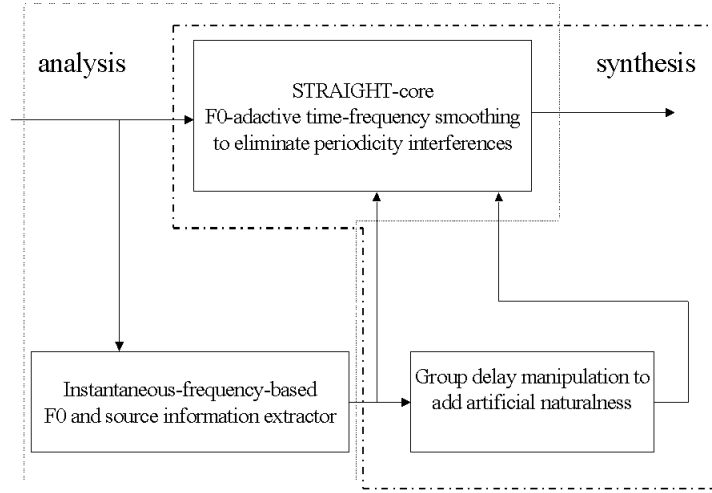


図 2.2: STRAIGHT の全体図

$$h_t(\lambda, \tau) = \frac{1}{4} \left(1 - \left| \frac{\lambda}{\omega_0(t)} \right| \right) \left(1 - \left| \frac{\tau}{\tau_0(t)} \right| \right) \quad (2.1)$$

$$-\omega_0(t) \leq \lambda \leq \omega_0(t), -\tau_0(t) \leq \tau \leq \tau_0(t)$$

ただし、基本周期のときの角周波数を $\omega_0(t)$ 、周波数を λ 、時間を τ とおく。
平滑化関数 $h_t(\lambda, \tau)$ を用いて、基本周波数に適応した平滑化スペクトログラム

$$S(\omega, t) = g^{-1} \left(\int \int_D h_t(\lambda, \tau) g(|F(\omega - \lambda, t - \text{lambda})|^2) d\lambda d\tau \right) \quad (2.2)$$

2.2.3 F0 の制御モデル

齋藤らの F0 制御モデルを用いる [1]。齋藤らの F0 制御モデルを図 2.3 に示す。

齋藤らの F0 制御モデルはメロディから 12 平均律によりステップ状の入力を作成し、それにあるオーバーシュート、ヴィブラート、予備的変動、歌声の微細変動を付け加える F0 モデルである。

F0 制御モデルはオーバーシュート、ヴィブラート、予備的変動の歌声動的成分を次式のような制動 2 次系のインパルス応答 $H(s)$ で記述する。

$$H(s) = \frac{\Omega}{s^2 + 2\zeta\Omega s + \Omega^2} \quad (2.3)$$

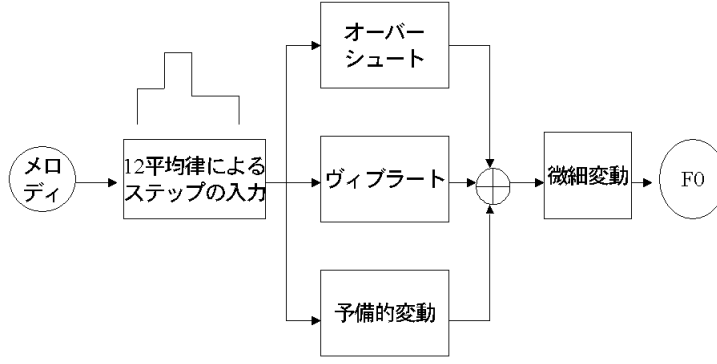


図 2.3: 齋藤らの F0 制御モデル

ζ は減衰振動、 Ω は固有周波数である。

このインパルス応答 $h(t)$ は次のように場合わけされる。

1. $|\zeta| > 1$ のとき

$$h(t) = \frac{\Omega}{2\sqrt{\zeta^2 - 1}}(e^{\lambda_1 \Omega t} - e^{\lambda_2 \Omega t}) \quad (2.4)$$

$$\lambda_{1,2} = -\zeta \pm \sqrt{\zeta^2 - 1}$$

2. $|\zeta| < 1$ のとき

$$h(t) = \frac{\Omega}{\sqrt{\zeta^2 - 1}}e^{-\zeta \Omega t} \sin(\sqrt{1 - \zeta^2} \Omega t) \quad (2.5)$$

3. $|\zeta| = 1$ のとき

$$h(t) = \Omega t e^{-\Omega t} \quad (2.6)$$

4. $|\zeta| = 0$ のとき

$$h(t) = \sin(\Omega t) \quad (2.7)$$

表 2.1: 歌声動的成分のパラメータ

歌声動的成分	Ω [rad/ms]	ζ
オーバーシュート	0.031	0.52
ヴィブラート	0.033	0
予備的変動	0.028	0.72

実際の歌声データと 各歌声動的成分において F0 モデルのフィッティングを行うことにより最適なパラメータ Ω 、 ζ を求めている。F0 制御モデルのオーバーシュート、ヴィブラート、予備的変動の歌声動的成分のパラメータは次表に示す。また、微細変動成分は 10Hz 以下の不規則変動であり、白色雑音をローパスフィルタに通して作成した [8]。

今回の研究では母音、子音のすべてに適用させるため、齋藤らの F0 制御モデルを使い、まずすべて有声音であるとして F0 を作成し、無声子音の部分のところの F0 をないものとした (図 2.4 参照)、何故すべて有声音として F0 を作成するのかというと、無声子音部分は声帯が振動していないだけ、無声子音部分においても目指す音高に向かっている歌を歌っていると考慮したからである。

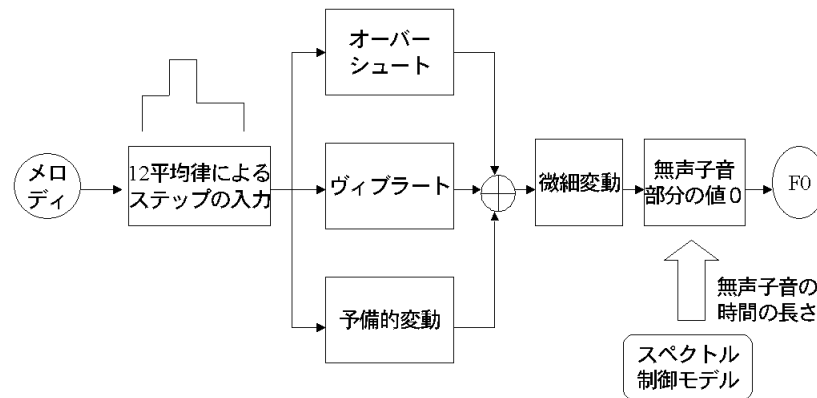


図 2.4: 提案する F0 制御モデル

実際の歌声 F0 と F0 モデルにより作成した歌声 F0 を図 2.5, 図 2.6 にそれぞれ示す。

2.3 話声のスペクトルを歌声のスペクトルに変形する制御

歌声は自分の思いどおりの音色を出しながら、声の高さを大きく変化させ、かつ声の強さも激しく変化させている。声の高さを大きく変化させながらも自分の音色を保っていることは、基本周波数の変化にともないスペクトルも変化していると考えられる。より高品質な歌声合成を目指すためには、基本周波数の変化に対応したスペクトル包絡を制御する

必要がある。基本周波数を考慮することにより話声のスペクトルから歌声のスペクトルに変形するスペクトル制御を行うこととする。メロディ、話声の F_0 、話声のスペクトル、歌声 F_0 制御モデルにより作成した F_0 を入力として スペクトルの時間方向の変形、スペクトルの周波数方向への変形を行うことにより歌声のスペクトルを出力とするスペクトル制御モデルを提案する。スペクトル制御モデルを図 2.7 に示す。

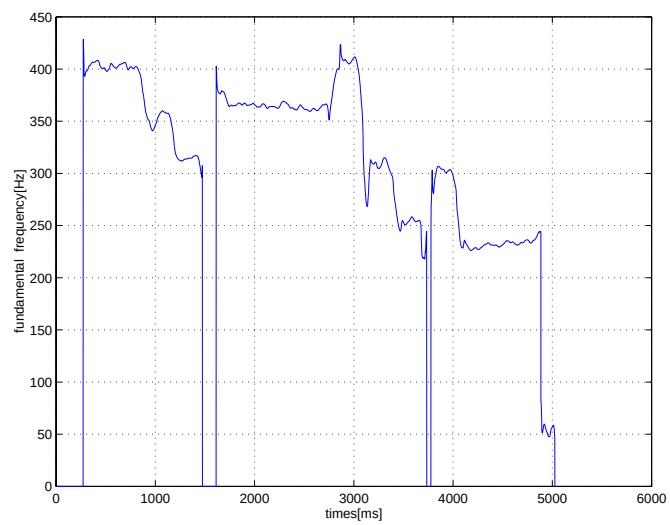


図 2.5: 実際の歌声の F0

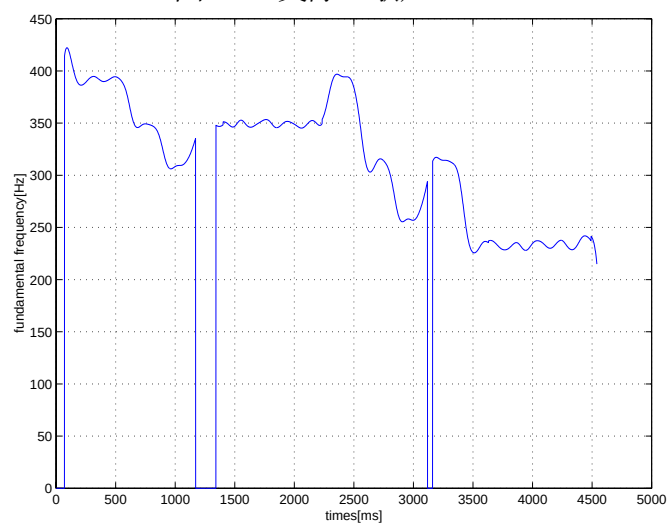


図 2.6: F0 制御モデルにより作成された歌声の F0

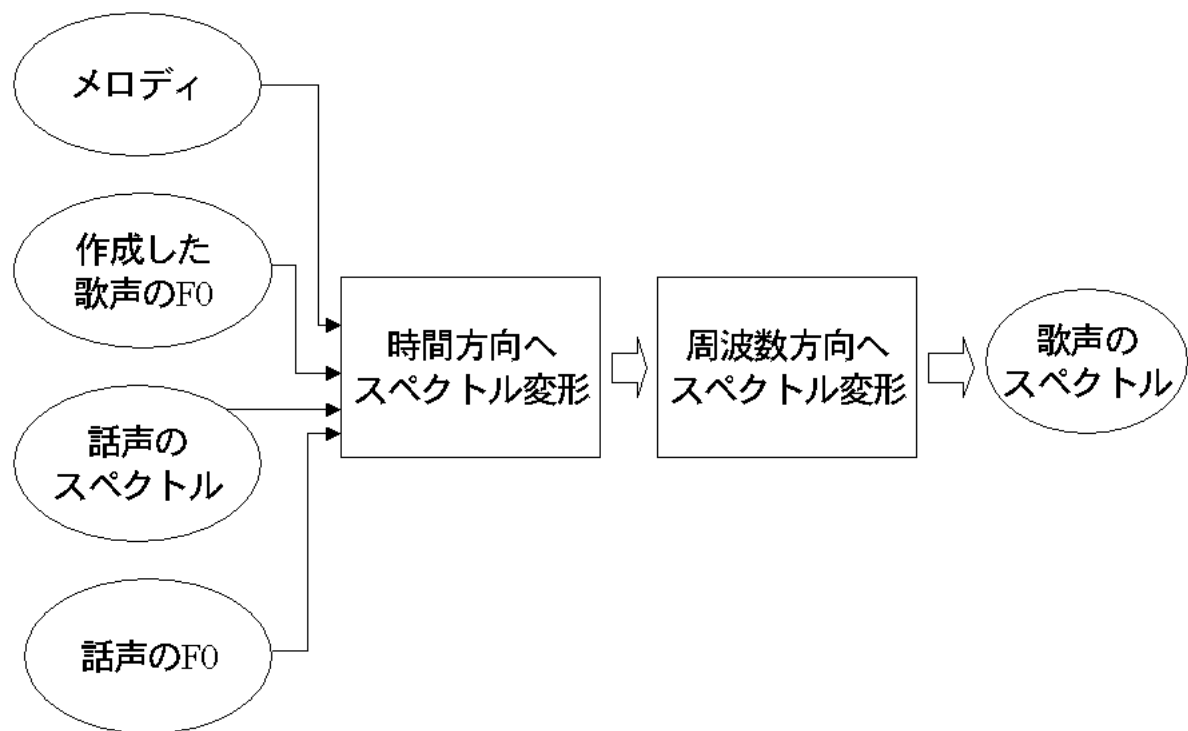


図 2.7: スペクトル制御モデル

第3章 歌詞のついた歌声と話声の分析結果

話声から歌声を合成するために、話声、歌声の違いを分析する必要がある。歌声の F_0 、話声のスペクトルから歌声のスペクトルの制御を行うので、話声と歌声の F_0 、スペクトル包絡を分析した。最初に話声と歌声の子音の時間の長さを比較した。子音の時間の長さを比較した理由は、メロディにより歌声での音節の長さは決まっており、1つの音符に対して子音+母音、母音のいずれかであり、歌声の子音の長さが制御できればどうじに母音の時間の長さも決定できると考えたからである。次に話声と歌声のスペクトルに関連して、今回は基本周波数に対応したスペクトル包絡の制御を行うので、話声と歌声の基本周波数に対し、スペクトルの特徴の1つであるフォルマントを分析対象にすることとした。

歌詞のついた歌声と話声の分析結果として時間波形、 F_0 、スペクトルの図を図 3.1、図 3.2 にそれぞれ示す。

3.1 歌声資料

北風らがボーカルスクールに通う学生である女性2名から採取した七つの子の歌声データと歌詞を朗読した話声のデータを用いた。歌声データは防音室においてサンプリング周波数 48[kHz]、量子化ビット数 16[bit] で DAT によりデジタル録音された七つの子の歌声データを 12[kHz] にダウンサンプリングしたデータである。話声のデータは防音室においてサンプリング周波数 48[kHz]、量子化ビット数 16[bit] で DAT によりデジタル録音された七つの子の歌詞を朗読した話声のデータを 12[kHz] にダウンサンプリングしたデータである。

3.2 歌声と話声の子音の時間の長さの比較

話声のスペクトルから歌声のスペクトルを作成するために、音素ごとに話声の子音の時間の長さに対する歌声の子音の時間の長さを求めた。サウンドスペクトルグラムにより分析を行い、話声と歌声の同じ音韻を比較する。話声と歌声のサウンドスペクトルグラムをそれぞれ図 3.3、図 3.4 に示す。

子音の調音位置と調音方式ごとに話声の時間に対する歌声の時間を並べてみた。子音の

音素ごとに調音位置と調音方式を考え話声の時間に対する歌声の時間を並べた表 3.1 を示す。子音の調音方式ごとに子音の時間の長さを制御する。

表 3.1: 話声の子音の時間に対する歌声の子音の時間

	口唇		歯、歯茎		硬口蓋		軟口蓋		声門
	有声	無声	有声	無声	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	/w/ 2.61		/r/ 2.12						
鼻音	/m/ 1.35		/n/ 1.50						

データの確からしさを示すために、摩擦音、破裂音、半母音、鼻音についてサンプル数と標準偏差を計算した。摩擦音、破裂音、半母音、鼻音について話声の時間に対する歌声の時間とその標準偏差をそれぞれ表 3.2、表 3.3、表 3.4、表 3.5 に示す。

話声の子音のスペクトルから歌声の子音のスペクトルを作成するためには、話声のスペクトルを摩擦音のとき 1.28 倍、破裂音のとき 1 倍、半母音のとき 2.37 倍、鼻音のとき 1.43 倍、/y/ のとき 1.22 倍にすればよい。

表 3.2: 摩擦音の話声の時間に対する歌声の時間とその標準偏差

音素	サンプル数	標準偏差	話声の時間に対する歌声の時間 [倍]
/s/	9	0.12	1.18
/h/	6	0.16	1.28
/z/	4	0.28	1.37

表 3.3: 破裂音の話声の時間に対する歌声の時間とその標準偏差

音素	サンプル数	標準偏差	話声の時間に対する歌声の時間 [倍]
/k/	35	0.31	0.97
/t/	5	0.33	1.09
/g/	3	0.17	1.14
/d/	2	0.00	1.00

表 3.4: 半母音の話声の時間に対する歌声の時間とその標準偏差

音素	サンプル数	標準偏差	話声の時間に対する歌声の時間 [倍]
/r/	11	0.93	2.12
/w/	7	0.83	2.61

表 3.5: 鼻音の話声の時間に対する歌声の時間とその標準偏差

音素	サンプル数	標準偏差	話声の時間に対する歌声の時間 [倍]
/n/	26	0.66	1.50
/w/	4	0.25	1.35

3.3 歌声と話声の F0 に対するスペクトル

話声のスペクトルから基本周波数に対応した歌声のスペクトルを制御するために、スペクトルの特徴量であるフォルマントに着目し基本周波数に対する話声のフォルマントと歌声のフォルマントを求める。基本周波数に対する歌声と話声のフォルマントを分析し、基本周波数に対する話声と歌声のフォルマントの傾向を見るために、それぞれについて式を最小 2 乗直線で求める。

フォルマントの計算は次のように行った。自己相関法から予測係数を求め、全極型のフィルタと考えフォルマントを計算することとする。

振幅スペクトル $X(\omega)$ 、パワースペクトルを $S(\omega)$ 、自己相関を $R(\tau)$ 、予測係数を a_k とする、LPC の次数を p 、サンプリング周波数を F_s とする。

$$S(\omega) = |X(\omega)|^2 \quad (3.1)$$

$$R(\tau) = IFFT(S(\omega)) \quad (3.2)$$

正規方程式は

$$R((\tau))a_k = -R(\tau + 1) \quad (3.3)$$

$$A(Z) = 1 + \sum_{k=1}^p a_k z^{-k} = 0 \quad (3.4)$$

$A(Z)=0$ の根を求め、 Z の振幅が単位円に近いものをフォルマントと考え、 Z の位相 $W[\text{rad}]$ により フォルマント周波数を計算する。フォルマントの周波数 F は次式のように求める。

$$F = \frac{W}{\pi} \times \frac{F_s}{2} \quad (3.5)$$

話者 A の母音 /a/ における基本周波数に対する歌声のフォルマント周波数を図 3.5 に示し、話声のフォルマントを図 3.6 に示す。話者 B の母音 /a/ における基本周波数に対する歌声のフォルマントを図 3.7 に示し、話声のフォルマントを図 3.8 に示す。

歌声の母音 /a/ の F_1, F_2, F_3 は次の式のようなである。

$$F_1 = 1.1F_0 + 6.5 \times 10^2 \quad (3.6)$$

$$F_2 = 1.0F_0 + 1.5 \times 10^3 \quad (3.7)$$

$$F_3 = 1.5F_0 + 2.6 \times 10^3 \quad (3.8)$$

話声の母音 /a/ の F_1, F_2, F_3 は次の式のようなである。

$$F_1 = 1.3F_0 + 5.2 \times 10^2 \quad (3.9)$$

$$F_2 = -0.6F_0 + 1.7 \times 10^3 \quad (3.10)$$

$$F_3 = 1.0F_0 + 3.0 \times 10^3 \quad (3.11)$$

歌声の母音 /i/ の F_1, F_2, F_3 は次の式のようなである。

$$F_1 = 0.36F_0 + 2.1 \times 10^2 \quad (3.12)$$

$$F_2 = -0.25F_0 + 3.0 \times 10^3 \quad (3.13)$$

$$F_3 = 0.85F_0 + 3.5 \times 10^3 \quad (3.14)$$

話声の母音 /i/ の F1,F2,F3 は次式のようなものである。

$$F_1 = 0.71F_0 + 1.4 \times 10^2 \quad (3.15)$$

$$F_2 = 1.6F_0 + 2.2 \times 10^3 \quad (3.16)$$

$$F_3 = 2.8F_0 + 2.7 \times 10^3 \quad (3.17)$$

歌声の母音 /u/ の F1,F2,F3 は次式のようなものである。

$$F_1 = 0.27F_0 + 3.0 \times 10^2 \quad (3.18)$$

$$F_2 = 1.9 \times 10^3 \quad (3.19)$$

$$F_3 = 0.85F_0 + 2.7 \times 10^3 \quad (3.20)$$

話声の母音 /u/ の F1,F2,F3 は次式のようなものである。

$$F_1 = 0.10F_0 + 3.0 \times 10^2 \quad (3.21)$$

$$F_2 = -1.4F_0 + 2.0 \times 10^3 \quad (3.22)$$

$$F_3 = 0.85F_0 + 2.9 \times 10^3 \quad (3.23)$$

歌声の母音 /o/ の F1,F2,F3 は次式のようなものである。

$$F_1 = 0.90F_0 + 4.3 \times 10^2 \quad (3.24)$$

$$F_2 = 1.7F_0 + 7.5 \times 10^2 \quad (3.25)$$

$$F_3 = 1.1F_0 + 2.7 \times 10^3 \quad (3.26)$$

話声の母音 /o/ の F_1, F_2, F_3 は次式のようなものである。

$$F_1 = 0.16F_0 + 6.1 \times 10^2 \quad (3.27)$$

$$F_2 = -0.85F_0 + 1.5 \times 10^3 \quad (3.28)$$

$$F_3 = 0.45F_0 + 2.9 \times 10^3 \quad (3.29)$$

話声は F_0 が高くなるほど F_1, F_3 が高くなる傾向が見られ、 F_2 が低くなる傾向が見られた。歌声は F_0 が高くなるほど F_1, F_2, F_3 が高くなる傾向がみられた。

3.4 フォルマント制御法

3.4.1 話声のスペクトルと歌声のスペクトルの関係

話声のスペクトルと歌声のスペクトルの関係をそれぞれ最小 2 乗直線と最小 2 乗 2 次曲線を引いたものを図 3.9、図 3.10 に示す。話声のスペクトルを FSP 、歌声のスペクトル I 、話声の基本周波数に対する歌声の基本周波数の比を $\bar{F}_0$ とする。図 3.9、図 3.10 の点は歌声の平均の F_0 に対する母音 /a/、/i/、/u/、/o/ に対する上で求められた最小 2 乗直線の F_1 、 F_2 、 F_3 によって求められたものである。

最小 2 乗直線の式は

$\bar{F}_0=0.8$ のとき

$$I = 0.91FSP + 8.1 \times 10 \quad (3.30)$$

$\bar{F}_0=1.0$ のとき

$$I = 0.93FSP + 1.0 \times 10^2 \quad (3.31)$$

$\bar{F}_0=1.5$ のとき

$$I = 0.98FSP + 1.5 \times 10^2 \quad (3.32)$$

$\bar{F}_0=2.0$ のとき

$$I = 1.0FSP + 2.0 \times 10^2 \quad (3.33)$$

3.4.2 2 変数を用いた歌声のスペクトルの制御

話声のスペクトルと話声の基本周波数に対する歌声の基本周波数の比を変数として歌声のスペクトルを制御する。話声のスペクトル、話声の基本周波数に対する歌声の基本周波数の比、歌声のスペクトルのデータの点を図 3.11 に示す。話声のスペクトル、話声の基本周波数に対する歌声の基本周波数の比から歌声のスペクトルを求める式をデータから最小 2 乗によって求めた。

歌声のスペクトル制御の式は話声のスペクトルを P とし、話声の基本周波数に対する歌声の基本周波数の比を $\bar{F}_0$ とすると次式のように表せる。

$$I = 0.13\bar{F}_0P + 0.84P \quad (3.34)$$

最小 2 乗法により歌声のスペクトル I の求め方は図 3.9 から $I = a\bar{F}_0P + bP$ とおくことができると考え、次式の誤差を最小にするような a, b を求める。

$$E = \sum_{i=1}^N (a\bar{F}_{0i}P_i + bP_i - I_i)^2 \quad (3.35)$$

平均 2 乗誤差 E は 278[Hz] であった。2 人の それぞれ 81 個の実際のデータ 間の平均 2 乗誤差は 406[Hz] であった。2 人により適応したスペクトル制御の周波数方向への写像の式を提案できた。

平均 2 乗誤差は 次の式で求めた。

$$E = \sqrt{\sum_{i=1}^N (z_i - I_i)^2} \quad (3.36)$$

3.5 F0 のヴィブラート成分と F1 の微細変動の相関

F0 のヴィブラートがかかっている部分と F1, F2, F3 の概形を見てみると、F0 と F1 の相関が高そうで、F0 と F2, F3 は相関があまりなさそうだった。よって、STRAIGHT において分析したときの F0 のヴィブラート成分と F1 の微細変動の相関があるかどうかを調べた。

/karasu/ の /u/ の F0 と F1 をそれぞれ図 3.12 と図 3.13 に示す。

/karasu/ の /u/ の歌声の ヴィブラート部の F0 を図 3.14、F1 の微細変動を図 3.15 に示す。また、/nakuyo/ の /o/ の歌声の ヴィブラート部の F0 を図 3.16、F1 の微細変動を図 3.17 に示す。a F0 と F1 の時間変化にバンドパスフィルタを通し、ヴィブラート部分の F0 と F1 の微細変動を求めて相関をとる。歌声における ヴィブラート部分の F0 と F1 の微細変動が相関があるかを求めるために相関係数を用いた。相関係数を求める式を式 3.37 に示す。表 1 に F0 と F1 の相関係数を示す。表 1 から F0 と F1 の相関は時間の

ラグなしのときに最大になり、F0 と F1 は時間のずれなく相関が持っていることがあきらかになった。

$$r_{xy}(\tau) = \frac{R_{xy}(\tau)}{\sqrt{R_{xx}(0)R_{yy}(0)}} \quad (3.37)$$

表 3.6: ヴィブラート部分の F0 と F1 の微細変動との相関

音素	相関係数の最大値	最大となる F0 と F1 の時間差 [ms]
/i/	0.6383	1
/u/	0.6950	0

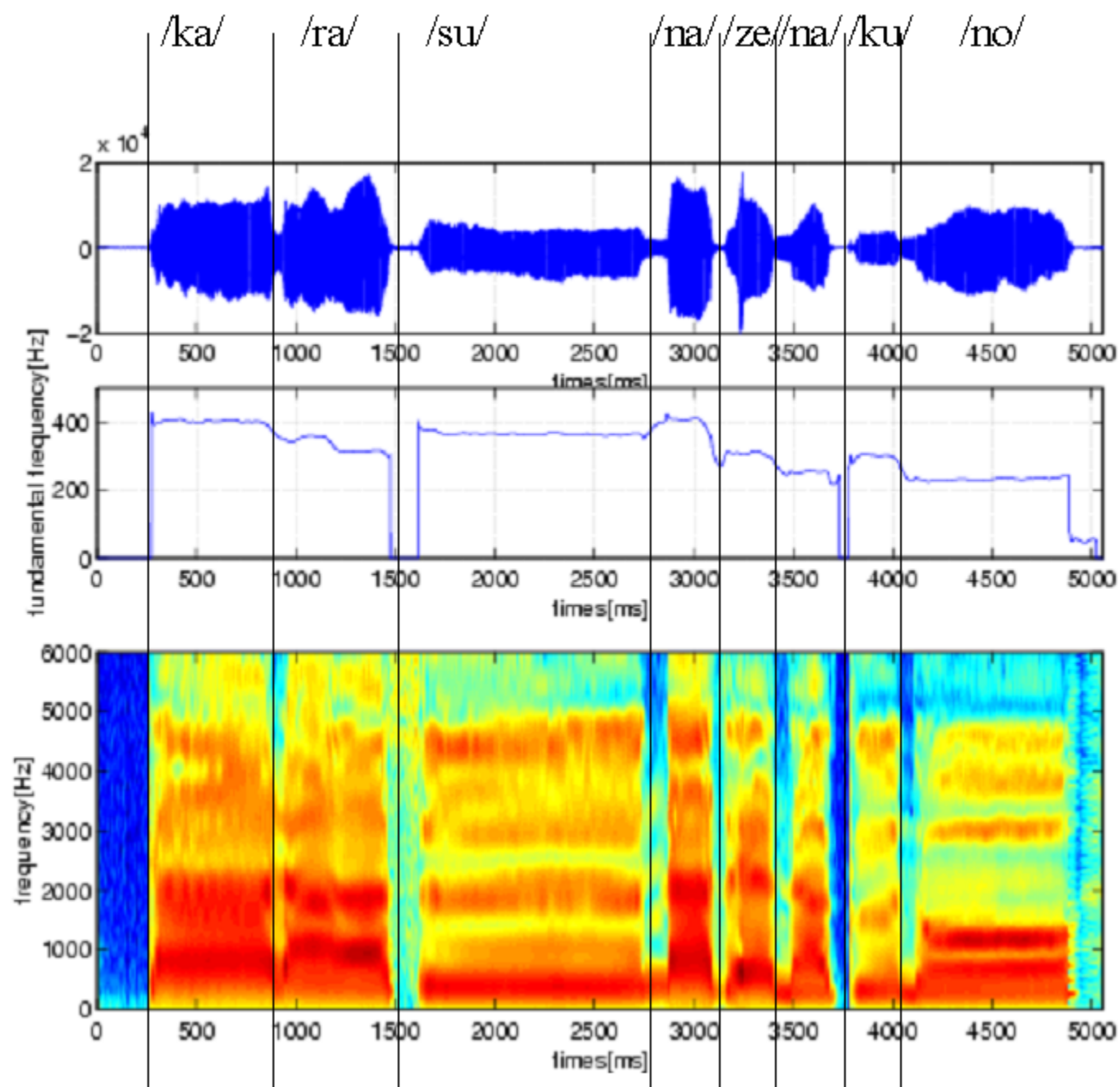


図 3.1: 歌声の分析結果

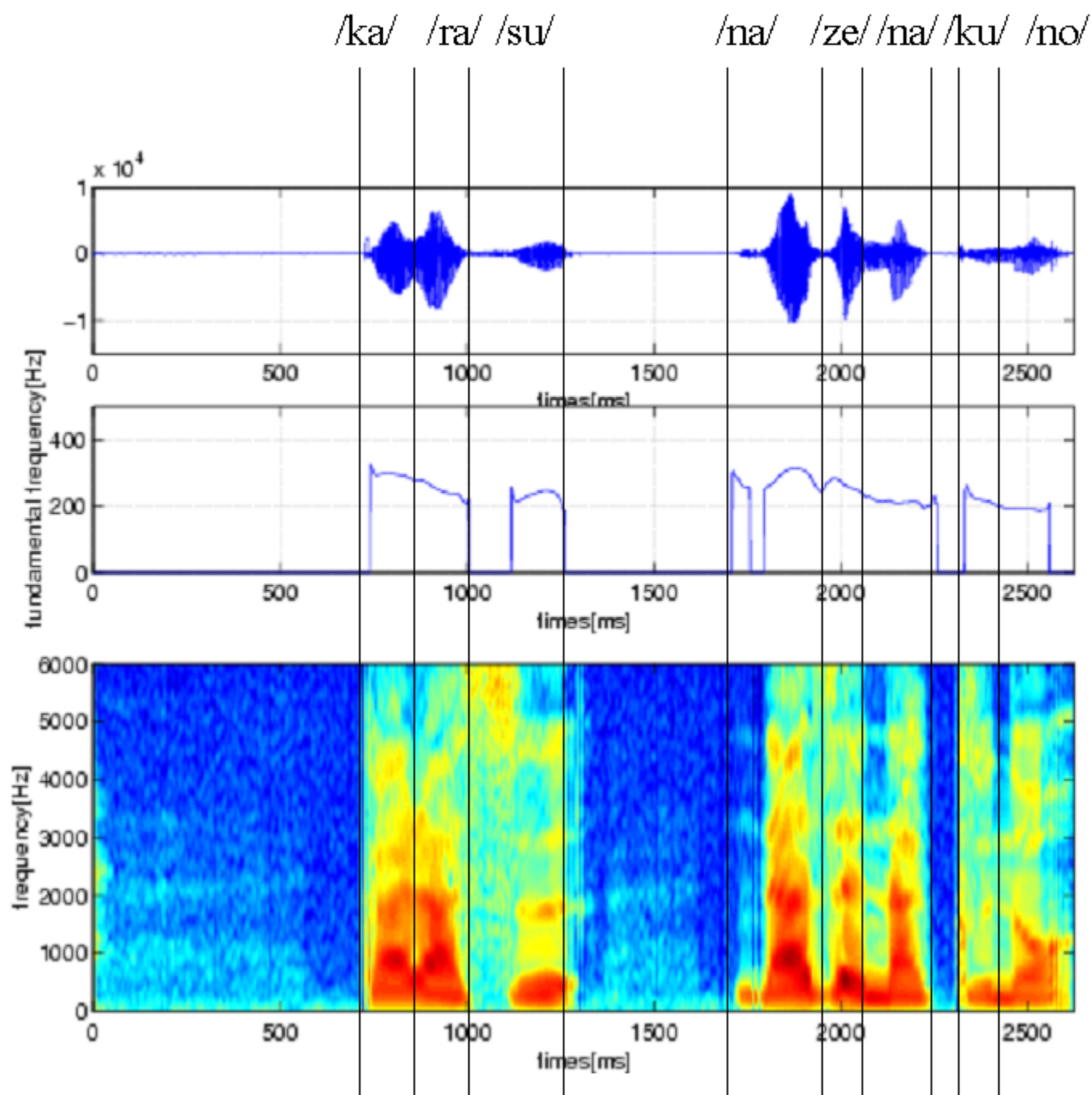


図 3.2: 話声の分析結果

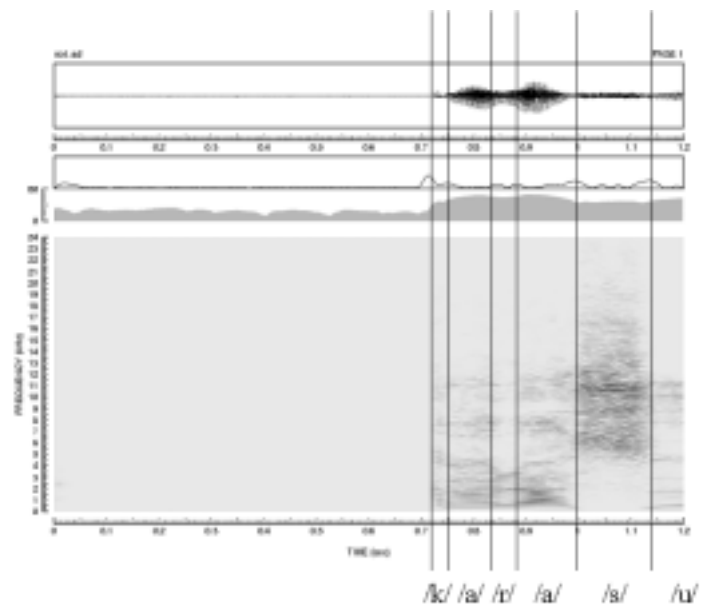


図 3.3: 話声のサウンドスペクトルグラム

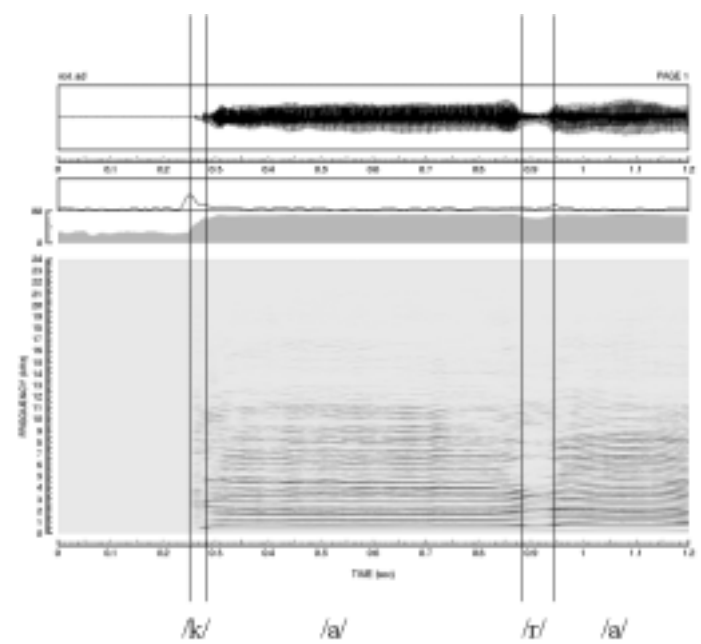


図 3.4: 歌声のサウンドスペクトルグラム

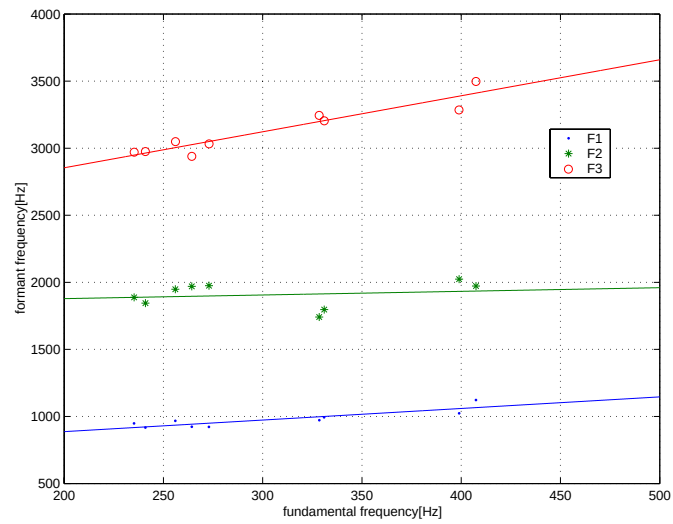


図 3.5: 話者 A の歌声における基本周波数に対する /a/ のフォルマント

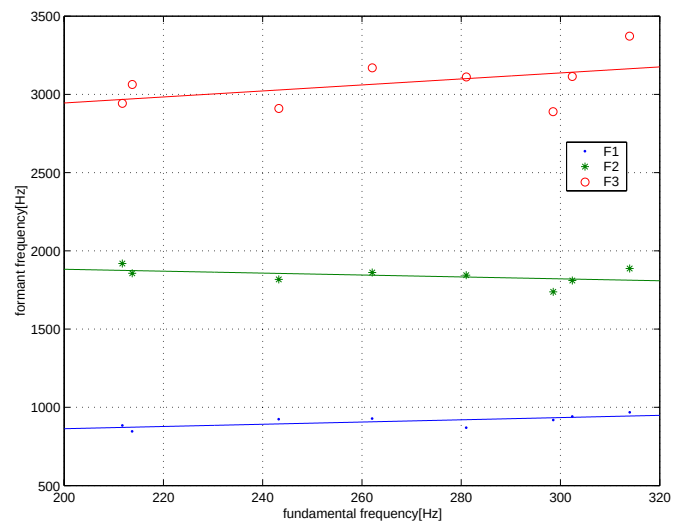


図 3.6: 話者 A の話し声における基本周波数に対する /a/ のフォルマント

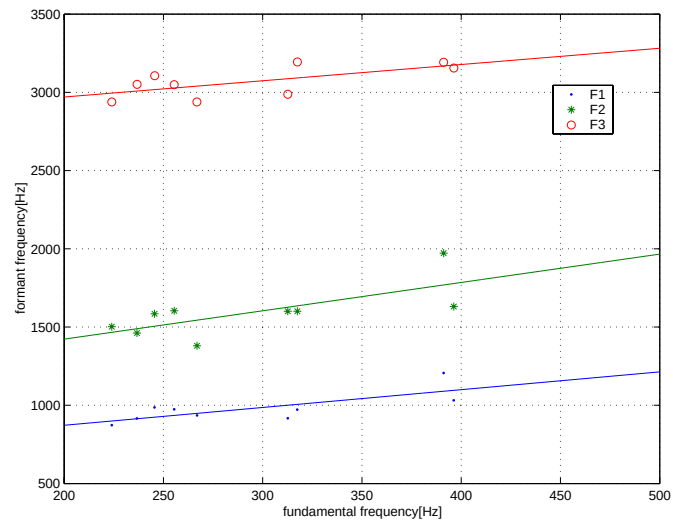


図 3.7: 話者 B の歌声における基本周波数に対する /a/ のフォルマント

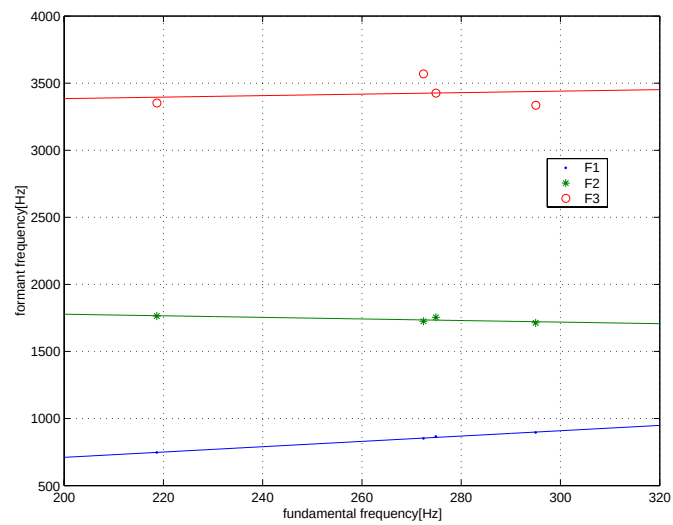


図 3.8: 話者 B の話し声における基本周波数に対する /a/ のフォルマント

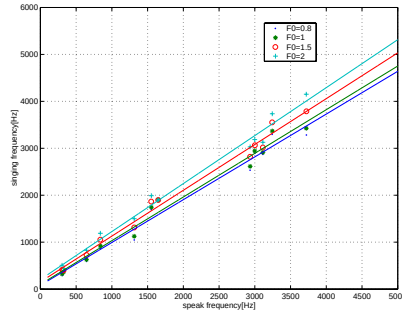


図 3.9: 最小 2 乗直線による話声のスペクトルと歌声のスペクトルの関係

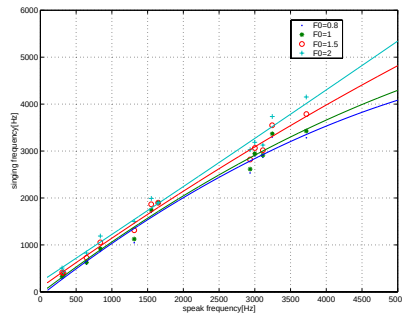


図 3.10: 最小 2 乗 2 次曲線による話声のスペクトルと歌声のスペクトルの関係

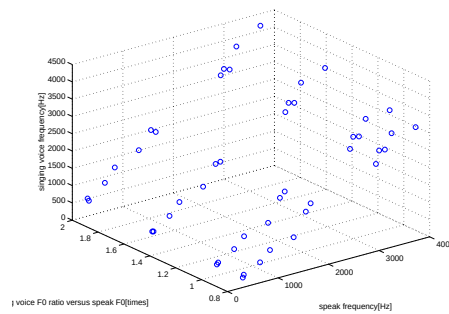


図 3.11: 話声のスペクトルと話声の基本周波数に対する歌声の基本周波数の比に対する歌声のスペクトル

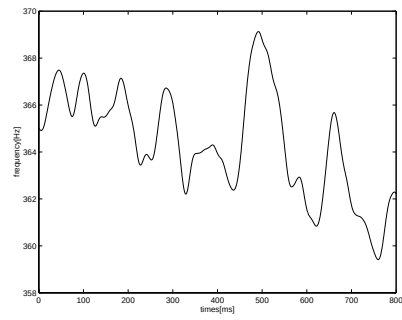


図 3.12: /karasu/の/u/の部分での基本周波数

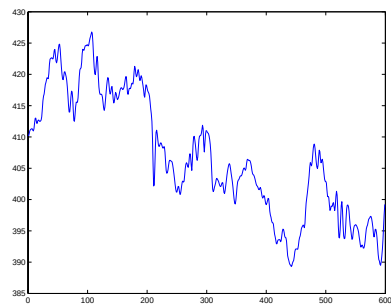


図 3.13: /karasu/の /u/の部分での F1

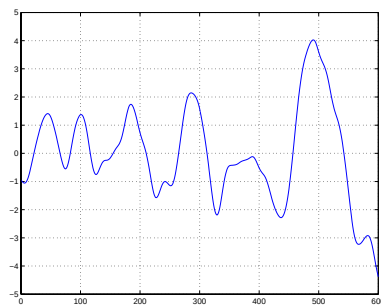


図 3.14: /karasu/の /u/の部分での ヴィブラート部の F0

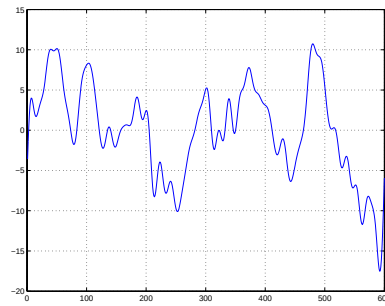


図 3.15: /karasu/の /u/の部分での F1 の微細変動

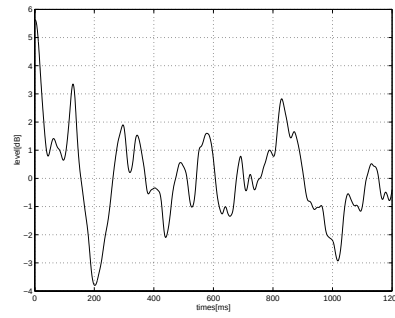


図 3.16: /karayo/の /o/の部分での ヴィブラート部の F0

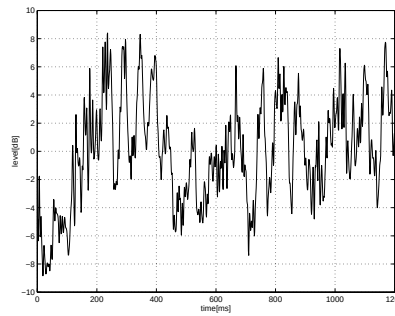


図 3.17: /karayo/の /o/の部分での F1 の微細変動

第4章 歌声の合成

話声とメロディを入力とし、歌声の F0、スペクトル包絡、非周期成分を作成し、STRAIGHT により歌声の合成を行う。

歌声の F0 はスペクトルの時間方向の制御が終り、無声子音の時間区間がわかってから、メロディと無声子音の時間を入力として F0 制御モデルにより作成される。歌声のスペクトルは STRAIGHT により分析された 話声のスペクトル とメロディからスペクトルの時間方向の制御が行われ、時間方向の制御が行われたスペクトルと歌声の F0、話声の F0 から周波数方向の制御が行われることにより作成される。歌声の非周期成分はスペクトルの時間方向の制御とともに話声の非周期成分を同じ時間の区間だけ制御を行う。

4.1 歌声の合成条件

STRAIGHT において 12[kHz] でダウンサンプリングした話声データを FFT の長さを 8192 で分析し、FFT の長さ 8192、群遅延の標準偏差 1[ms]、群遅延の空間周波数 5000[Hz] で歌声を合成した。使用した STRAIGHT のバージョンは V30kr16 である。

4.2 話声のスペクトルを用いた時間方向へのスペクトル変形手法

時間方向における話声のスペクトルから歌声のスペクトルへの変形として、子音の定常部、調音結合部 40[ms]、母音定常部にセグメンテーションする。歌声のスペクトルの制御は調音結合部はそのまま話声のものを扱い、話声の母音、子音の定常部のスペクトルを伸長する。図 4.1 の話声の /ze/ のスペクトルから調音結合部はそのまま話声のものを扱い、話声の母音、子音の定常部のスペクトルを伸長することにより作成された歌声のスペクトルを図 4.2 に示す。

分析結果から歌声の子音の長さは話声の子音の時間の長さを摩擦音のとき 1.28 倍、破裂音のとき 1 倍、半母音のとき 2.37 倍、鼻音のとき 1.43 倍、/y/ のとき 1.22 倍すればよい。歌声の母音の時間の長さは音符の時間の長さか音符の時間の長さから子音の時間を引いたものから決定される。母音、子音の定常部の時間の長さを伸ばす手法として 3 つを提案する。1 つ目は話声の定常部の 1 点を取ってきて、それを伸ばしたい時間の長さだけ並べることにより歌声の定常部を作成する。2 つ目は 8192 点の FFT で分析された

話声の定常部を伸ばしたい時間の長さにするために各周波数方向へチャンネルごとに時間方向へ線形補間を行う。3 つ目は 8192 点の FFT で分析された話声の定常部を伸ばしたい時間の長さにするために各周波数方向へチャンネルごとに時間方向へスプライン補間を行う。

3 つの方法でスペクトルの制御を行って作成された歌声合成音の受聴により、スプライン補間を行ったものが最も高品質だと思われたので、スペクトルに時間方向への制御はスプライン補間を採用することとした。

どのようにスペクトルの時間方向への制御を行っているかを示すために、8192 点の FFT で分析された話声における /karasu/ の /ra/ の /a/ の周波数方向における 100 チャンネル目の時間に対する振幅を図 4.3、話声における /karasu/ の /ra/ の /a/ の周波数方向における 8192 点の FFT で分析された 100 チャンネル目の時間に対する振幅をスプライン補間によりスペクトルの時間方向への制御を行った例を図 4.4 に示す。

STRAIGHT の非周期成分の制御をスペクトルの時間の変形とともに行う。母音、子音の種類により非周期成分の割合が異なるので、スペクトルの時間方向に変形とともに非周期成分の制御を行わなければならない。非周期成分の制御はスペクトルと同じ時間区間で各チャンネルごとに時間方向へスプライン補間を行った。話声の非周期成分と非周期成分の時間方向への制御を行ったものを図 4.5、図 4.6 にそれぞれ示す。

話声のスペクトルとスペクトルの時間制御をした歌声のスペクトルを図 4.7、図 4.8 にそれぞれ示す。

4.3 話声のスペクトルを用いた周波数方向へのスペクトル変形手法

3 章での分析から話声のスペクトルと話声の F_0 に対する歌声の F_0 の関係から歌声のスペクトルを算出した。分析に基づきスペクトルの周波数方向への制御は話声のスペクトルの周波数方向への写像を行い、ヴィブラート部の F_0 に対応した F_1 の微細変動をつけく加える。周波数方向への流れを図 4.9 に示す。

4.3.1 話声のスペクトルの周波数方向の写像

3 章の分析により得られた話声のスペクトルから基本周波数により歌声のスペクトルの計算を行える $I = 0.13\bar{F}_0P + 0.84P$ の式を用い、1[ms] ごとに話声のスペクトルを写像することにより歌声のスペクトルを作成する。

/karasu/ の /u/ における話声のスペクトルを時間方向のスペクトル制御を行ったものと $\bar{F}_0 = 1.51$ で時間方向と周波数方向のスペクトル制御を行ったスペクトルを図 4.10 に示す。スペクトル制御を行った歌声のスペクトルを図 4.11 に示す。

4.3.2 ヴィブラート部の F0 に対応した F1 の微細変動

歌声の分析結果からヴィブラート部の F0 と F1 の微細変動が時間のずれがなく相関があることがわかった。(3.5 節参照) ヴィブラート部の F0 に対応した揺れを同期して F1 の微細変動として F1 に付け加えた。/karasu/ の /u/ の ヴィブラート部の F0 において、F1 に微細変動を加えた例を示す。/karasu/ の /u/ の F0 を図 4.12、/karasu/ の /u/ ヴィブラート部の F0 を図 4.13、ヴィブラート部の F0 に対応した揺れを F1 の微細変動を図 4.14 にそれぞれ示す。

4.4 F0 とスペクトルの時間における関係の調整

F0 とスペクトルの調整は、次のように行った。F0 はメロディにより作成されるので、メロディの始まりにおいてスペクトルは子音、母音か母音になる。F0 とスペクトルの調整の例を図 ?? に示す。

4.5 歌声合成音

モデルにより作成された歌声の時間波形、F0、スペクトル包絡を図 4.16 に示す。歌声の分析合成音の時間波形、F0、スペクトル包絡を図 4.17 に示す。

4.6 歌声合成の評価

聴取実験により品質の良い歌声合成を検討する。シェッフェの一対比較実験を用い聴取実験により歌声合成音の評価を行うこととした。

歌声合成の聴取実験を行う。実験方法として 2 つずつ対に判断することにより、人間にとって比較的簡単であり、判断が簡単なため判断の信頼性が高く心理尺度を求める方法である一対比較法を採用する。サーストン法などの一対比較法は多くの被験者を必要とするので、被験者が判断した点を評価点の序数尺度のまま統計的検定を行い多くの被験者を必要としないシェッフェの一対比較法を採用する。実験の教示は次のように行う。判断の評価尺度は図 4.18 のように 5 段階評価で行う。被験者は 正常の聴力を有した大学院生 3 名で行った。実験は DAT-LINK を通して音を再生し、ヘッドフォンを通して両耳受聴により防音室内で行われた。

- 実験の教示

ヘッドフォンにより 2 つの音声を聞かせます。前の音声より後の音声の方がどの程度歌声らしいかを判断してください。判断は 2. 前の音声の方が歌声らしい、1. 前の音声の方がやや歌声らしい、0. 同じくらい歌声らしい、-1. 後の音声の方がやや

歌声らしい、-2. 後の音声の方が歌声らしい の項目どれかの選択をしてください。
対はつぎつぎと提示されますので、遅れないように判断してください。

実験に使用した器具を表 4.1 に示す。

- 使用器具

表 4.1: 実験における使用器具

使用器具	メーカー	型番
DAT	SONY	57ES
アンプ	SONY	F333ESJ
DAT-LINK+	<i>TOWNSHENDCOMPUTERTOOLS</i>	DAT LINK V1.6
ヘッドフォン	SENNHEISER	HDA200
アンプ	SANSUI	<i>AU - α907MR</i>

- 刺激条件

実験データとして 5 つのデータを作成した。5 つのデータの種類の後で述べる。5 つのデータを 2 つずつ対して 20 対作成しそれを 1 セットとする。ランダムな順番で 1 人に対し 3 セット行った。比較する 2 つのメロディ間は約 2 秒、対間は約 5 秒とした。使用する刺激の音圧は 人工耳を持ちいて測定し、75dB SPL である。

- 刺激

1. timechange

話声のスペクトルの母音、子音の定常部を時間方向にスプライン補間し作成された歌声のスペクトルを用いた歌声合成音

2. freqchange

話声のスペクトルの母音、子音の定常部を時間方向にスプライン補間し $I = 0.13\bar{F}_0P + 0.84P$ の式を用いてのみの周波数方向への制御を行い作成された歌声のスペクトルを用いた歌声合成音

3. all

話声のスペクトルの母音、子音の定常部を時間方向にスプライン補間し $I = 0.13\bar{F}_0P + 0.84P$ を用い、 F_0 のヴィブラート成分に対する F_1 の微細変動の周波数方向への制御を行い作成された歌声のスペクトルを用いた歌声合成音

4. sing

歌声の分析合成音

5. speak

実際の話声

4.6.1 結果の処理

実験データの結果の処理の方法として、芳賀の変法を用いる [11]。母数の値が大きいほど、より歌声らしい音になる。実験結果を図 4.19 に示す。

4.7 歌声合成の結果

歌声のスペクトル制御モデルとしては、freqchange が最もよかった。聴取実験の結果から歌声らしさは話声、time-change、all、freqchange、歌声の分析合成音である。モデルの歌声合成音は歌声の分析合成音には及ばなかった。モデルの歌声合成音の freqchange、all、time-change が話声より大いに歌声らしかったことから、モデルのどの歌声合成音も歌声らしいことが言える。また、モデルの歌声合成音の freqchange、all が time-change よりも歌声らしかったことからスペクトル制御において周波数方向の制御の有効性を示せた。all が freqchange より歌声らしさがなく、ヴィブラート部の F0 に対応した揺れを同期して F1 の微細変動として F1 に付け加えるスペクトルの周波数方向の制御の有効性が示せなかった。

モデルの歌声合成音は歌声に聞こえることが明らかにされ、話声から歌声への制御モデルが提案できた。

4.8 まとめ

話声とメロディから高品質な歌声合成を目的とした。話声から歌声へ合成する手法を提案した。話声から歌声の高品質な歌声合成手法を実現するために、スペクトルの時間方向、周波数方向の制御を提案した。そして、話声とメロディから歌声の F0、スペクトル、非周期成分を制御することにより歌声合成した。

手法による歌声合成音を聴取実験により知覚的に評価した。聴取実験の方法としては、被験者にわかりやすく人数、被験者の人数が少なくてすむシェッフェの一対比較法を採用した。聴取実験により明らかにしたいことは、スペクトル制御モデルの有用性と話声から歌声が合成できているかである。

スペクトル制御モデルの有用性は、手法により作成されたすべての合成音が話声に比べて歌声らしいという評価からあることより証明された。スペクトルの周波数方向の制御の有用性はスペクトルの時間方向と周波数方向を制御した歌声合成音 freqchange、all がスペクトルの時間方向を制御した歌声合成より歌声らしいことから示された。スペクトルの時間方向と周波数方向を制御した歌声合成音 freqchange、all がスペクトルの時間方向を制御した歌声合成より歌声らしいという結果からスペクトルの周波数方向の制御の有用性が示された。ヴィブラート部の F0 に対応した揺れを同期して F1 の微細変動として F1 に付け加えるスペクトルの周波数方向の制御の有用性は all が freqchange より歌声らしくないと結果から示すことができなかった。これは freqchange を行った時点で基本周

波数によりスペクトルを制御している影響かもしれない。歌声合成音が話声から歌声へスペクトル制御において振幅方向を制御していないことにより、実際の歌声より高域の振幅が小さいことによりあまり周波数方向の変換が効いていないことによる起因しているのではないかと考えられる。

話声から歌声を合成する手法で作成された歌声合成音は話声よりはるかに歌声らしいという聴取実験の結果から、話声とメロディから歌声を合成することはできることが証明された。

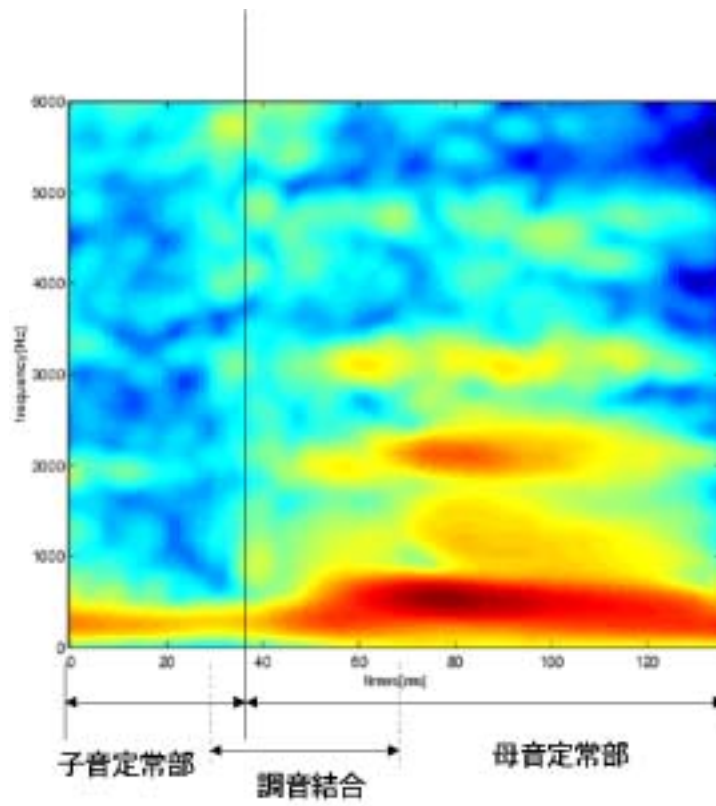


図 4.1: /ze/ の話声のスペクトル

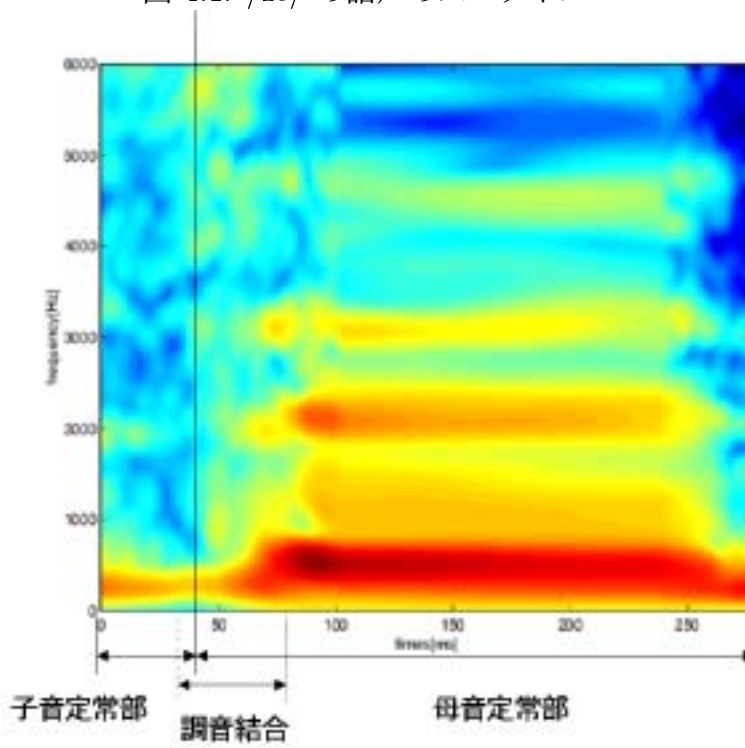


図 4.2: /ze/ の実際の歌声のスペクトル

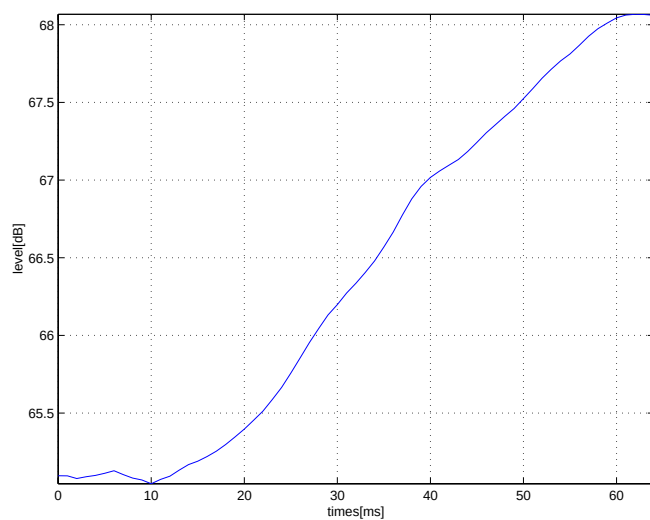


図 4.3: 8192 点の FFT で分析された話声における /karasu/ の /ra/ の /a/ の周波数方向における 100 チャンネル目の時間に対する振幅

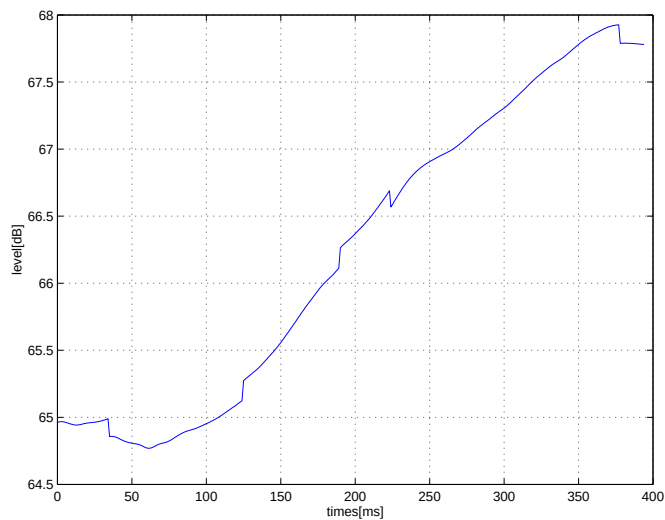


図 4.4: 8192 点の FFT で分析された話声における /karasu/ の /ra/ の /a/ の周波数方向における 100 チャンネル目を時間方向へスプライン補間したものの時間に対する振幅

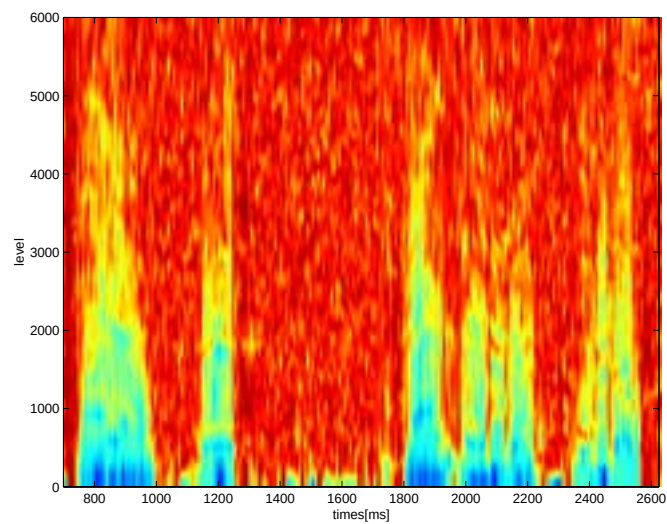


図 4.5: /karasunazenakuno/ の話声の非周期成分

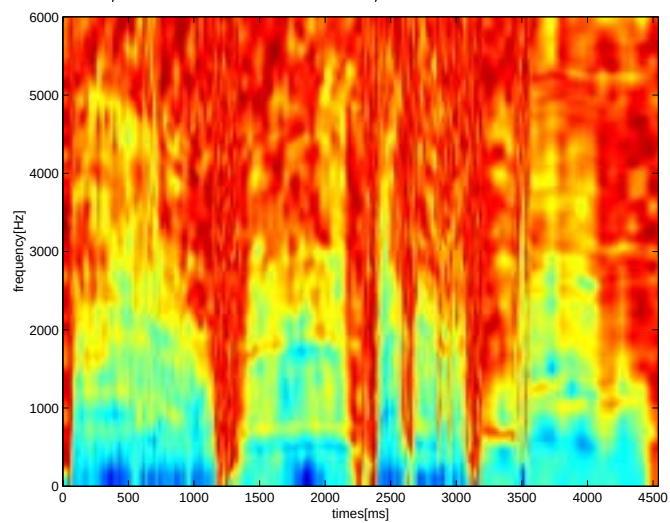


図 4.6: /karasunazenakuno/ の話声の非周期成分を時間方向へ制御を行ったもの

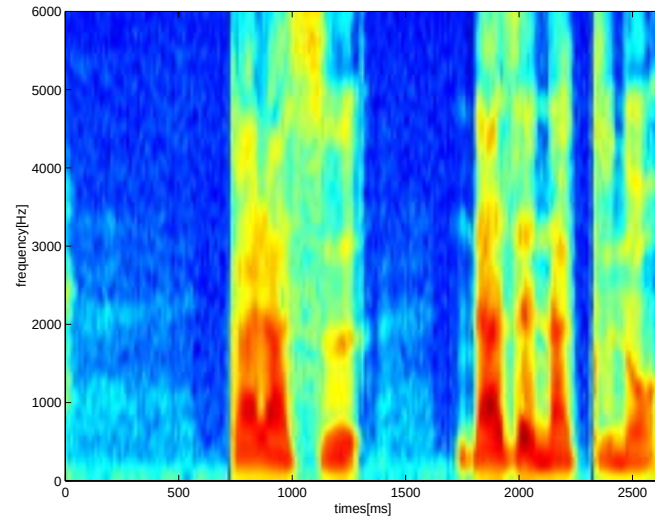


図 4.7: /karasunaznakuno/ の話声のスペクトル

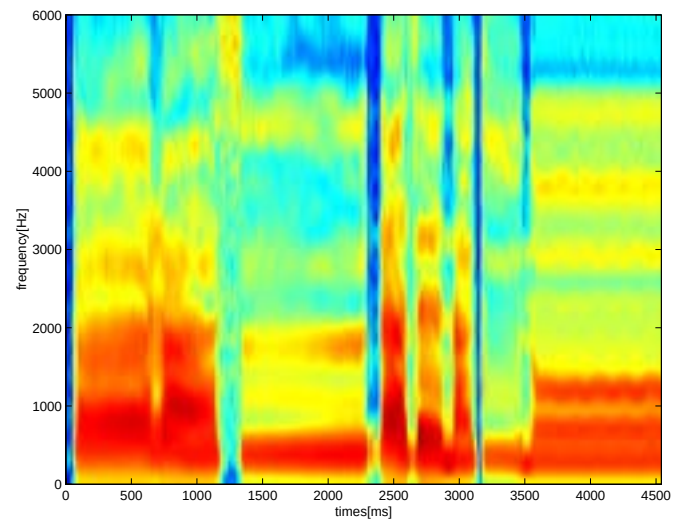


図 4.8: /karasunaznakuno/ の話声のスペクトルを時間制御したもの

スペクトルの周波数方向の制御

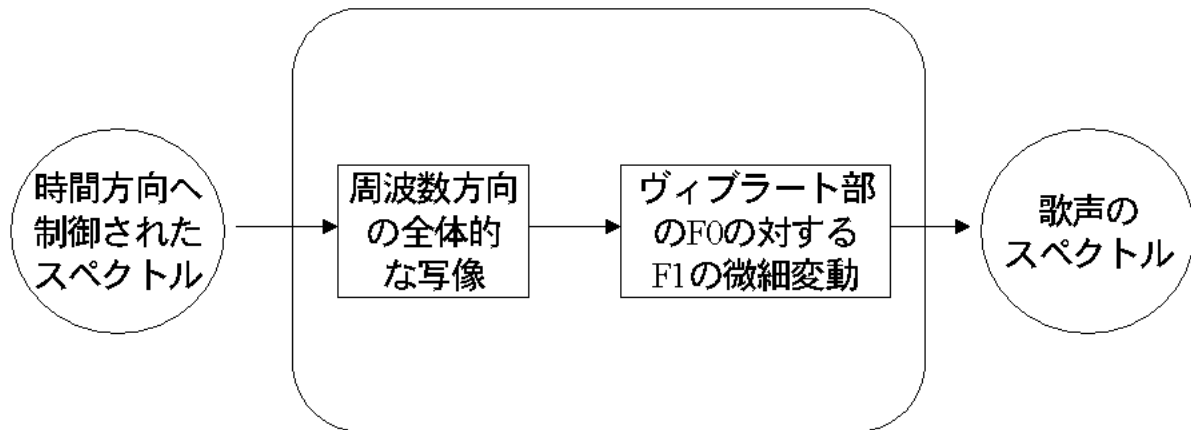


図 4.9: 周波数方向の制御の流れ

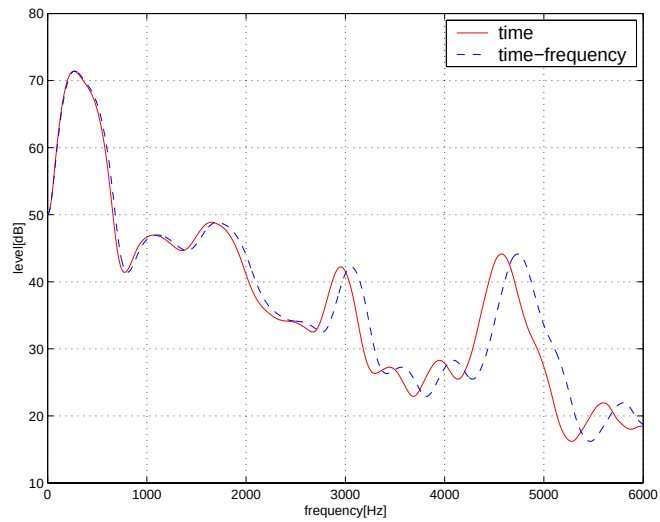


図 4.10: /karasu/ の /u/ における時間制御を行った歌声スペクトルと時間方向と周波数方向の制御を行った歌声のスペクトル

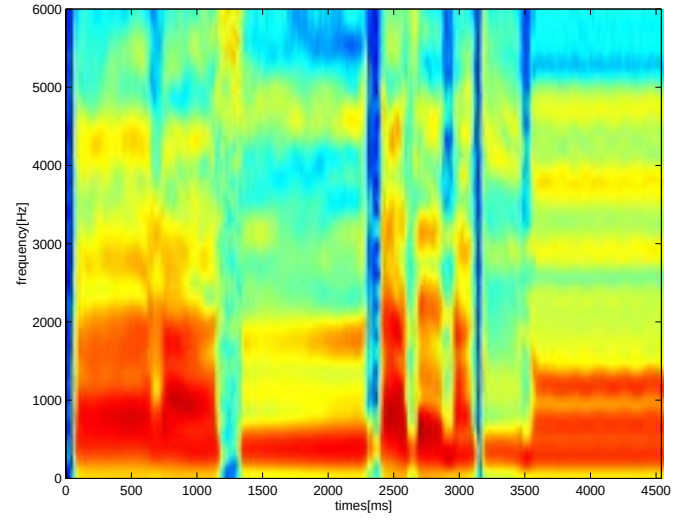


図 4.11: スペクトル制御を行った歌声のスペクトル

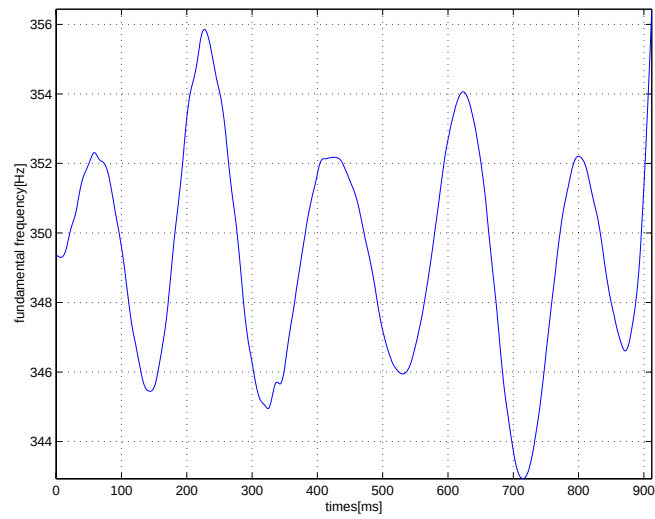


図 4.12: /karasu/ の /u/ の F0

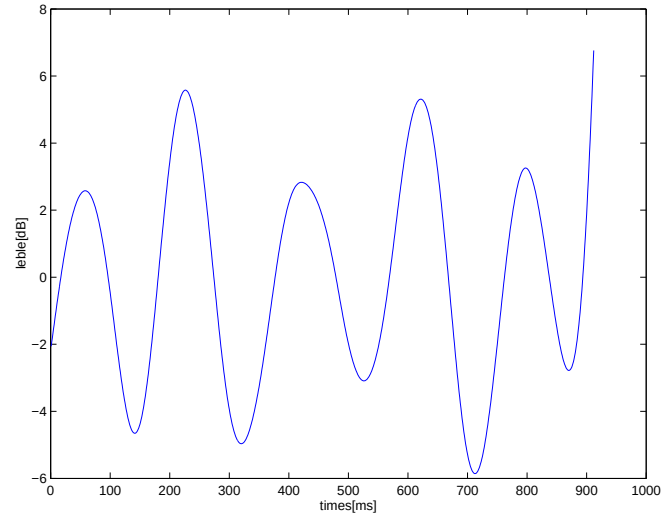


図 4.13: /karasu/ の /u/ ヴィブラート部の F0

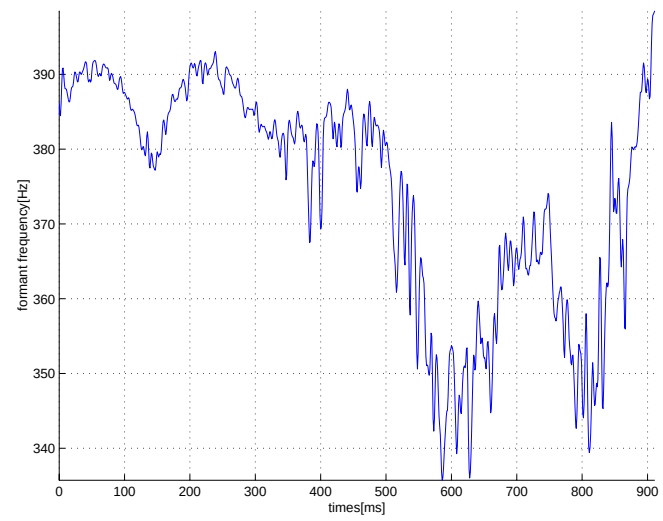


図 4.14: ヴィブラート部の F0 に対応した揺れを付け加えた F1

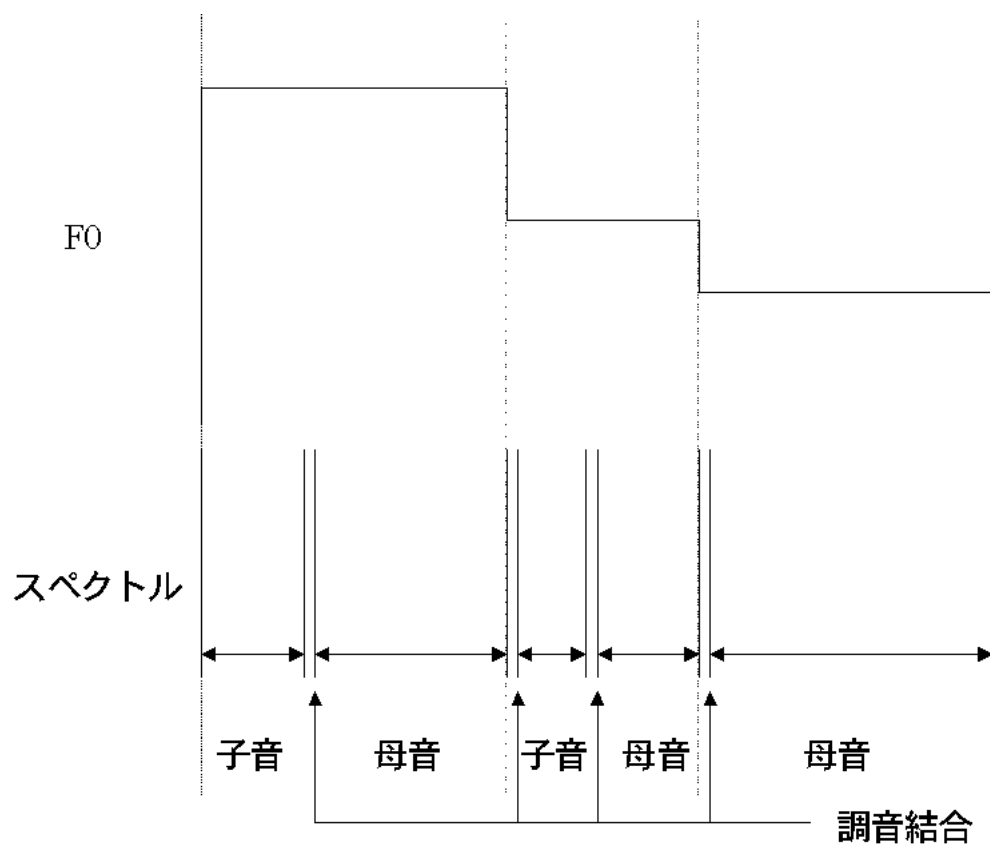


図 4.15: F0 とスペクトルの時間関係

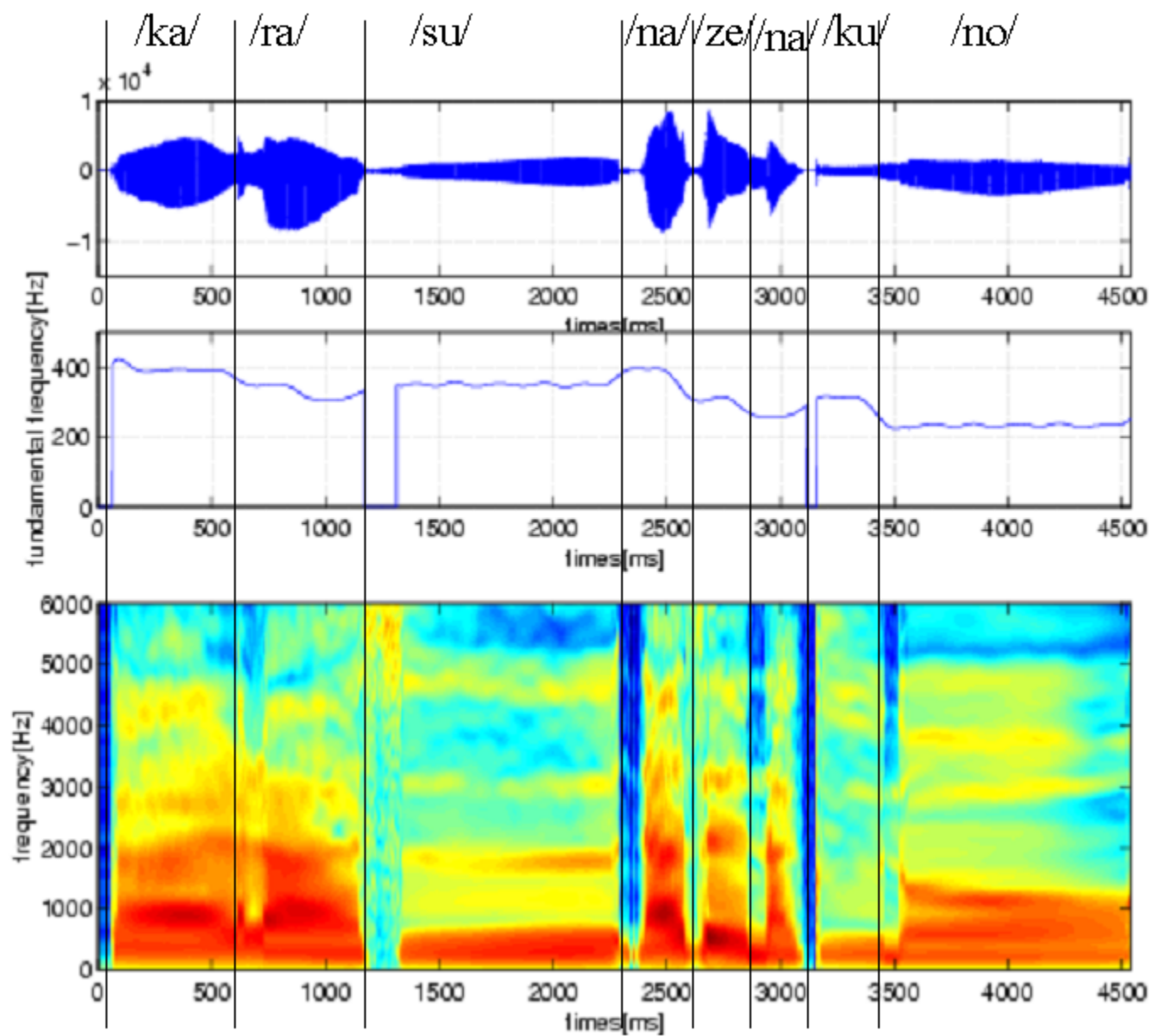


図 4.16: モデルに作成された歌声の時間波形、F0、スペクトル包絡

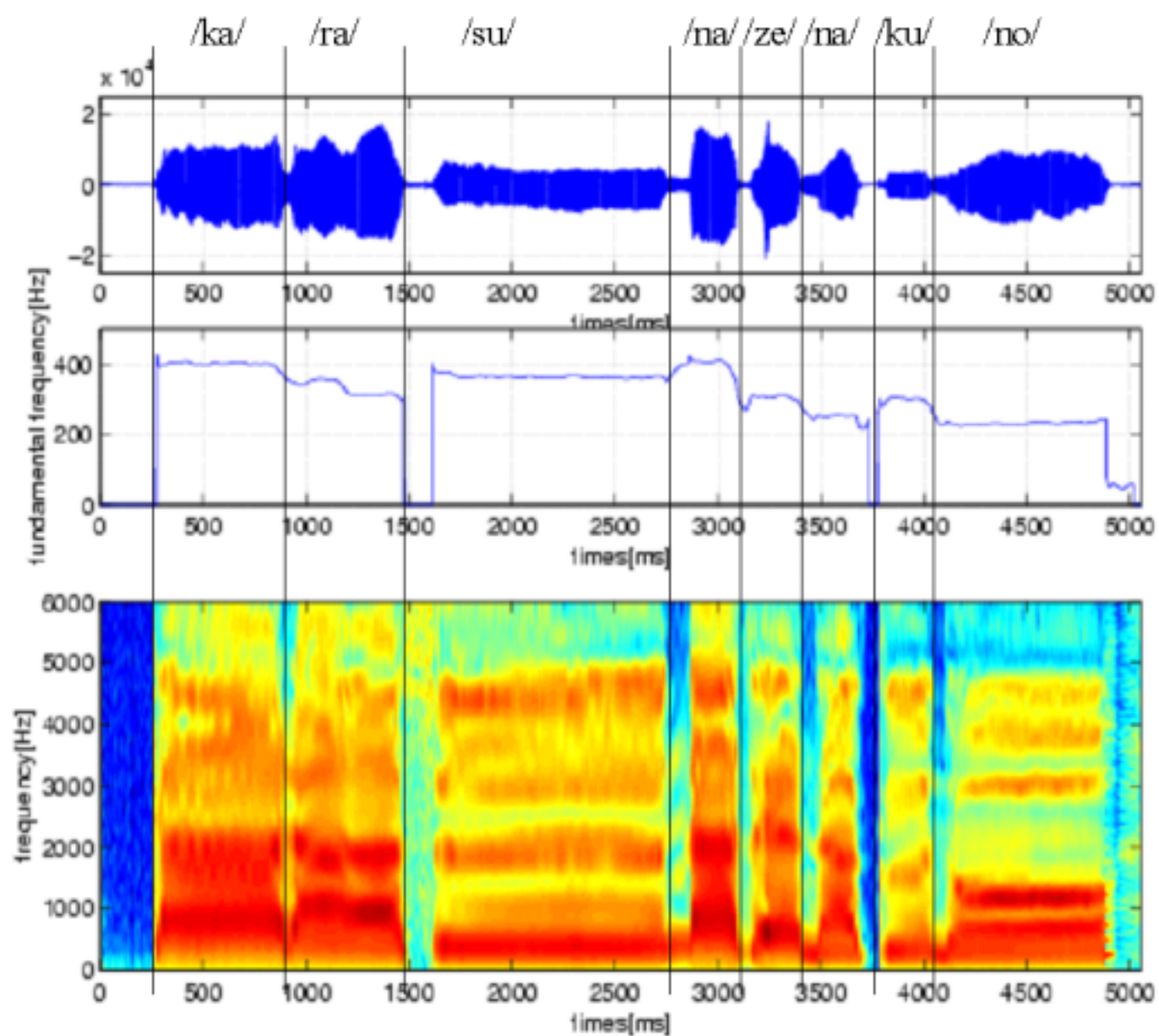


図 4.17: 歌声の分析合成音の時間波形、F0、スペクトル包絡

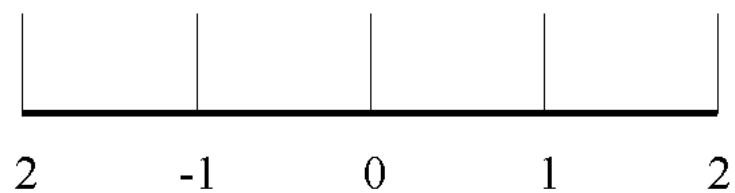


図 4.18: 評価尺度

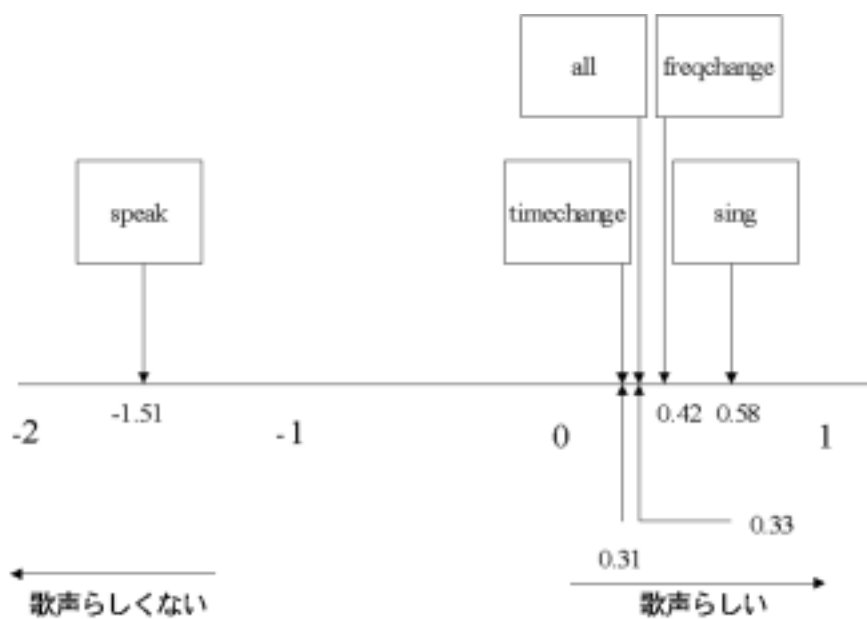


図 4.19: 実験結果

第5章 結論

5.1 本研究の要約

話声とメロディから 歌声の F_0 、スペクトル包絡を制御することにより高品質な歌声合成を実現することを目的とする。話声から歌声のより高品質な歌声を合成するためには、声の高さを制御するパラメータである F_0 とともに、声の音色などに関係するスペクトル包絡も制御しなければいけない。話声から歌声を合成するために、歌声の F_0 制御は齋藤らの F_0 制御モデルで制御できるので、歌声のスペクトル包絡制御をどのように行うかが問題である。人間は歌を歌うとき自分の音色を保ちつつ、メロディにあわせ音程を激しく変化させている。つまり、人間は歌を歌うとき基本周波数を大きく変化させながら音色をたもっている。音色を保つためには基本周波数の変化にともないスペクトルが変化しているのではないかと考えられる。そこで、基本周波数の変化に対応してスペクトルを制御することを考える。

本研究で行った大きなことは以下の 2 つである。

1. 話声とメロディから 歌声の F_0 、スペクトル包絡を制御することにより歌声を合成したことである。
2. 高品質な歌声合成をするために、スペクトルの時間方向、周波数方向の制御を提案した。

スペクトルの時間方向の制御に関しては話声と歌声の時間の長さが異なるので、話声と歌声のデータの子音の時間の長さを分析し、話声の時間の長さに対する歌声の時間の長さを求めて、歌声の子音の時間が話声の時間から計算できるようにした。歌声の子音の時間が決定すれば、母音の時間の長さはメロディの音符から子音の時間の長さを引いたものになる。そして、歌声の子音は話声の子音から分析で得られた結果から時間の長さを計算し、歌声の母音はメロディにより歌声での音節の長さから子音の時間の長さを引いた時間の長さである。そして、求められた時間の長さを持つ歌声のスペクトルは、分析された話声のスペクトルを各周波数方向のチャンネルごとに伸ばすことにより得られる。

スペクトルの周波数方向の制御に関しては F_0 に対してスペクトルを制御できるのではないかと考え、 F_0 に対する話声と歌声の フォルマントを計算した。そして、話声のスペクトルと話声の F_0 に対する歌声の F_0 から歌声のスペクトルを算出する式 $FSI = 0.13\bar{F}_0 FSP + 0.84FSP$ (話声の F_0 に対するまた、ヴィブラート部の F_0 と F_1 の微細変動に時間ラグなしに相関がみられるという分析結果を得た。分析結果から、スペクトルの

周波数方向の制御は時間制御された歌声のスペクトルに $FSI = 0.13\bar{F}_0FSP + 0.84FSP$ の周波数方向の制御を行い、ヴィブラート部の F0 の揺れを F1 に付け加える周波数方向の制御を行うこととした。

話声から歌声を作成するという手法により合成した歌声合成音を聴取実験により歌声らしさを評価した。そして、手法により作成された合成音の中では話声のスペクトルの母音、子音の定常部を時間方向にスプライン補間し $FSI = 0.13\bar{F}_0FSP + 0.84FSP$ の式で話声の F0 に対する歌声の F0 である $\bar{F}_0$ と話声のスペクトル FSP において歌声のスペクトル FSI を算出するという周波数方向への制御を行い作成された歌声のスペクトルを用いた歌声合成音が最も歌声らしいという結果からスペクトルの周波数方向の制御が有効性が示せた。ヴィブラート部の F0 と対応した F1 の微細変動を付け加えた歌声合成音が付け加えていないものより歌声らしくないという結果が出て、ヴィブラート部の F0 と対応した F1 の微細変動の有効性を示せなかった。

また、すべてのモデルの歌声合成音が話声よりかなり歌声らしいという結果からスペクトル制御の有効性を示せ、話声から歌声を合成する手法を提案することができた。

5.2 今後の課題

- 今回の研究ではスペクトル包絡制御において振幅方向の制御が行われていない。スペクトル包絡制御の振幅方向の制御を行うことは、歌声の高域の周波数における振幅が話声より大きいと考えられるので、よりスペクトルの周波数方向の制御を生かすことができる。スペクトル包絡制御がより高品質な歌声合成の助けになると考えられる。
- 聴取実験の被験者の数、分析した話声のデータの種類の数、歌声のデータの種類の数が少ないものである。よって、より分析したものではない多くの話声データとメロディから歌声が合成できるモデルが適応することができるかを調べる必要がある。また、スペクトル包絡の制御も歌声を合成するために多くのデータに適応して行われているかどうかを検証する必要がある。
- 提案したスペクトル包絡制御モデルから話声と歌声のスペクトルの違いを検証し、また生理学的にはどのような意味あいをもつかを考慮することである。そして、話声、歌声に適応した生理学的モデルに発展させ、人間はどのように歌っているのかを解明することである。
- 齋藤らの F0 制御モデルは母音 /a/ のみで考えている。歌詞をつけた歌声は子音を含んでいるので、F0 制御モデルのオーバーシュート、予備的変動のパラメータを見直すことにより高品質な歌声合成が提案できるものと考えられる。

謝辞

研究を遂行するあたり、多大なご指導、ご鞭撻を賜りました赤木正人教授に深く感謝いたします。

また、貴重なご助言を頂きました党建武助教授、鵜木祐史助手に深く感謝いたします。ご多忙のなか、聴取実験にご参加頂いた方々に深く感謝いたします。研究の手助けや協力をして頂いた赤木研究室の皆様にも深く感謝いたします。学生生活を支えて頂いたすべての方々に深く感謝をいたします。

参考文献

- [1] 齋藤毅. 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修士論文, 2001.
- [2] 中山一郎, 小林範子. 歌の声, 日本音響学会 52 巻 5 号, pp.383-388,1996.
- [3] 寺西立年, 大串健吾, 宮崎謙一監訳. ドイツ音楽の心理学 (上), 西村書店, pp72-pp.117,1987.
- [4] 新美成二. 第 2 版 声の検査法 基礎編, 日本音声言語医学会編, pp.173-pp.182,1994.
- [5] 高良富夫, 國吉徹也, 木下正仁, 泉和人, 長山格. ケプストラム法を用いた歌声合成におけるピッチパターンとスペクトルデータの検討, 信学技報, SP99-58,1999.
- [6] Slawson,A.W. Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. Journal of the Acoustical Society of America , 1968,43,87-101
- [7] 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.
- [8] 西澤信行, 広瀬啓吉. 基本周波数の影響を考慮したフォルマント音声合成日本音響学会講演論文, 1-7-10,2000.
- [9] 河原英紀. 高品質音声分析変換合成法 STRAIGHT, ATR 人間情報通信研究所, 和歌山大学, 2002.
- [10] 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.
- [11] 難波精一郎, 桑野園子. 音の評価のための心理学的測定法, コロナ社, 1998.