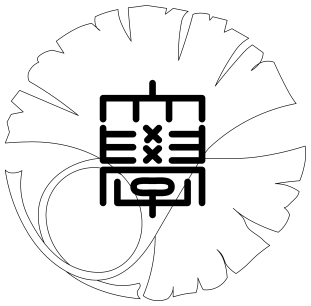
修士論文

対話型進化計算を用いた
コンピュータ歌唱の表情付け



2010年2月9日

指導教員　　伊庭 斉志 教授

東京大学大学院 工学系研究科

電気系工学専攻
37-086539

渡辺 晃生

Abstract

Today, researches for singing by computer have attracted attention. VOCALOID is an application to realize that aim. By inputing lyrics and melody, users can make songs sung by the computer. In order to make the singing voice more “ human ”, users must control frequency curve very carefully. Comparing with inputing lyrics or melody, this controlling presents heavy overhead for users. In this research, we propose a system for easily optimizing frequency curves with Interactive Evolutionary Computation (IEC). We compared various frequency model by evaluating the convergence performance of GA to fit one of the frequency curve of real human singing, and found a suitable one for achieving our goal. And we made a questionnaire to compare previous interface and our IEC interface. From the result of this questionnaire, we found our IEC interface can optimize frequency curve more easily than previous interface.

内容概要

今日，コンピュータによる歌唱が注目を集めている．VOCALOID はそれを実現したアプリケーションの一つである．歌詞とメロディラインを入力することで，ユーザはコンピュータに自作の歌を歌わせることができる．しかし，より人間らしい歌声を作るためには，メロディラインに乗せる周波数曲線の繊細な調整が必要である．これは歌詞，メロディラインの入力と比べて大きなユーザ負担となる．そこで私は対話型進化計算 (Interactive Evolutionary Computation:IEC) を用いて簡単に周波数曲線の最適化を行うシステムを提案する．この研究において私はいくつかの周波数モデルを試験的に導入し，遺伝的アルゴリズム (Genetic Algorithm:GA) を用いてモデル中のパラメータを最適化して人間の歌声の周波数曲線を再現するという実験を通して各モデルの収束性を調べ，少ない計算量で再現が可能なモデルを提案した．さらに調整の簡便さについて，VOCALOID 上の既存インタフェースと IEC インタフェースをアンケート評価により比較した．このアンケートの結果から，IEC インタフェースは周波数曲線の調整に優れたものであることが確認できた．

目次

図目次

表目次

第１章

序論

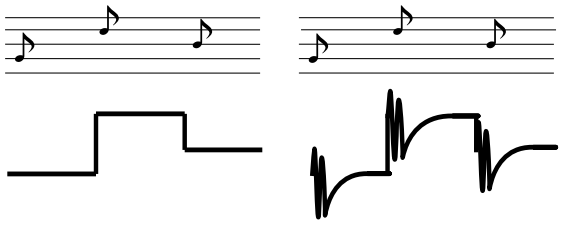
1.1 研究の背景

1.1.1 VOCALOID について

従来，多くの研究者が音楽や歌声について研究を行ってきた．例えば丹治らは，コンピュータにより拍節構造を解析する方法を提案した [12]．そしてそれらを基に,「コンピュータによる歌唱」についてもいくつかの先行研究が存在する．VOCALOID [1][1]はコンピュータ歌唱を実現したアプリケーションである．この VOCALOID において，ユーザは歌詞とメロディラインを与えることでコンピュータに歌を歌わせることができる．これまで，作曲の支援をするようなアプリケーションはいくつかあったが [2]，VOCALOID により曲だけではなく歌詞のついた歌を作ることが可能になったことから，ユーザの表現の幅が広がった．今日では多くのユーザによって歌が作られ，WEB 上に公開されており，ユーザ同士でこれらを評価しあうというような現象も起きており，コンピュータ歌唱への注目が伺える．

1.1.2 コンピュータ歌唱における周波数曲線

より人間らしく，表情豊かな歌声を目指す場合，歌のメロディライン上にのせる周波数曲線の繊細な調整が必要になってくる．しかしこの調整には歌声に関する知識が必要であり，一般のユーザにとっては非常に難しいタスクであった．WEB 上のほとんどの VOCALOID 曲に関してこの周波数曲線の調整が全く成されていないことからも調整の複雑さが伺える．図 1.1 が周波数曲線のイメージ図である．(a) が未調整のもので，(b) が調整したものの例である．人間の歌声の周波数曲線にはオーバーシュートやビブラート等の特徴がある [3] ため，もし未調整の (a) のような周波数曲線の歌を聴いた場合，コンピュータ歌唱であることは容易に判別出来てしまう．



(a) 未調整　　(b) 調整済

図 1.1: 周波数曲線の調整例

1.2 周波数曲線の調整に関する従来研究

この周波数曲線の調整に関する従来の研究を紹介する．

[1] http://www.crypton.co.jp/mp/pages/prod/vocaloid/

VocaListener

中野らは，人間の歌唱を真似することによって人間的特徴をもった周波数曲線を VOCALOID 曲に付与する VocaListener というシステムを開発した [4]．このシステムは周波数曲線を人間の歌声から抽出したものと比較して反復更新する．VocaListener によって調整された VOCALOID 曲はウェブ上でも非常に人間らしいと高い評価を得ている．

ところがこのシステムを使う場合，コンピュータに歌わせる曲を人間が歌ったものを用意する必要があり，通常の場合ユーザは自分で作った歌を歌うことを要求される．これはユーザにとって負担であるとともに，コンピュータ歌唱を求められる状況において結局人間の歌唱が必要であることは大きな制約であると考えられ，全体としてボイスチェンジャーのように声色を変える役割にしかなっていないという意見もある．

SingBySpeaking

斎藤らは，人間の話声から歌声を作る SingBySpeaking というシステムを開発した [11]．このシステムの目的は人間の声をベースにしない VOCALOID とは異なるが，歌声を作る際に斎藤らは，ある周波数モデルを一つ定め，これを全ての音符に付与した．

この研究における周波数モデルは人間の歌声に関する研究に則ったものであり，その意味ではある程度の人間らしさを付与できると考えられる．しかし実際の人間の歌声には様々な周波数曲線のパターンがあり，ただ一つの曲線で再現できる物では無いと考える．SingBySpeaking は同じ周波数曲線を曲中の全音符にあてはめるため，全ての音符が同じ表情付けをされてしまう．同じく斎藤らによって，周波数曲線におけるパラメータ (後述) は同一人物でも曲中で変化することが報告されており，単一の周波数曲線パターンでは不十分であると考えられる．図 1.2 は人間の歌声の周波数曲線パターンの例である．基本周波数成分は除去しているため，音階は見えていない．横軸は時間であり，この中に 13 音符が含まれている．これを見ても音符毎に様々に曲線が変化していることが分かる．

1.3　本研究の目的

従来研究を踏まえ，この研究では，人間の歌声を用いず，音符毎に周波数曲線を調整できるシステムを提案する．この際，歌声の知識の無いユーザでも調整が簡単に行えることを目指す．

1.4　本論文の構成

本論文は全 8 章で構成されている．
各章の内容は以下の通りである．

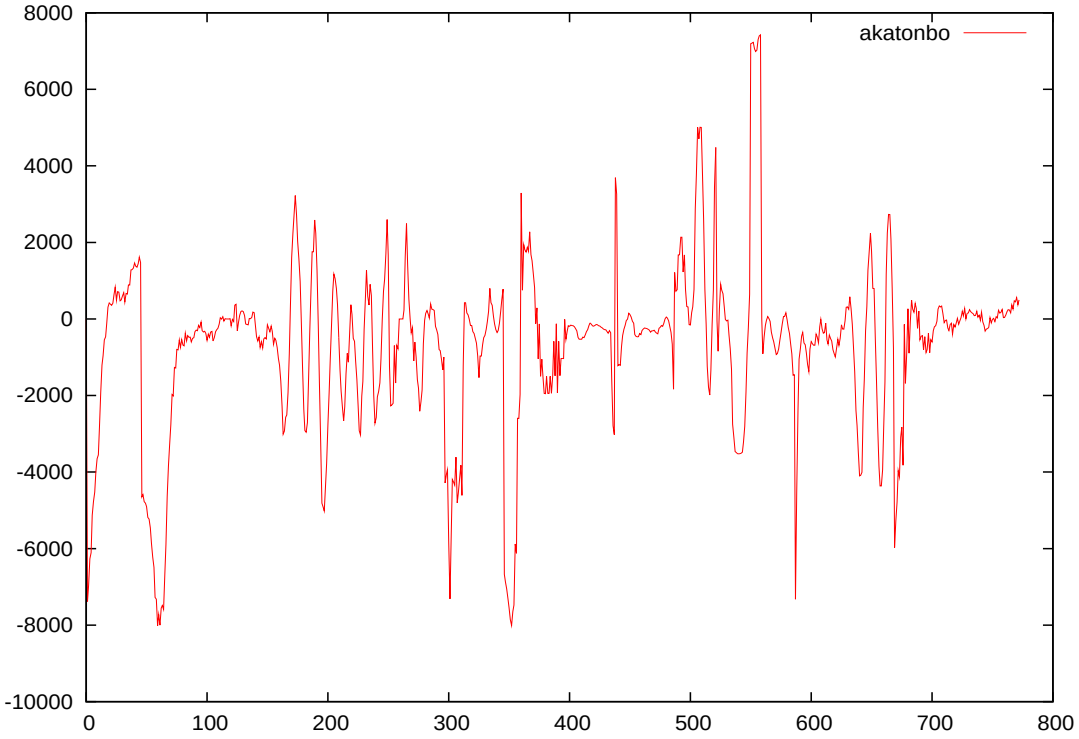


図 1.2: 人間の歌声における周波数曲線の例

第 1 章 序論

本章である．研究の背景を述べ，目的を明らかにする．

第 2 章　IEC の概要と特徴

本研究で用いた手法である IEC の概要とその特徴について述べ，この手法における解決すべき問題について触れる．

第 3 章　提案システム

提案システムの仕様とそのインタフェースにおける特徴について述べる．

第 4 章　周波数モデル

IEC を用いて最適化をする周波数モデルに関して検討する．その後モデルの評価実験を行い，結果を考察する．

第 5 章　提案システムの収束性の検証

局所探索との比較から，IEC における GA 探索の有効性を示す．また，人間が評価する場合のノイズが収束性に与える影響を調べる．

第 6 章　インタフェース

従来インタフェースの問題点を述べた後に従来インタフェースと IEC インタフェースの比較をアンケート評価によって行い，実際の使用感を調べ，結果を示す．

第 7 章　考察

実験から得られた結果を基にして考察を行う．

第 8 章　まとめ

まとめを行い，本研究から得られた知見について述べる．最後に，今後の課題・展望を記す．

第２章

IECの概要と特徴

2.1　IEC の概要

この研究で用いた対話型進化計算 (Interactive Evolutionary Computation:IEC) について説明する．IEC とは，ユーザとコンピュータの相互のやりとりによって，従来コンピュータでの解決が不可能とされてきた作曲や作画などといったタスクを可能にする手法である．この手法は遺伝的アルゴリズム (Genetic Algorithm:GA) に基づいたものであるため，まず GA について説明する．

2.1.1　遺伝的アルゴリズム (GA)

GA とは生物の進化における交叉や突然変異を模して解探索を行う手法である．具体的な手順を説明する．2.1 が GA のフローチャートである．

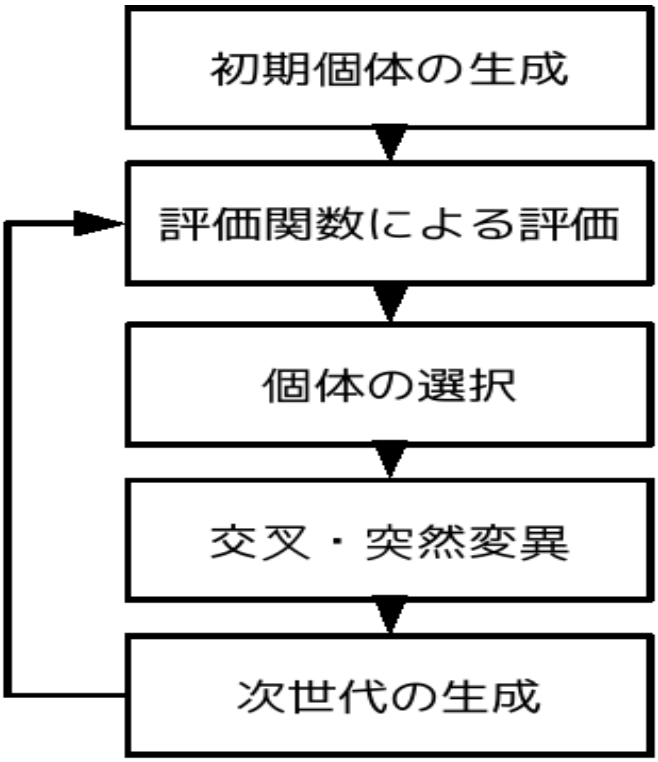


図 2.1: GA のフローチャート

まず，ランダムにいくつかの解をサンプリングする．GA においてはこの解を個体，最初に作った個体群を第一世代と呼ぶ．次にその中からあらかじめ決めた評価関数の良い個体をいくつか選ぶ．これは生物界における自然選択を模している．ここで選ばれた個体のみが後世に遺伝子を残すことが出来る．選択方法には，ルーレット選択，トーナメント選択等があるが，本システムではトーナメント選択を用いた．トーナメント選択とは，個体群から少数の個体をランダムに抽出し，その中で最も優れているものを選択する方法である．また，GA における遺伝子とは個体の持つパラメータの事である．選ばれた個体のパラメータを受け継いで，第一世代と同じ個体数の

次の世代を作る．パラメータを受け継ぐ際，交叉，突然変異という操作を行う．GA における交叉とは，上記の選択法によって選ばれた 2 個体 (以後これを親と呼ぶ) の持つパラメータを参照することで，親の特徴を受け継いだパラメータを作り，子とする操作である．また突然変異とは，子のパラメータのうちいくつかを，ランダムに生成する操作である．これらの操作により親の世代と同じ個体数の子世代を作り，次は子世代に選択，交叉を行うことで次に世代を作る．この操作を繰り返すことによって，世代が進むにつれて徐々に自然界に適応した，すなわち評価関数の高い個体が出来てくる．GA による探索は，多峰性のあるランドスケープにおいて局所解に陥りにくいという性質をもっている．

2.1.2　IEC

図 2.2 が IEC のフローチャートである．コンピュータが生み出した作品をユーザが評価し，高い評価を得た個体を元にして，それと良く似た個体を生み，またそれを評価させるというプロセスを繰り返して完成度を高めていくというシステムである．分かりやすく言えば，通常の GA における評価関数を人間の感覚に置き換えた手法ということが出来る．それ故，評価関数を数式として定義する必要が生じない．

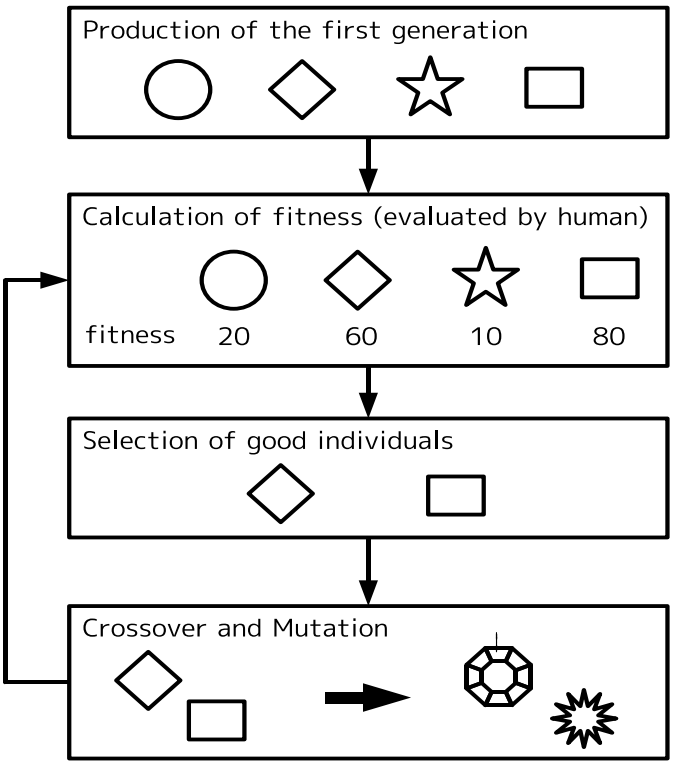


図 2.2: IEC のフローチャート

2.2　IEC の特徴

IEC の特徴として第一に，最適化に評価関数を必要としないことが挙げられる．本研究の歌声の調整の場合において，どのような歌声が良く，どのような歌声が悪いのかということは人の好みや時代によっても変化するため，定義することは不可能であると言える．そのため本研究では IEC のような対話型アプローチが不可欠である．システムを設計する側にとっては，対象とするタスクに関する知識を持たなくてもユーザを満足させるシステムを作ることが可能になる．

コンピュータにより解探索が支援されることにより，ユーザは比較的簡単に最適化問題を解くことができる．本研究でも，マウスでドラッグをして周波数曲線を作成しなければならない既存インタフェースと異なり，どの周波数曲線が良いかを選ぶだけで最適化が成される．人間が解探索を行う場合，多くは局所探索に陥り，あるパターンから抜け出せなくなる場合があるが，IEC システムには発想支援という特徴が含まれており，ユーザが思いつかないような解を生み出せることも大きなメリットの一つである．

デメリットとしてはユーザの疲労度が挙げられる．完全にコンピュータで解ける問題と異なり，IEC システムのユーザは多数回の評価を強いられる．今回の歌声調整というタスクはコンピュータのみで解けるものではないため，ユーザの負担を必要とする．そのため，従来インタフェースと比べて負担を軽減することが本研究の目標である．

また，人の評価は曖昧性を有し，個体の正確な採点が出来ない場合がある．例えば 100 点満点で採点をさせた場合，70 点と 71 点の差は人間には判断できず，両方を 70 点とする場合等である．この事によって進化が遅れた場合，タスク達成まで多くの計算時間が必要になる．これは解くべき問題の解の収束先が粗くてもよく，ある程度の広さをもった大域的最適解で良いというメリットでもある．

2.3　IEC の応用例

音楽の記憶の復元

大まかなイメージは覚えているが細部を忘れてしまった音楽をユーザに思い出させ，その音楽をデータベースから検索するというアプリケーション [16] を紹介する．まずシステムは，あらゆる音楽情報の入ったデータベースから音楽を 5 つ抜き出し，ユーザに聴かせる．それを聴いたユーザは，自分の思い出そうとしている音楽との類似性を評価する．システムはユーザの評価の高い音楽と似たものをデータベースから探し，数個抜き出して再度ユーザに聴かせる．以上のプロセスを繰り返し，最終的にはユーザの求めている音楽をデータベースから抜き出すという仕組みである．

図 2.3 がこのアプリケーションのスクリーンショットである．様々な種類の音楽において，どの音程がどの程度の割合で現れているかを解析すると，音楽の種類ごとに明確な差が見られる．そ

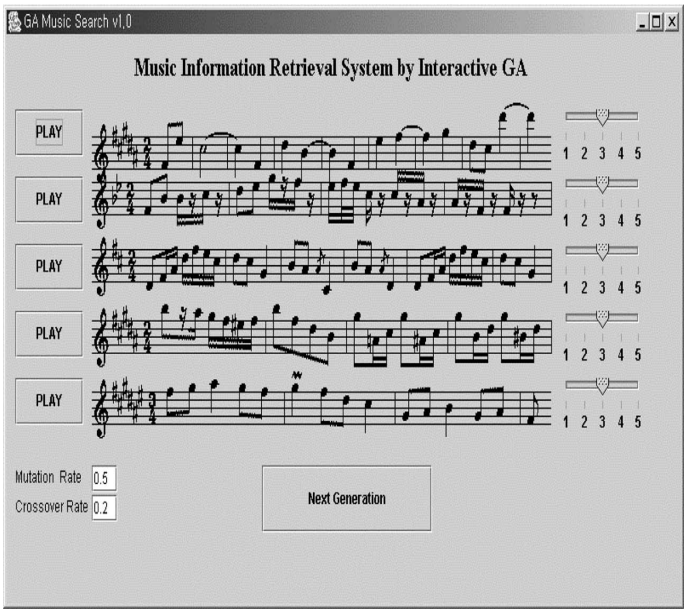


図 2.3: GA Music Search

こでこの音程の現れる頻度の情報を遺伝子に格納し，遺伝子を更新していくことでユーザの求める音楽の種類を探索する．

木をモチーフにしたデザイン

木をモチーフにした描画を行うアプリケーションを紹介する．このアプリケーションはフラクタル図形をベースにし，枝の長さや枝の開く角度等をユーザ好みに最適化することで木をデザインする．ユーザは 8 個体から 2 個体を選択し，その個体の特徴パラメータが子供に引き継がれる．各パラメータは低確率で突然変異する．図 2.4 がこのアプリケーションのスクリーンショットである．

IEC において突然変異は探索範囲を広げるために行われ，例えばこのアプリケーションでは多様な木を生成することが出来るが，あまりにも突然変異の幅が広い，もしくは確率が高いと，全く木らしくない奇妙な物が出来てしまったり，なかなかユーザの求めるデザインに収束しなくなったりする．このことからシステムの設計者は突然変異の確率や範囲に注意をしながら実装をするべきである．

作曲支援システム

安藤らは，IEC を用いた作曲支援システムである CACIE を開発した [9]．このシステムは，プロフェッショナルな音楽家の作曲を支援することは勿論，作曲に関して知識を持たないユーザも本格的な音楽制作を行えることを目的として開発された．このシステムを使った作曲でユーザが

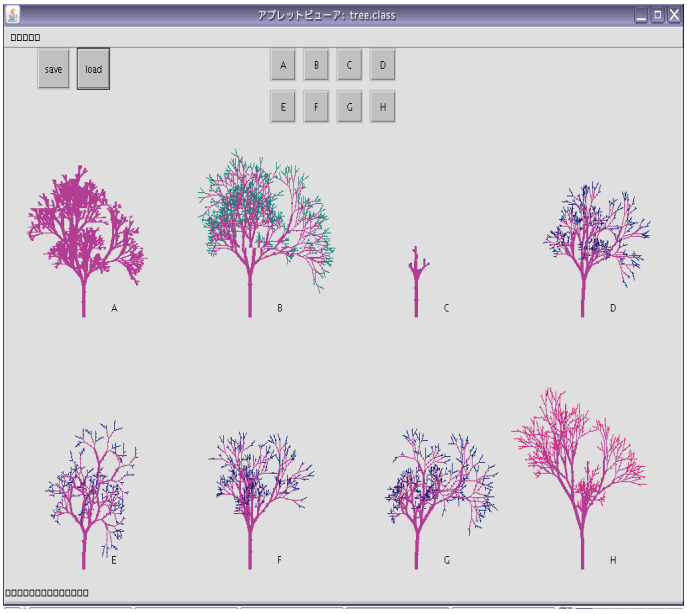


図 2.4: 木のデザイン

すべきことは曲の評価だけであるため，負担がかからず，簡単である．

補聴器のパラメータ調整

音楽やデザインといった芸術分野だけではなく，工学的な応用も行われている．補聴器のパラメータ調整はその一つの例である．補聴器内で行われる信号処理のパラメータは，ユーザ毎に最適なものが異なるため，調整を他人が行うことは困難である．そこで大崎らは IEC を用いて使用者が簡単にパラメータ調整を行えるような技術を導入しようとしている．

2.4　IEC の研究動向

2.4.1　フィードバック機構を用いた IEC

ユーザ負担というデメリットを考え，より少ないユーザの評価回数で満足度の高い個体を作るための IEC の研究が進められている．この目的のためにはユーザの操作から情報を効率良く取得し，システムに反映することが必要である．通常の IEC システムではユーザの評価情報は次の世代を生む際にのみ使われ，2 世代以前の評価情報は全く使われていなかった．Wang らによって開発されたシステムは評価情報をデータベースに格納し，良い評価，悪い評価をされた個体群の特徴を学習する [14]．この学習を Relevant Feedback(RF) と呼ぶ．RF により，同じ情報でもユーザに関して正確な情報を抽出し，システムに与えることが可能になっている．

この手法は学習の為に計算時間が多く必要になることが予想されるが，IEC システムにおける計算時間はユーザが個体を評価する時間が支配的であるため，その間に学習を行うことは十分可

能である．

RF の効果の検証実験

このシステムの評価として Wang らは，20 代の学生 20 人に約 1300 個の風景の入ったデータベースから好みの風景を見つけるという問題において，RF を用いた場合と用いない場合の評価回数を調べた．この実験における RF は以下の働きをする．

1. 評価済みの個体を，良い悪いの 2 値に分類し，記憶しておく．
2. 良い個体として記憶された個体と類似度の高い個体を提示されやすくする．

システムは毎世代 12 個体をユーザに示し，12 個体中 9 個体以上をユーザが良いと判断するまでの世代数の平均を取った．その結果，RF を用いた場合の平均世代数は 4.6 世代，RF を用いない場合の平均世代数は 6.4 世代であった．この結果から，RF を用いた場合のほうが少ない疲労度でユーザの満足度を得られることが分かる．

2.4.2 インタフェースデザインの影響

ユーザが関与する IEC システムの解探索において，インタフェースのデザインは疲労度に大きく影響すると考えられる．IEC においては様々な評価方法，進化方法を用意しておくことが重要である．例えば作曲の場合，全体的なイメージを変えたい場合もあれば，一部分だけを変えたい場合もある．ところが，常に全体を進化させるシステムであれば，変えたくなかった部分までも変化してしまい，ユーザの満足度を低くする原因になる．人によって作曲のプロセスは異なるため，その違いを吸収できるのが望ましい．

また，意味の無い評価を極力減らすことも重要である．上記で紹介した CACIE のインタフェースにおいては，提示されたいくつかの個体のうち類似度の高いものはまとめて評価させるという工夫が成されている．

さらに，見た目に関して，現在の個体の持つパラメータがどのような特徴を持っているかが直感的にユーザに示されていることが重要である．畦原らが開発した作曲支援システム [15] のインタフェースにおいては，全ての音の高さと長さの情報が視覚的に与えられているため，何度も聴き直さなくても，場合によっては全く聴かなくてもメロディを理解することが可能である．また進化においてパラメータの変わった部分を色分けすることで，どの部分に変化が起きたのかを瞬時に把握することが出来る．

図 2.5 が改良前のインタフェース，図 2.6 が改良後のインタフェースである．改良前はいくつかのバーが置いてあり，それをクリックすると音楽が流れるというシンプルなものになっているが，改良後は音楽の情報が視覚的に得られるようになっている．こうしたインタフェースの工夫によるユーザ疲労度の改善が確認されている．

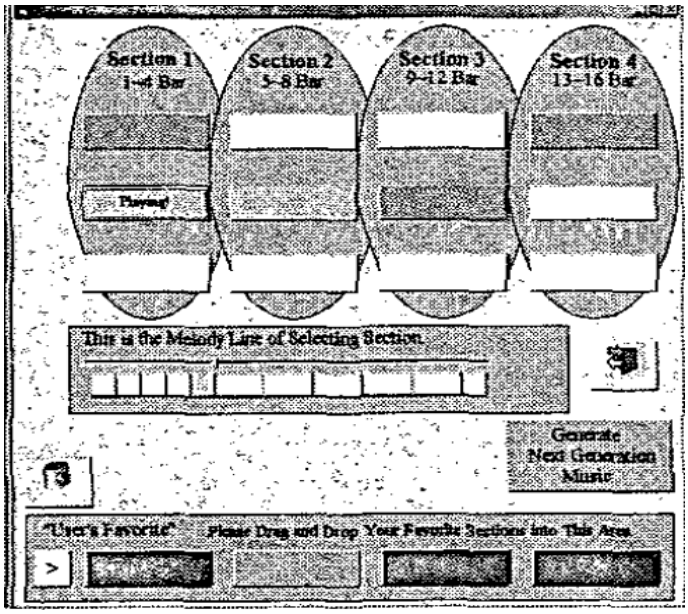


図 2.5: 個体情報の見えないインタフェース

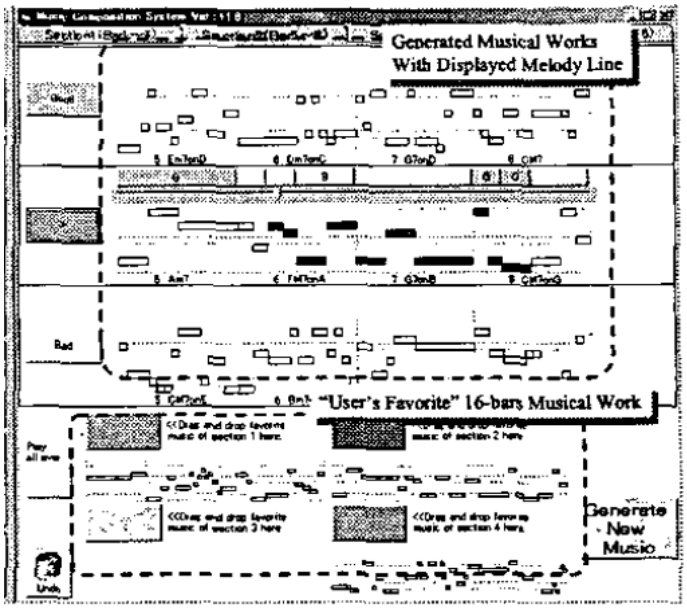


図 2.6: 個体情報の見えるインタフェース

2.4.3 ユーザの評価方法の検討

ユーザの評価方法には，世代の中から好きなものを少数選ぶ方法や，各個体に採点をしていく方法がある．好きなものを選ぶ方法はユーザにとってシンプルであるが，選ぶか選ばないかの 2 値評価になるため，収束効率の悪さが問題である．各個体に採点をする方法は細かい点差を区別できない，比較的ユーザ負担が大きいという問題がある．細かい点差を区別できないことを回避するために，対話型差分進化という手法が提案されている [13]．この手法においてユーザは 2 個体を提示され，どちらが好みかを選ぶだけでよいため，点差が小さくとも評価をしやすく，収束性においてもパフォーマンスの向上が報告されている．

第3章

提案システム

3.1　IEC を用いた本システム実装方法

周波数曲線の調整は，工学的にはある目的関数（ここでは歌声の「よさ」）を最大化する為のパラメータ最適化問題として考えることが出来る．そこで今回 IEC を用いてこの最適化問題を解くシステムを実装した．本研究のシステムにおける IEC の実装方法を具体的に説明する．まずコンピュータにランダムなパラメータを持った 8 つの個体を生成させる．各個体のもつパラメータを特に進化計算において遺伝子と呼ぶ．最初の 8 個体を第 1 世代と呼び，この時点では遺伝子はユーザにとって好ましい物ではない．次に個体の持つパラメータから 8 つの周波数曲線を生成し，VOCALOID 曲に付与する．このようにして作成された 8 つの VOCALOID 曲をユーザに提示し，評価を得る．本システムにおいて評価は 100 点満点でそれぞれの個体を採点するという仕様である．各個体の評価が終わった後，システムはそれを元に次世代の生成を行う．こうして出来た次世代は，ユーザの好みを反映したものになっている．その後，再びその世代の 8 個体の評価を得る．これを繰り返して遺伝子のもつパラメータを最適化していく．

3.2　周波数曲線の付与のしかたについて

元の音階の基本周波数を F0 としたとき，周波数曲線 Fcurve(t) を付与した後の周波数 f(t) は以下のように表される．

$$f(t) \quad = \quad F0 * 2^{Fcurve(t)/8000} \tag{3.1}$$

今回 IEC によって最適化するのはこの Fcurve(t) である．基本周波数が 2 倍になることは，音階が 1 オクターブ上がることを意味する．つまり Fcurve(t) が 8000 増えるたびに音階は 1 オクターブずつ上がる．

3.3　インタフェースの説明

図 3.1 が本システムのインタフェースである．各個体の持つ周波数曲線が画面上に表示され，クリックすることでその曲線が付与された歌声を聴くことができる．ユーザはこれを聴いて，曲線横のバーで個体に評価を与え，評価が終わった後，Reproduce ボタンを押して次世代に進む．

一瞬で各個体の特徴を把握できる絵などと異なり，音楽や動画のような時系列メディアは評価に一定の時間を要し，これがユーザの負担となる場合がある．本インタフェースにおいては曲線を視覚的に把握出来るようにしたことで，聴くまでも無く悪い個体であることが分かる場合など，ユーザに無意味な評価をさせないという効果を期待している．

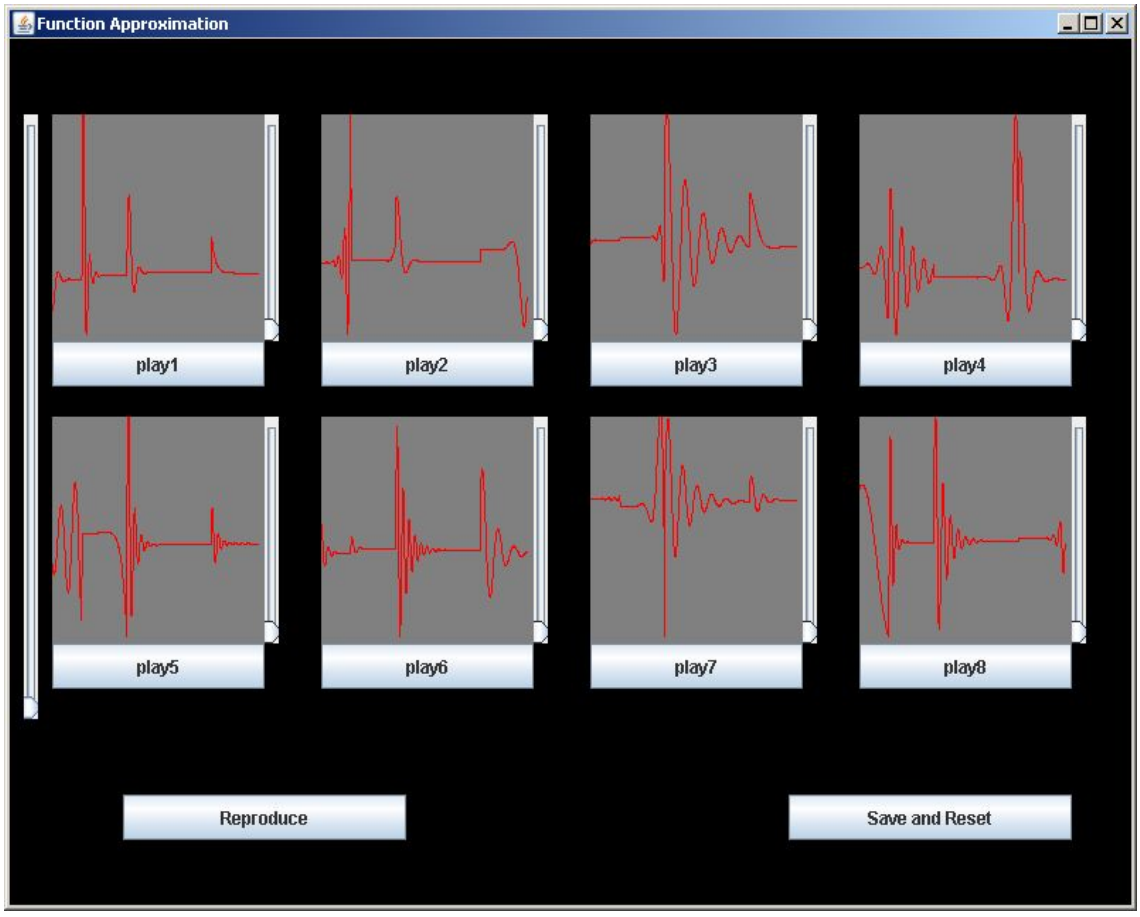


図 3.1: IEC インタフェース

第4章

周波数モデル

既存インタフェースにおいては多くのユーザが周波数曲線の調整を諦めてしまっていた．本研究の主目的はこの調整を簡単化し，一般ユーザでも様々な歌声表現の作成が成せるようにすることである．これは第 2 章で述べた IEC システムを用いることによる問題の簡単化だけでは達成できない場合がある．調整の終了までにユーザは何度かの個体評価を必要とするが，この個体評価の回数によってはユーザに過剰な負担がかかる場合があるからである．

この問題を解消するために，最適化問題における探索空間を適切にとる必要がある．もし探索空間が広すぎる場合，ユーザが求める解を探索するまでに多くの評価回数を必要とする．また，狭すぎる場合，ユーザの求める解が探索空間中に無い可能性がある．このことは人によって最適解の異なる本研究のタスクのような場合に考慮しなければならない問題である．

そこで本章では，適切な周波数モデルを選択することによって探索空間の適切な設定を目指す．モデルにおいて重要な点は以下である．

1. 人間の歌声における周波数パターンを網羅するだけの自由度を含むこと
2. ユーザに負担を与えない程度の評価回数で良い解を得られること

私は本章においていくつかの周波数モデルを提案した．上記の点においてモデルを評価するための実験をした．実験内容は以下である．

1. 人間の歌声から周波数曲線を抽出する．これを目的曲線とする．
2. 各モデル中のパラメータを GA により最適化する．この時使う評価関数は目的曲線との平均絶対誤差である．
3. 各モデル 8 個体 50 世代までにできた曲線と目的曲線との誤差の少ないものを選択する．

今回は目的曲線として，人間によって歌われた童謡「赤とんぼ」の一部分の周波数曲線を抽出した．個体数と世代数の積がユーザの評価回数であるため，この実験により最適解を得られるまでにかかるユーザの負担を推定することが出来る．

4.1　実験したモデルの説明

私は従来研究として紹介した SingBySpeaking において用いられたモデルを参考にし，いくつかのモデルを作成した．

まず，SingBySpeaking において用いられたモデル (以下，斉藤らのモデルと呼ぶ) は次のものである．

$$Fcurve(t) = \frac{k}{\sqrt{1-\zeta^2}} exp(-\zeta\omega t) sin(\sqrt{1-\zeta^2}\omega t) \tag{4.1}$$

ここで t は時間であり，k，ζ，ω は定数である．このモデルによって得られる曲線を図 4.1 に示す．SingBySpeaking はこの曲線を全音符にあてはめることによって人間らしい歌声を目指した．

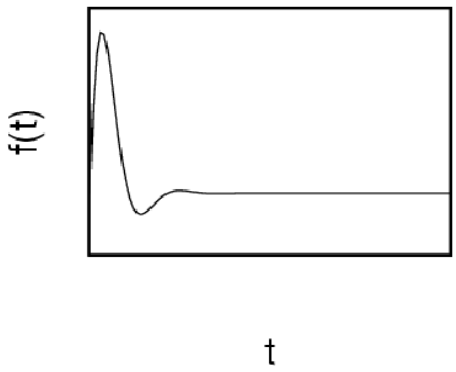


図 4.1: SingBySpeaking 中の周波数曲線

IEC システムによって斉藤らのモデル中における定数を変化させ，音符ごとに最適化することは可能であるが，人間の歌声を再現するには自由度が足りず，改良が必要であると考えた．

解析のために (2)(3)(4) を式 (1) に代入することでモデルを簡単化した．

$$k' = \frac{k}{\sqrt{1-\zeta^2}} \tag{4.2}$$

$$\zeta' = \zeta\omega \tag{4.3}$$

$$\omega' = \sqrt{1-\zeta^2}\omega \tag{4.4}$$

この操作により，式 (5) を得た．

$$Fcurve(t) \quad = \quad k'exp(-\zeta' t)sin(\omega' t) \tag{4.5}$$

式 (5) を見ると，この式はサイン関数のエンベロープを指数関数で表現し，次第に減衰するサインカーブを表現していることが分かる．これを基に以下の 4 つのモデルを作成した．

1. GP エンベロープモデル
2. シンプル GP モデル
3. 3 次多項式エンベロープモデル
4. 5 次多項式モデル

GP エンベロープモデル

式 (5) における指数関数の部分を遺伝的プログラミング (Genetic Programming:GP) [7] により一般的な関数の中から探索するモデルである．エンベロープが指数関数に限定される斉藤らのモ

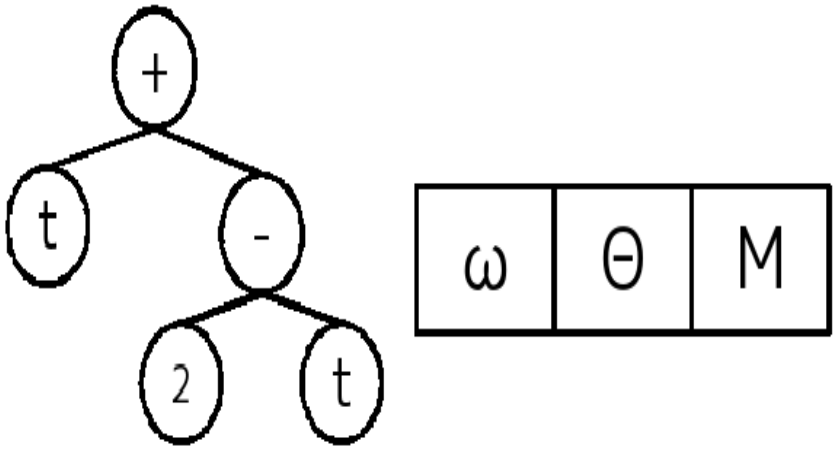


図 4.2: GP エンベロープモデルの遺伝子

デルよりも自由度が高く，多様な曲線を構成可能である．このモデルの遺伝子を図 4.2 により示す．

GP によって構成される木構造の例がこの図の左側に示されている．この例では t と $(2-t)$ という部分木が和の演算子によって結びついているため，全体としては $t+(2-t)$ を意味する．GP の交叉は部分木を他の個体のものと入れ替えることによって行われ，突然変異は部分木をランダムに入れ替えることによって行われる．この研究で使った GP のパラメータを表 4.1 に示す．今回 GP のノードとして t，四則演算，サイン関数，指数関数を用いた．また，モデルのサイン関数の中の波長 ω と位相 θ についても GA によって最適化を行う．これらについては，UNDX [8] という交叉方法を参考にした．図 4.3 が交叉によるパラメータ生成のイメージ図である．今回のシ

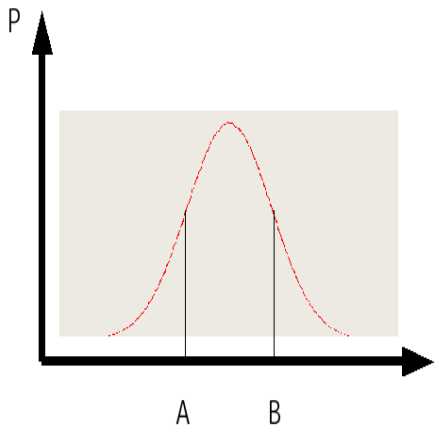


図 4.3: 交叉のイメージ図

ステムにおいては 2 つの親個体のパラメータの中間値を平均，中間値から片方の親までの距離を分散としたガウス分布を確率分布として子個体のパラメータを生起する．GA のパラメータを表

表 4.1: GP Parameter

Population size	8
Selection	tournament
Tournament size	2
Crossover rate	0.8
Mutation rate	0.2
Elite size	1
Max depth	6
Min depth	3

表 4.2: GA Parameter

Population size	8
Selection	tournament(size 2)
Mutation rate	0.1
Elite size	1
Crossover	UNDX

4.2 に示す．このモデル以外においても，GA と GP のパラメータは同様である．

GP によって関数を作成する際，値が非常に大きくなってしまうことが高確率で起こるため，スケーリングを行う必要がある．また，元のメロディを変えないために平均値は 0 である必要がある．そのため遺伝子中に定数 M を加え，以下の操作を行った．

1. 関数全体から，その関数の平均値を引く．この操作で関数の平均は 0 になる．
2. 最も絶対値の大きい点の値が M になるように関数全体に定数を掛ける．

シンプル GP モデル

このモデルは GP によってのみ構成され，GP エンベロープモデルのサイン関数を含むという制限を取ったものである．遺伝子は図 4.4 のように GP の木構造とスケーリングの際に用いる M のみである．

3 次多項式エンベロープモデル

このモデルはサイン関数のエンベロープを 3 次多項式で構成するものである．3 次多項式は，関数中の 4 点を GA によって得た後，それらを補完することによって得る．4 点の t 座標は，音符の始まりの時刻を t=0，終わりの時刻を t=Tend として 0，Tend/3，Tend*2/3，Tend とする．遺伝子はこの 4 時点でのエンベロープの値をそれぞれ k1〜k4 として，図 4.5 のようになる．

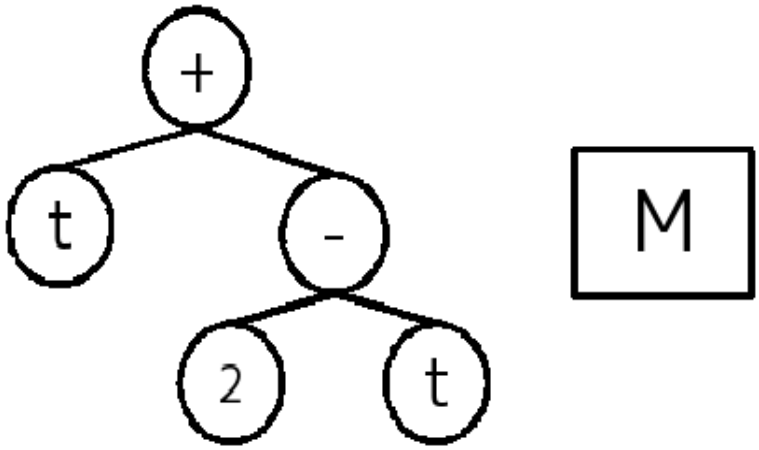


図 4.4: シンプル GP モデルの遺伝子

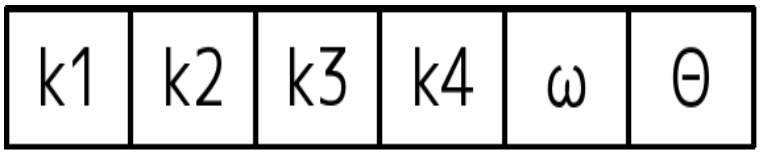


図 4.5: 3 次多項式エンベロープモデルの遺伝子

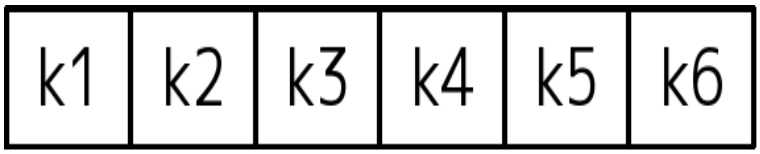


図 4.6: 5 次多項式モデルの遺伝子

5 次多項式モデル

このモデルは 5 次多項式のみによって周波数曲線を決定する．5 次多項式は，関数中の 6 点を GA によって得た後，それらを補完することによって得る．6 点の t 座標は，音符の始まりの時刻を t=0，終わりの時刻を t=Tend として 0，Tend/5，Tend*2/5，Tend*3/5，Tend*4/5，Tend とする．遺伝子はこの 6 時点でのエンベロープの値をそれぞれ k1〜k6 として，図 4.6 のようになる．パラメータの数は 3 次多項式エンベロープモデルと同じで 6 つであるから，最適化にもほぼ同等の計算量がかかると予測される．

表 4.3: 各モデルの平均絶対誤差 (best)

音符	斎藤らのモデル	GP エンベロープモデル	シンプル GP モデル	3 次多項式エンベロープモデル	5 次多項式モデル
yu	1318.51	903.79	880.77	479.76	626.79
u	2277.34	2274.61	1993.25	538.52	1567.04
ya1	1262.02	1015.82	1201.76	951.05	1259.56
ke1	1078.76	722.37	710.62	602.7	672.04
ko	2436.73	2369.60	2374.51	743.33	1930.81
ya2	1682.71	792.20	940.62	655.89	978.54
ke2	573.07	350.07	363.46	404.54	513.39
e	354.66	338.15	227.04	173.6	340.67
no	1661.40	1460.54	1554.43	1288.9	881.6
a	1440.70	1406.04	1406.04	273.08	896.28
ka	1596.59	1426.52	1447.44	720.15	749.13
to	1045.28	888.20	1186.11	877.09	1141.73
n	561.18	421.78	459.01	360.06	536.52
total	17288.93	14369.69	14745.07	8068.65	12094.09
average	1329.92	1105.36	1134.24	620.67	930.31

4.2　各モデルの評価

これら 5 つのモデル (斎藤らのモデルを含む) を使い，人間の歌声の周波数曲線を再現する実験を行った．この実験では,「赤とんぼ」における音符のもつ周波数曲線を，1 音符ずつ最適化している．したがって全音符同時に最適化する場合とは結果が異なる．表 4.3 が各モデルの 50 世代における最優秀個体の目的曲線との平均絶対誤差である．従って値が少ないほど人間の歌声に近いことを意味する．第 3 章で説明したように，8000 で 1 オクターブ分の誤差である．全て 10 回の試行のベストの値である．

図 4.7 が各モデルの平均絶対誤差の世代毎の推移である．縦軸が平均絶対誤差，横軸が世代であり，下に行くほど良い．赤色が GP エンベロープモデル，青色がシンプル GP モデル，ピンク色が 3 次多項式エンベロープモデル，水色が 5 次多項式エンベロープモデルである．

図 4.8-4.12 はそれぞれの音符の最優秀個体を繋いだものである．赤線が目的曲線，青線が得られた優秀個体の曲線である．

表 4.4: 各モデルの平均絶対誤差 (average)

音符	斎藤らのモデル	GP エンベロープモデル	シンプル GP モデル	3 次多項式エンベロープモデル	5 次多項式モデル
yu	1768.98	1074.77	1044.23	822.44	1046.62
u	2292.67	2274.61	2105.80	662.66	2233.42
ya1	1325.82	1072.85	1211.38	1059.47	1432.61
ke1	1142.59	790.22	892.31	659.57	1010.23
ko	2515.31	2402.09	2390.54	1107.25	2381.05
ya2	1682.71	970.3	1168.27	893.19	1277.59
ke2	699.28	401.9	392.78	464.58	763.76
e	473.28	339.14	294.32	240.45	516.29
no	1837.45	1471.67	1556.92	1413.86	957.37
a	1535.01	1406.04	1442.11	463.79	1646.14
ka	1816.62	1520.68	1544.84	978.19	892.74
to	1257.84	1023.26	1216.41	955.4	1365.02
n	614.39	449.45	485.26	422.94	776.37
total	18961.94	15196.96	15745.17	10143.79	16299.2
average	1458.61	1169.00	1211.17	780.29	1253.78

表 4.5: 各モデルの平均絶対誤差 (standard deviation)

音符	斎藤らのモデル	GP エンベロープモデル	シンプル GP モデル	3 次多項式エンベロープモデル	5 次多項式モデル
yu	225.19	147.98	114.93	310.01	382.88
u	48.42	0.00	145.29	90.99	343.94
ya1	69.60	68.29	20.46	87.15	91.78
ke1	201.82	73.78	71.89	156.41	220.95
ko	31.61	29.08	17.55	218.56	296.30
ya2	0.01	165.21	144.01	175.46	265.87
ke2	399.01	86.04	31.32	44.72	200.73
e	251.18	1.96	59.70	57.20	123.44
no	61.86	20.47	3.57	88.99	49.66
a	106.25	0.00	38.88	217.72	469.88
ka	123.2	97.59	89.45	316.12	117.15
to	154.03	102.44	36.33	65.4	136.91
n	118.60	26.81	24.51	60.68	124.16

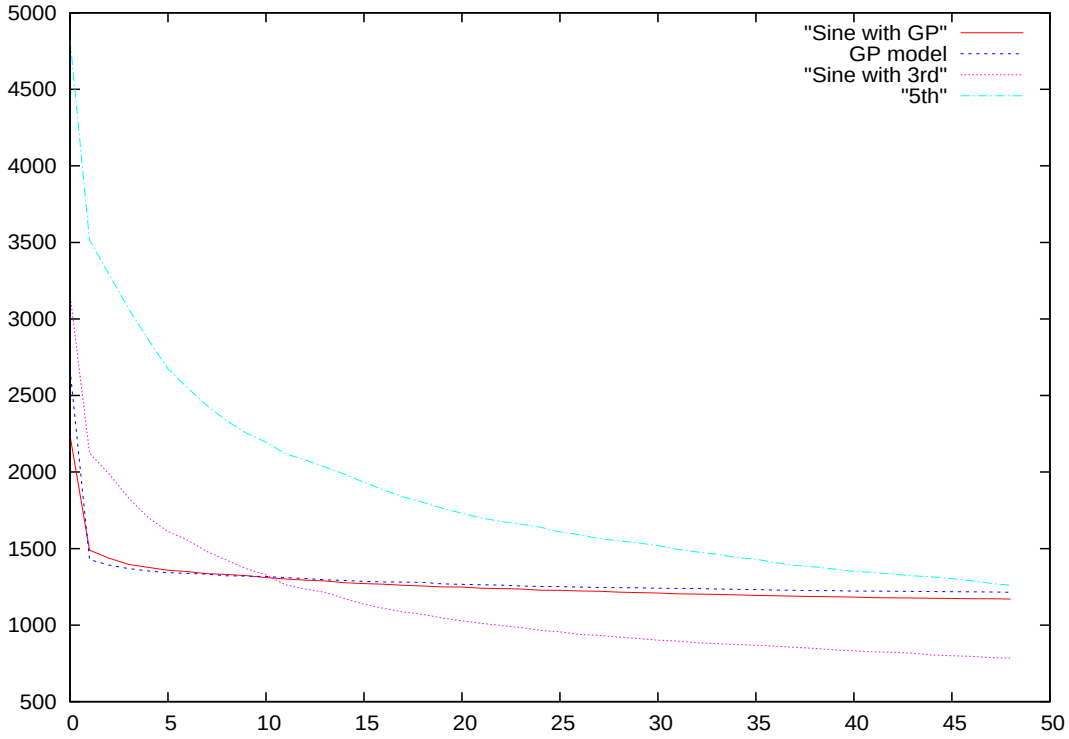


図 4.7: 各モデルの平均絶対誤差の推移 (average)

4.3　最適なモデルの選択

表 4.3 から，3 次多項式エンベロープモデルが最も少ない誤差の個体を出していることが分かる．斎藤らのモデルは自由度が足りず，図 4.8 のように多くの音符が表現できずに Fcurve(t)=0 となってしまっている．また，今回 GP を用いて 2 つのモデルを作成したが，平均絶対誤差を見ると結果はあまり変わらなかった．図 4.7 を見ると 10 世代までは良い結果を示しているが，図 4.9, 図 4.10 を見るとモデル自由度が高いにも関わらず斎藤らのモデルと同様に，いくつかの音符に関しては表現できないまま終わっている．これは探索空間が広すぎたことが原因で少ない世代数，個体数では良い解を探すことが出来なかったと考えられる．以上の結果から，IEC で許される個体数で探索を進めるには探索空間の広さを適切に設定することが非常に重要であることが分かる．図 4.11，図 4.12 を見ると，3 次多項式エンベロープモデルと 5 次多項式モデルに関しては，比較的良く再現が出来ていると言えるが，平均二乗誤差を見ると 3 次多項式エンベロープモデルのほうがより優れた結果を示している．この 2 つのモデルはどちらも 6 つの定数を GA で最適化するという点において計算量は同程度であることが予想されるが，結果に大きな差が出たことから，モデルにサイン関数を含めることは有効であると言える．5 次多項式モデルはサイン関数を含まないことから，ビブラートのある音符に関して特に悪い結果になってしまっており，それが”ya(1)”(図 4.14) や”to”(図 4.15) において見られる．”no”(図 4.13) の結果は最もこのモデルが良くなってい

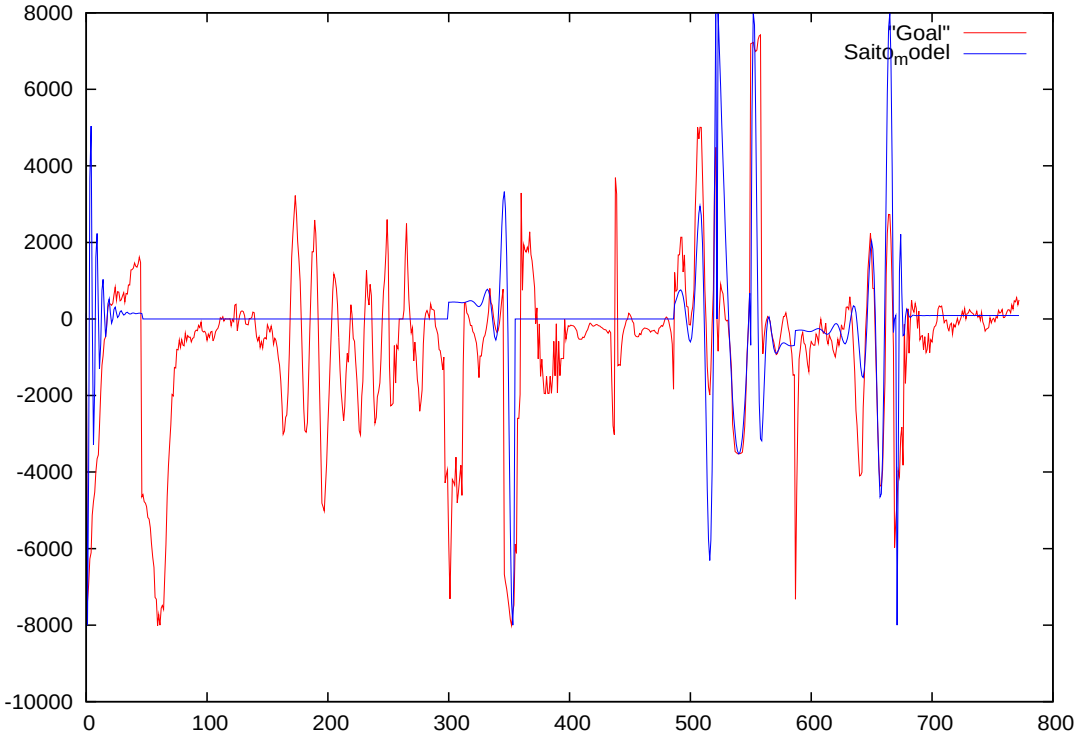


図 4.8: 斎藤らのモデルの最優秀個体

るが，波形を見ると，偶然 5 次多項式で近似可能であったことが分かる．そして表 4.5 から分かる通りこのモデルの標準偏差は比較的高く，またベストと平均の差も大きい．このことは，このモデルは収束までにより多くの世代数を必要とすることを意味する．

図 4.11 を見ると，3 次多項式エンベロープモデルはかなり元の周波数曲線を再現しており，このことは高々6 個の定数の調整で人間の歌声の周波数曲線が表現できることを意味している．

この実験を踏まえ，探索の終わる早さと十分な自由度という 2 つの観点から 3 次多項式エンベロープモデルが最も本システムに適切であると判断した．

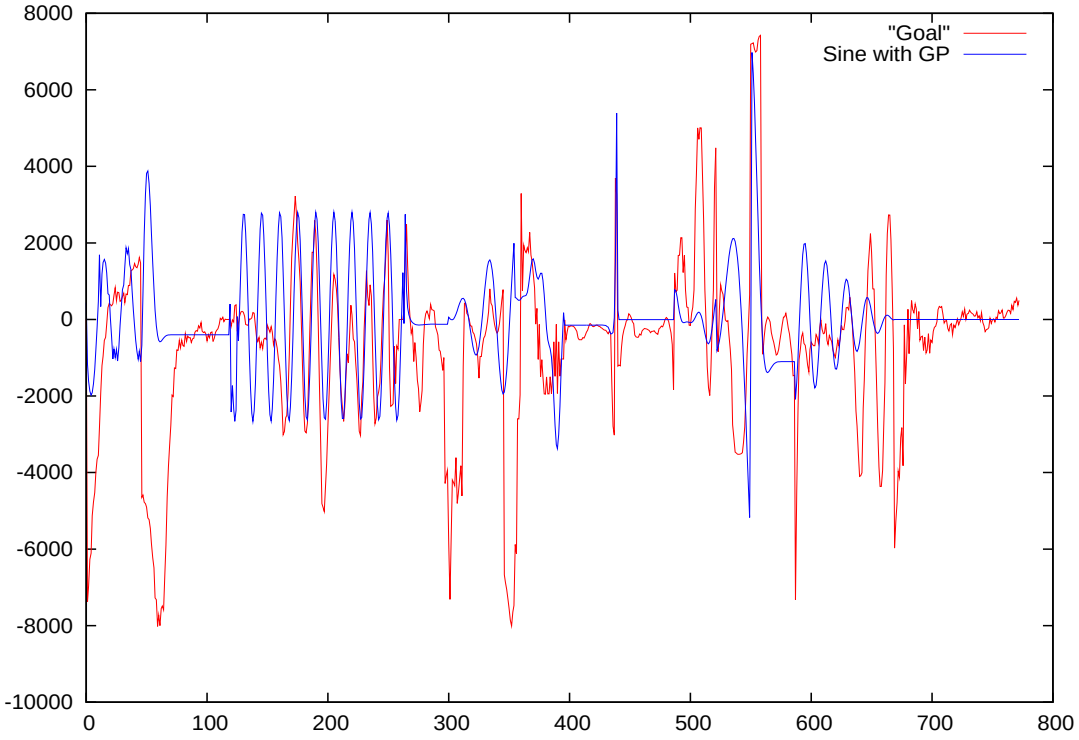


図 4.9: GP エンベロープモデルの最優秀個体

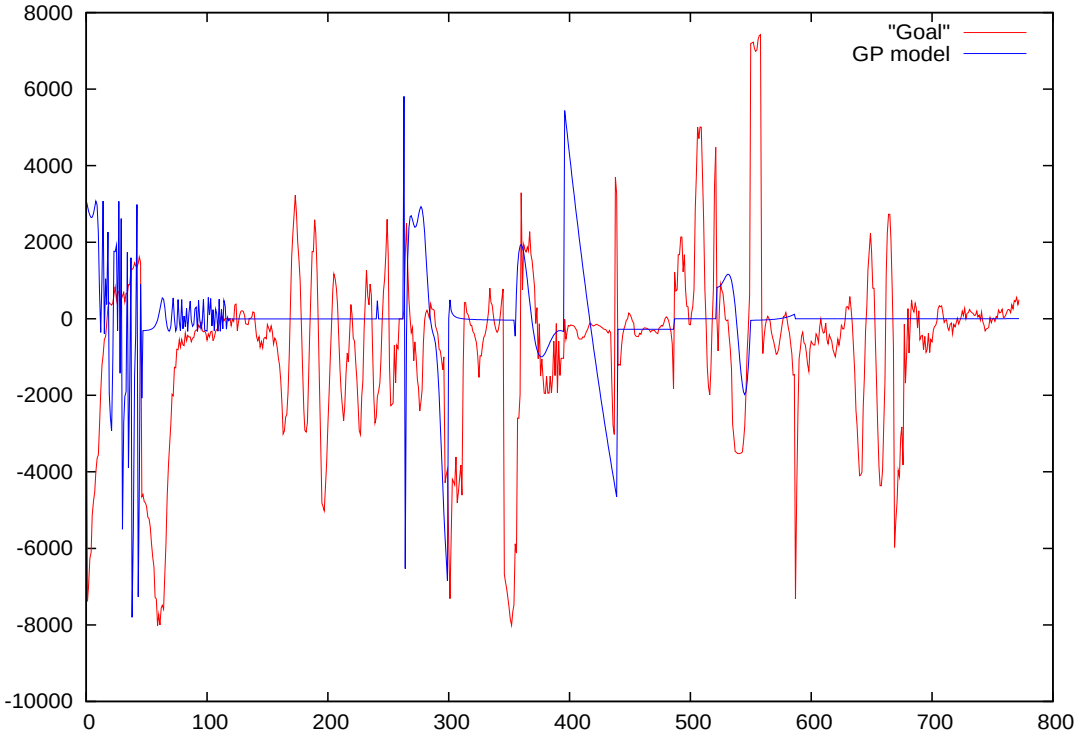


図 4.10: シンプル GP モデルの最優秀個体

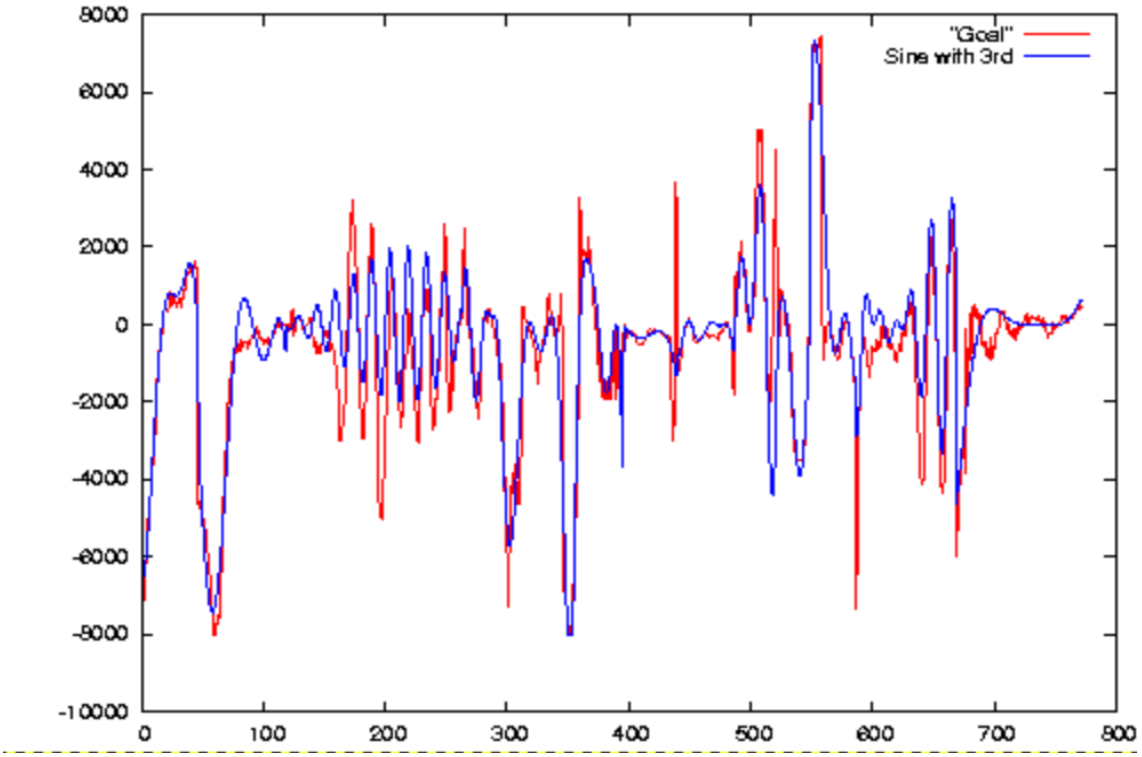


図 4.11: 3 次多項式エンベロープモデルの最優秀個体

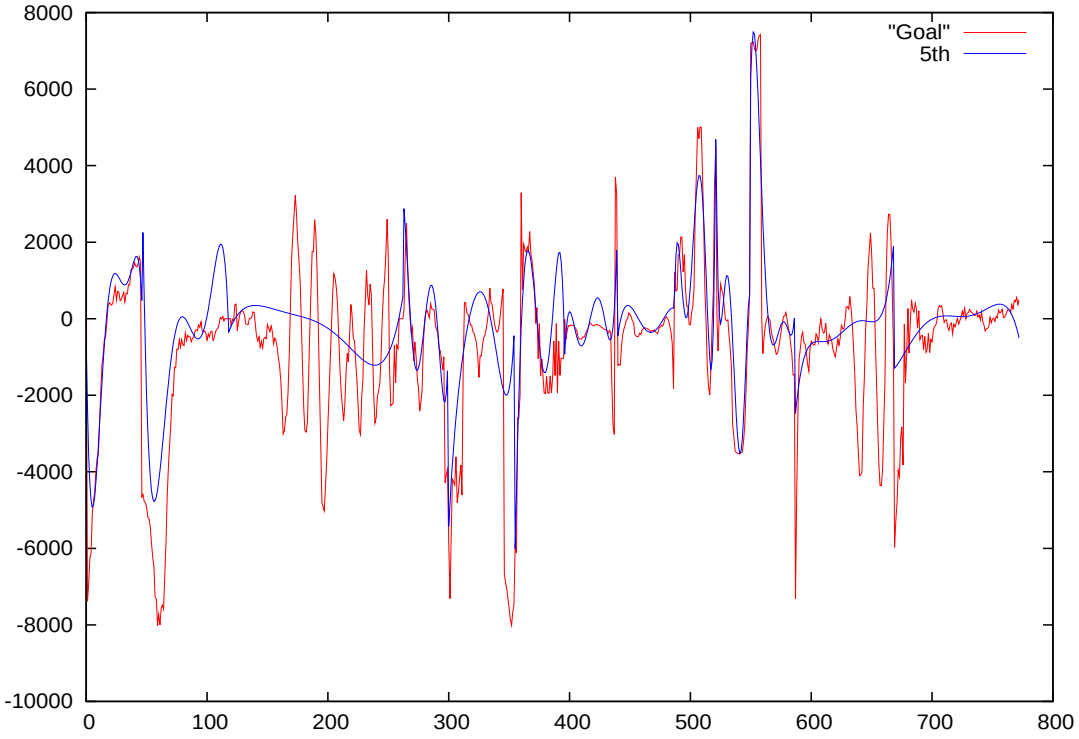


図 4.12: 5 次多項式モデルの最優秀個体

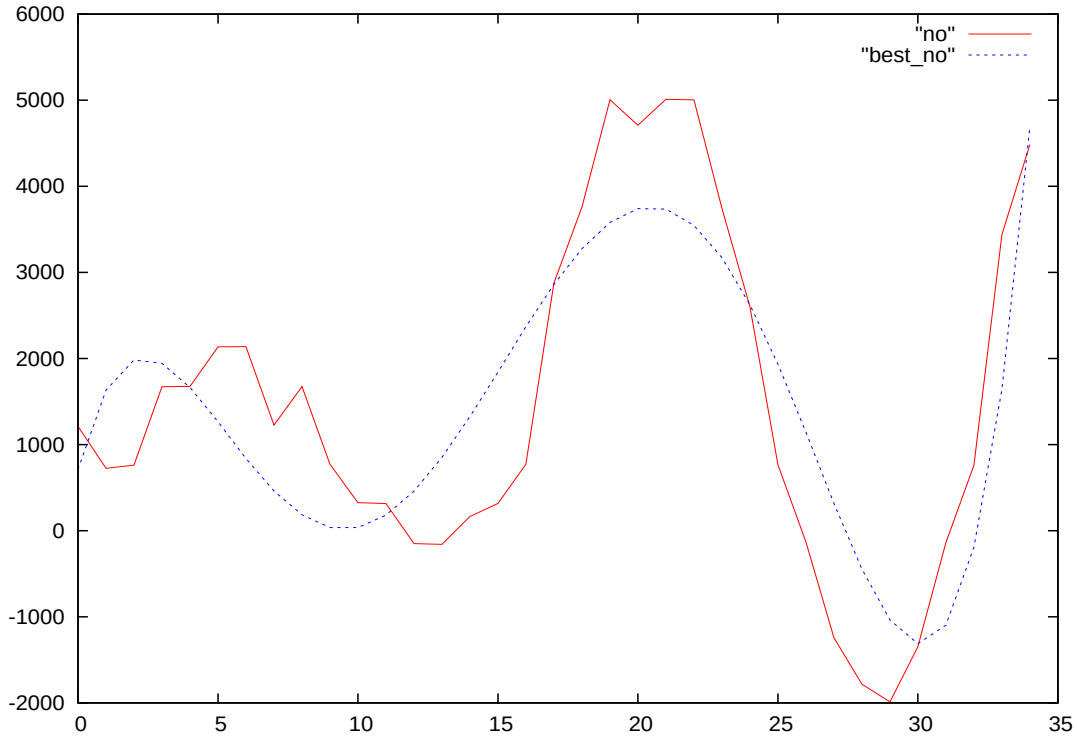


図 4.13: Frequency curve of ”no”

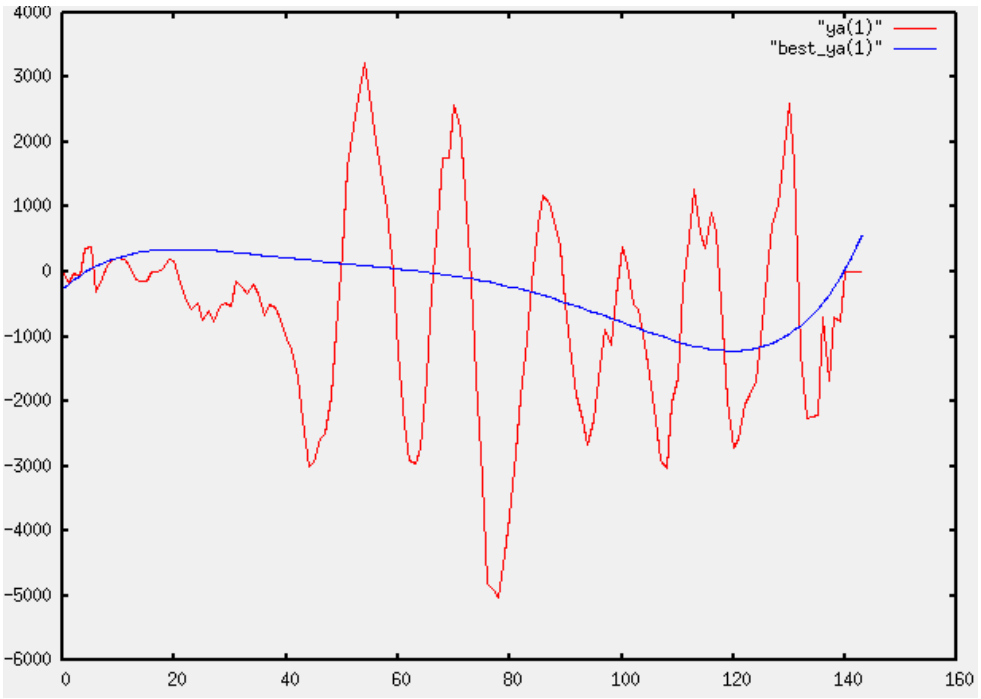


図 4.14: Frequency curve of ”ya(1)”

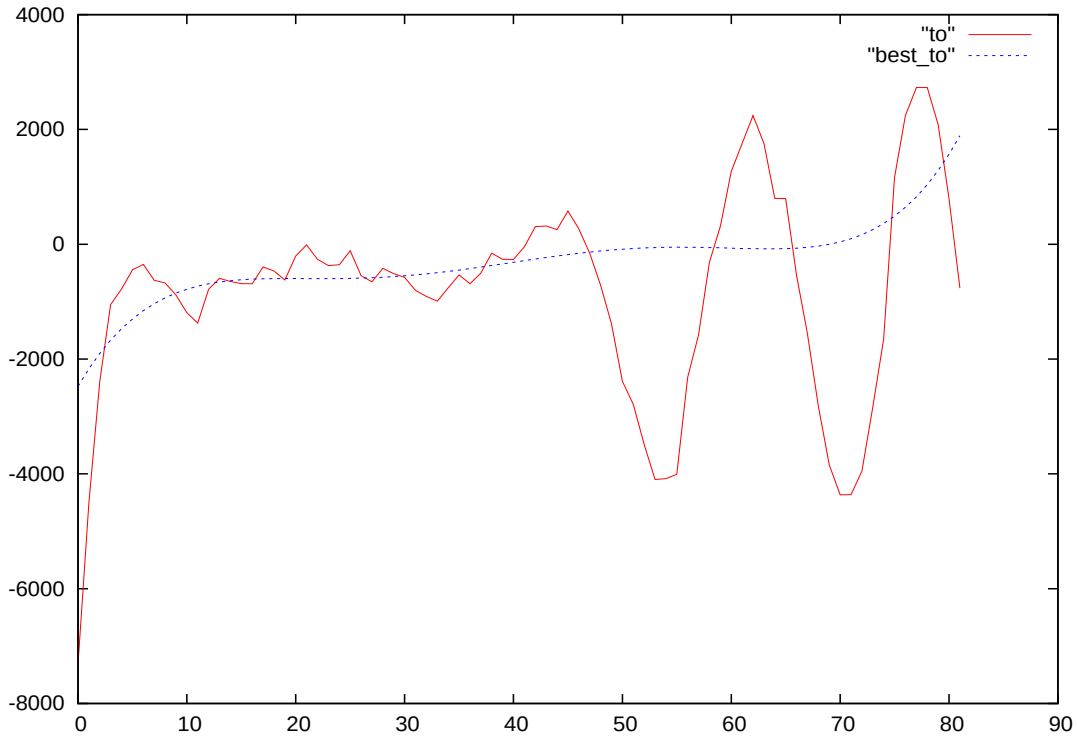


図 4.15: Frequency curve of ”to”

第５章

提案システムの収束性の検証

5.1　局所探索との比較

5.1.1　実験方法

本システムではIECを用いている為，GAによる探索で次世代を生み出している．ところが，ランドスケープが簡単である場合，評価関数の最適化は局所探索でも出来てしまう場合がある．そこで，前章の実験を局所探索を用いて行った場合にどのような結果になるかを調べた．用いたモデルは3次多項式エンベロープモデルである．モデル中の6つの定数を局所探索により最適化する．GAによる探索との比較を公平に行うため，以下の条件で実験を行った．

1. ランダムに8個の個体を作る．これを1世代目とする．
2. 8個中の最優秀個体の近傍に7個の個体を作る．これを2世代目とする．
3. 新たに出来た7個体に前世代の最優秀個体よりも良い物があればその個体を最優秀個体とし，その近傍に次の7個体を作る．
4. 手順3を50世代まで繰り返し，できた最優秀個体と人間の周波数曲線との誤差を測定する．

解の探索回数は前の実験と同様で8個体*50世代=400回である．

5.1.2　結果

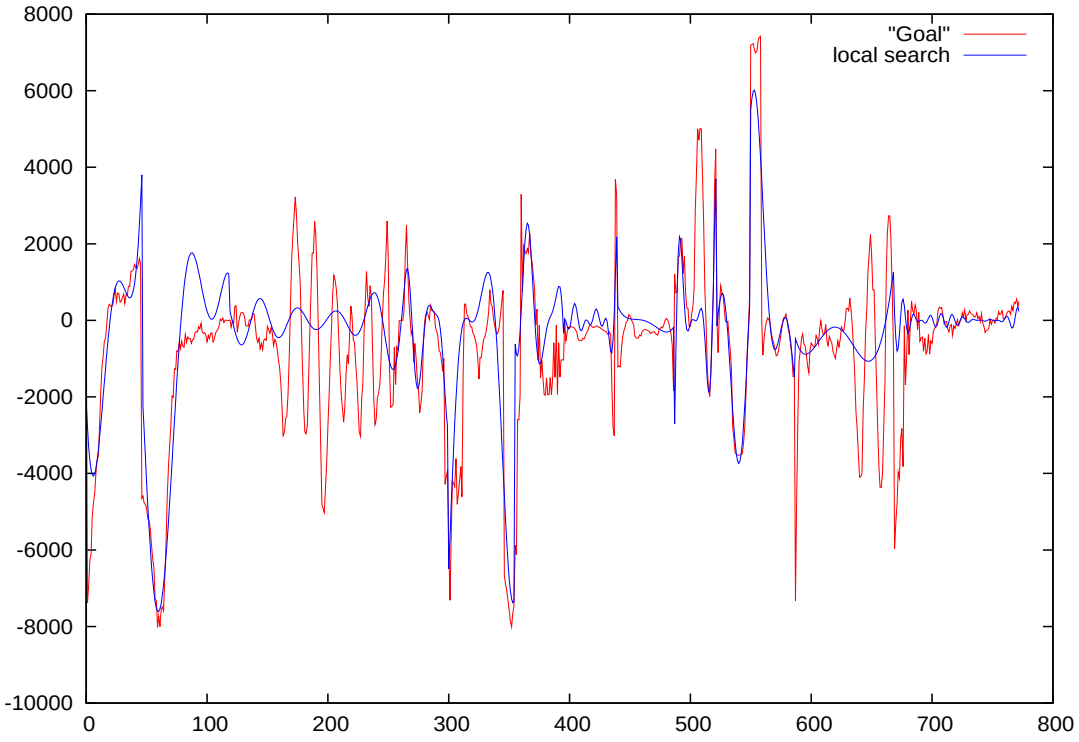


図 5.1: 局所探索での最優秀個体

表 5.1: 各探索手法での平均絶対誤差 (best)

音符	yu	u	ya1	ke1	ko	ya2	ke2	e
GA	479.76	538.52	951.05	602.7	743.33	655.89	404.54	173.60
局所探索	927.70	989.87	1267.00	516.55	1270.75	1314.59	478.87	358.98
音符	no	a	ka	to	n		total	average
GA	1288.90	273.08	720.15	877.09	360.06		8068.65	620.67
局所探索	1377.30	279.86	877.07	1161.98	550.96		11371.48	874.73

表 5.2: 各探索手法での平均絶対誤差 (average)

音符	yu	u	ya1	ke1	ko	ya2	ke2	e
GA	822.44	662.66	1059.47	659.57	1107.25	893.19	464.58	240.45
局所探索	1589.55	1896.71	1615.61	999.73	2239.94	1615.35	822.21	513.56
音符	no	a	ka	to	n		total	average
GA	1413.86	463.79	978.19	955.40	422.94		10143.79	780.29
局所探索	1561.65	1038.94	1797.69	1435.19	810.67		17936.79	1379.75

表 5.3: 各探索手法での平均絶対誤差 (standard deviation)

音符	yu	u	ya1	ke1	ko	ya2	ke2	e
GA	310.01	90.99	87.15	156.41	218.56	175.46	44.72	57.20
局所探索	345.23	557.61	266.82	260.78	438.46	355.46	290.43	142.80
音符	no	a	ka	to	n			
GA	88.99	217.72	316.12	65.40	60.68			
局所探索	145.73	600.62	485.18	255.74	234.17			

図 5.1 が局所探索によって得られた最優秀個体を繋げたものである．表 5.1 と表 5.2 が音符ごとの平均絶対誤差の比較である．表 5.3 が標準偏差である．世代ごとの推移を図 5.2 に示す．

5.1.3 収束性に関する考察

図 5.1 を見ると，最優秀個体に関しては局所探索でもそれなりに近いものが得られていることが分かる．ところが，平均絶対誤差について表 5.1 と表 5.2 を見ると，トータルの誤差に大きな差があり，表 5.3 から標準偏差も大きいことが分かる．このことから，局所探索でも何度かやれば良いものが得られることはあるが，平均的には GA 探索のほうが優れていることが分かった．図 5.2 を見ても，局所最適解に陥って世代が進んでも最適化があまり進んでいないことが分かる．以上の結果から，本研究の探索手法に関して GA による探索が適当であると判断した．この事は，

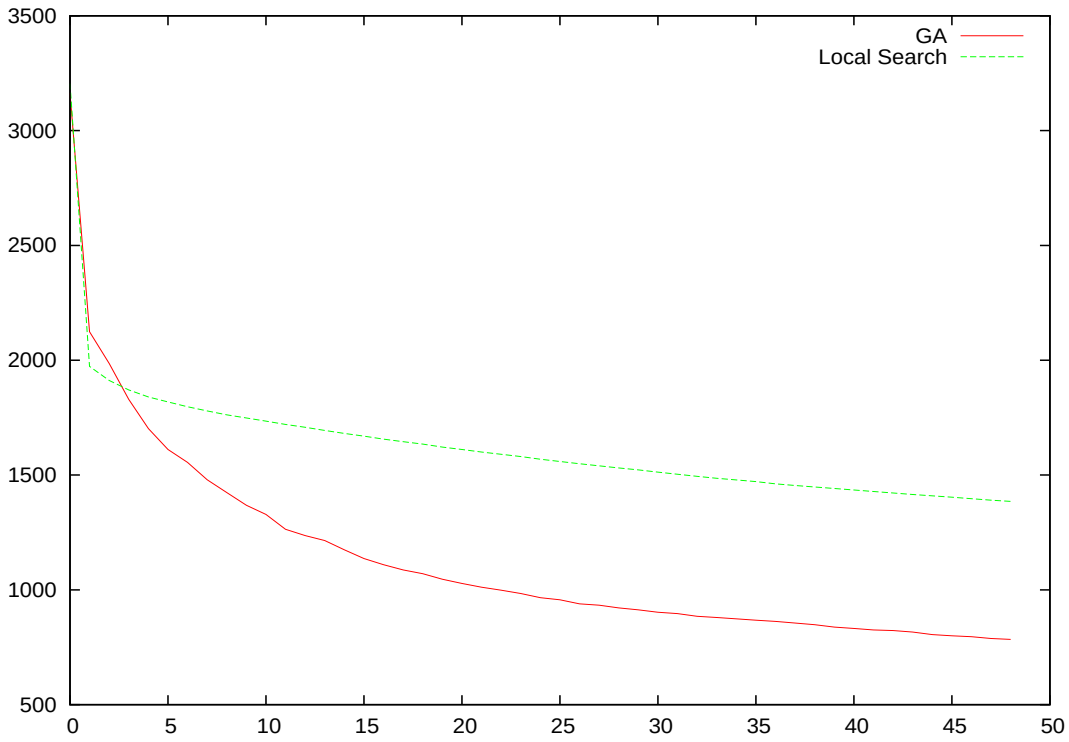


図 5.2: 局所探索と GA 探索の平均絶対誤差の推移

局所探索的な性質を持つ人手による調整でこのタスクを解くことの難しさを示しており，多くのユーザが調整を諦めた原因と考えられる．

5.2　ノイズを考慮した実験

5.2.1　人間が評価する際の問題点

第 2 章で述べたように，人間が評価する際，細かい点差が判断できない場合があり，それが解探索を妨げることがある．このノイズを考慮するために，次の操作を加えて前章と同じ実験を行った．

1. 8 個体のうち，最も良いものと最も悪いものの評価の間を N 分割する．
2. 同一の分割範囲にいる個体は同一の評価であるとする．

この操作により，評価は N 段階となり，人間が考慮出来ないような細かい評価の差は無視されるようになる．今回は 5 段階評価と 2 段階評価の場合において，ノイズの無い場合と比べどの程度結果が悪化するかを見た．

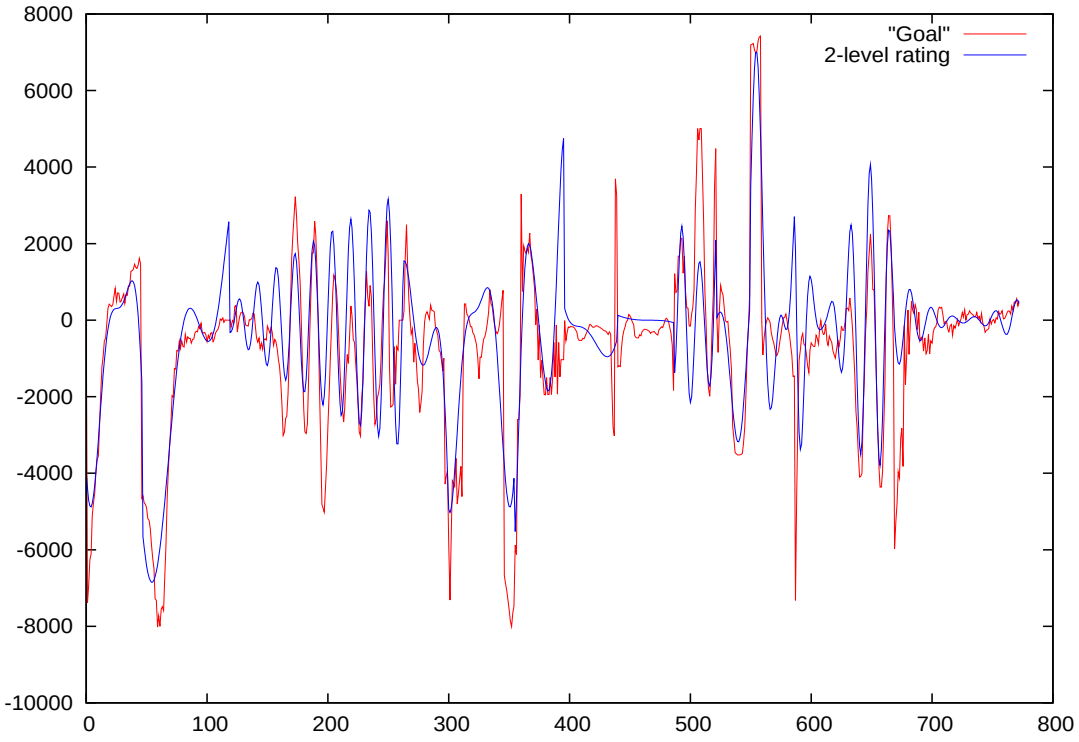


図 5.3: 2 段階評価での最優秀個体

表 5.4: ノイズごとの平均絶対誤差 (best)

音符	yu	u	ya1	ke1	ko	ya2	ke2	e
ノイズ無し (∞段階評価)	479.76	538.52	951.05	602.7	743.33	655.89	404.54	173.60
5 段階評価	619.53	794.09	1009.57	597.63	1065.13	721.77	420.67	314.74
2 段階評価	680.07	784.97	1165.58	699.81	1374.73	1164.66	618.09	364.89
音符	no	a	ka	to	n		total	average
ノイズ無し (∞段階評価)	1288.90	273.08	720.15	877.09	360.06		8068.65	620.67
5 段階評価	1481.91	493.09	839.28	974.07	420.56		9752.07	750.16
2 段階評価	1533.05	533.42	1337.66	1150.27	546.10		11953.28	919.48

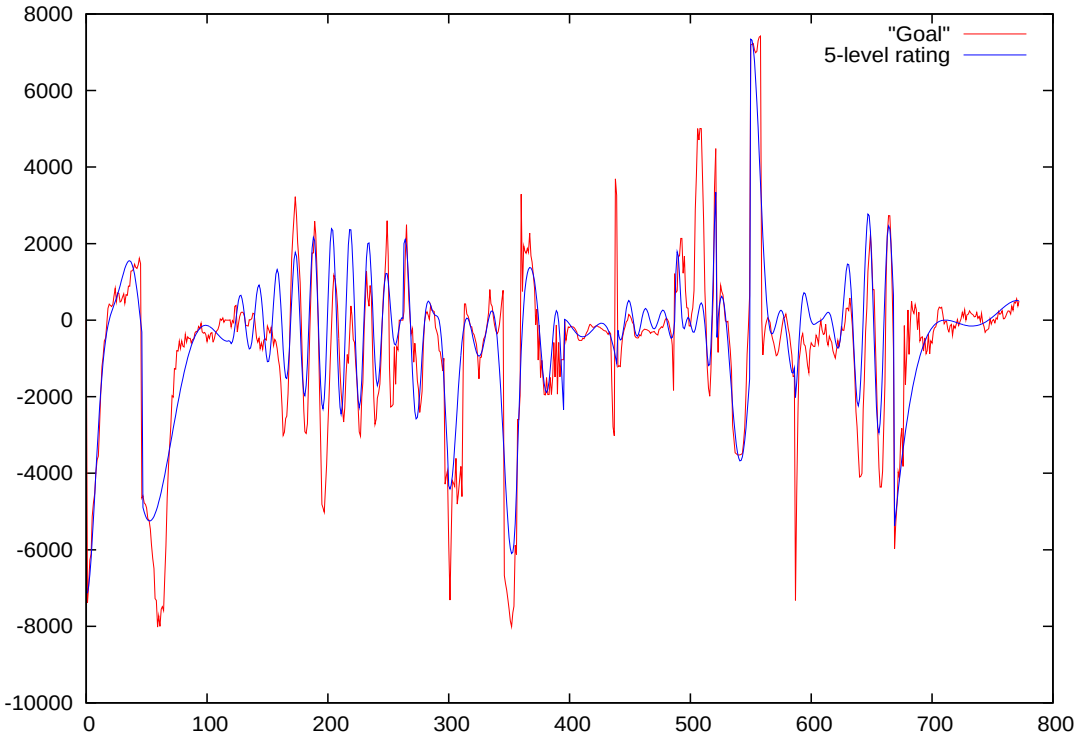


図 5.4: 5 段階評価での最優秀個体

表 5.5: ノイズごとの平均絶対誤差 (average)

音符	yu	u	ya1	ke1	ko	ya2	ke2	e
ノイズ無し (∞段階評価)	822.44	662.66	1059.47	659.57	1107.25	893.19	464.58	240.45
5 段階評価	1023.96	1115.60	1190.27	781.20	1564.97	1101.48	620.25	420.56
2 段階評価	1252.50	1482.35	1488.24	1021.57	1943.91	1401.10	1031.09	632.63
音符	no	a	ka	to	n		total	average
ノイズ無し (∞段階評価)	1413.86	463.79	978.19	955.40	422.94		10143.79	780.29
5 段階評価	1648.13	787.32	1251.44	1150.23	625.43		13280.83	1021.60
2 段階評価	1840.00	1074.12	1886.53	1459.53	864.34		17377.90	1336.76

表 5.6: ノイズごとの平均絶対誤差 (standard deviation)

音符	yu	u	ya1	ke1	ko	ya2	ke2	e
ノイズ無し (∞段階評価)	310.01	90.99	87.15	156.41	218.56	175.46	44.72	57.20
5 段階評価	223.73	414.58	106.40	115.94	360.25	196.95	156.62	67.23
2 段階評価	350.94	558.30	177.13	309.93	358.64	184.88	276.95	211.97
音符	no	a	ka	to	n			
ノイズ無し (∞段階評価)	88.99	217.72	316.12	65.40	60.68			
5 段階評価	108.26	342.70	292.02	138.26	142.86			
2 段階評価	223.31	340.45	418.45	233.23	221.89			

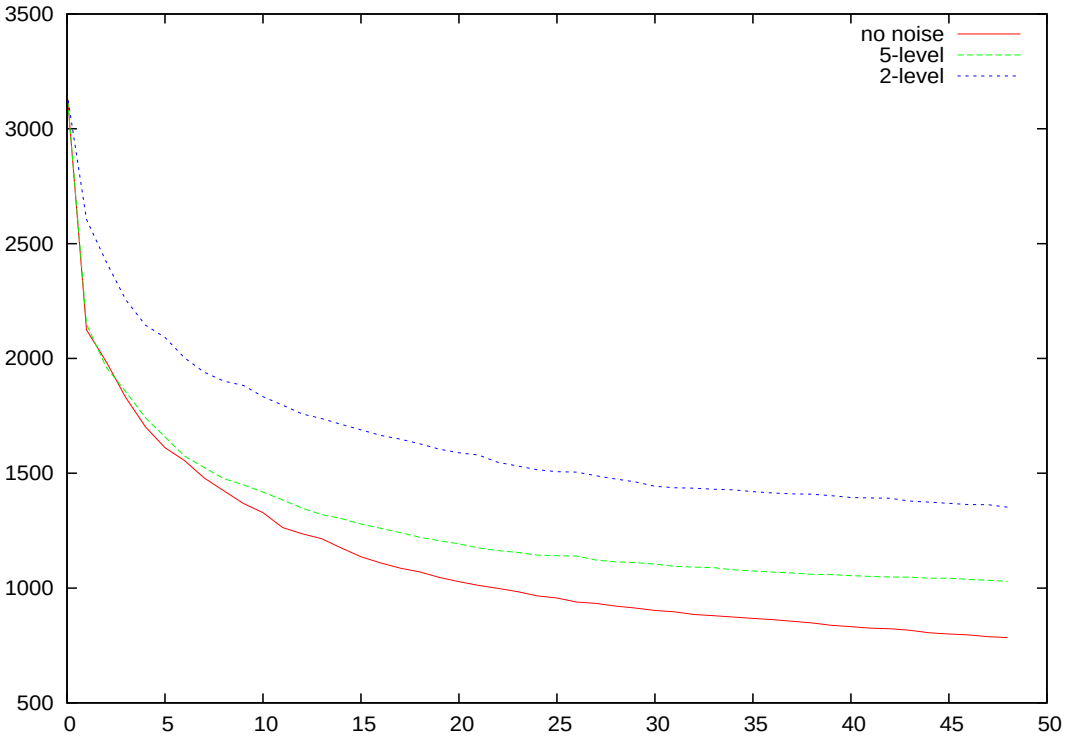


図 5.5: ノイズを含んだ場合の平均絶対誤差の推移

5.2.2 結果

図 5.3 が 2 段階評価, 図 5.4 が 5 段階評価によって得られた最優秀個体を繋げたものである．表 5.4 と表 5.5 が音符ごとの平均絶対誤差の比較である．表 5.6 が標準偏差である．世代ごとの推移を図 5.5 に示す.

5.2.3 ノイズの影響の考察

ノイズが大きくなるにつれて，先の実験の局所探索の場合と同じように平均の値が悪く，分散が大きくなってくることが分かった．人間が評価する場合，せいぜい 5 段階評価程度しか出来ないことを考えると，ノイズを含んだ場合の収束性がより正確なものであるといえる．50 世代でノイズ無しと比較して，平均絶対誤差は 5 段階評価で約 1.3 倍，2 段階評価で約 1.7 倍になっている．もしユーザが 2 段階評価しか正しく出来ない場合，局所探索と同程度の収束スピードに落ちてしまうことが分かった.

第6章

インタフェースの検証

既存インタフェースの説明

図 6.1 が VOCALOID における既存インタフェースである．画面上部がメロディラインと歌詞を入力する部分であり，下部が周波数曲線の調整部分である．ユーザはマウスで曲線を描き，曲を聴きながら調整を進める．

図 6.1: 既存インタフェース

どのような曲線を描けばよいかというガイドは無く，それゆえ未調整よりも好ましくない歌い方になってしまうことも多い．

6.1　アンケートによるインタフェースの評価

今回提案する IEC を用いたインタフェースを既存インタフェースと比較し，どの程度の負担軽減が出来ているかを調べるために，アンケートによる評価を行った．このアンケートは歌声に関する知識を持たない 11 人を対象にした．インタフェースを使う順番はランダムである．今回は以下の 4 つの質問に，それぞれ 5 点満点で採点をしてもらった．

Q1　既存インタフェースによって完成した曲は満足なものであったか．

Q2　既存インタフェースによる調整は簡単であったか．

Q3　IEC インタフェースによって完成した曲は満足なものであったか．

表 6.1: アンケート結果

	Q1	Q2	Q3	Q4
平均	2.36	2.55	3.36	3.64
標準偏差	1.12	1.37	0.92	1.21

表 6.2: インタフェース間の優劣

	IEC	Previous	draw
満足さ (Q3vsQ1)	7	3	1
簡単さ (Q4vsQ2)	7	2	2

Q4　IEC インタフェースによる調整は簡単であったか.

このアンケートによって得られた結果を表 6.1 に示す．また表 6.2 にはそれぞれのインタフェースにより高い点数をつけた人数を示している．満足さの行は Q1(既存) と Q3(IEC) を比べた場合であり，簡単さの行は Q2(既存) と Q4(IEC) を比べた場合である．

6.2　IEC インタフェースの優位性

表 6.1 を見ると，Q3 の平均点は Q1 の平均点より 1.00 点，Q4 の平均点は Q2 の平均点より 1.09 点高いことが分かる．そして表 6.2 を見ると，満足さの点で IEC インタフェースがより優れていると考えた人は 11 人中 7 人，簡単さの点でも 11 人中 7 人という結果になり，既存インタフェースが優れていると答えた人数と有意な差が見られることから，本研究の目的である IEC インタフェースによる周波数曲線の調整の簡単化が果たせたと考えられる．

そして既存インタフェースを先に使用した多くの被験者は既存インタフェースによる調整を行った際，数回の試行で調整を諦めてしまい，IEC インタフェースを先に使用した被験者から,「IEC インタフェースによりどのような曲線が良いのかが分かったので既存インタフェースでも調整がある程度可能になった.」というコメントを得られたことからも，発想支援という IEC の利点を生かせたと考えられる．

第7章

考察

7.1　評価実験からのシステム全体の考察

本章ではモデル比較実験，収束性評価実験，インタフェースについてのアンケートで得られた結果を基に考察を行う．

まずモデルについて，斉藤らのモデルは人間の歌声に関する知見から求められたものであったが，同じ曲線で全ての音符の表現を再現するのは不可能であるし，実際の歌声のパターンは非常に多様であり，このモデル中の定数を調整しても表現できないものも存在することが分かった．そこでモデルの自由度を高めることで多様な歌声表現を可能にしたが，GP を用いた場合あまりにも複雑になりすぎて，IEC での少ない世代数での収束は見られなかった．本研究では GP を多項式に置き換えることで探索空間を制限し，IEC システムが許容できる世代数での収束を確認できた．また，歌声にはビブラート表現のようにサイン関数の特徴を持つパターンがしばしば現れることからも推測できるように，モデル中にサイン関数を含めることでより正確な人間の歌声の再現が可能であることが分かった．

次に探索手法について，IEC システムで解けるタスクは，通常の GA で解くような最適化問題よりも遥かに少ない世代数，個体数で解くため，単純な局所探索でも，ともすれば GA よりも速く解ける可能性がある．そのため同じタスクで GA と局所探索について比較を行った．局所探索でも数回で人間の歌声に近いものが得られることもあるが，平均的には GA と大きく差があり，今回のシステムで用いるのは不適当と判断した．また評価の際のノイズを考慮した実験では，本システムにおいてもノイズが無視できず，2 段階評価で収束させるには多くの世代を必要とすることが分かった．

既存インタフェースとの比較においては，人手による調整の難しさを再確認すると共に，本システムの有効性を見ることが出来た．発想支援の点からも，探索に GA を用いることの有効性を確認できた．

7.2　今後の展望

今後の研究課題・展望を以下にまとめる．今回モデルの比較実験で人間の周波数曲線を再現し，平均絶対誤差の数値が低いものが得られたが，この誤差が人間の聴覚にどの程度感知できるものであるのかを調べる必要がある．また同様に，ノイズを考慮した実験で，本システムへのノイズの影響はあることが分かったが，この収束スピードの変化が実際のユーザ使用時にどの程度の差を生むのかについても考慮しなければならない．また，今後「赤とんぼ」以外の曲，歌手でも同様にこのモデルが通用するのかも測定する予定である．

探索手法については今回 GA と局所探索を比べ，GA が優れているという判断をしたが，今後他の探索手法についても比較し，最良のものを検討する．ユーザのノイズを考慮し，2 つの個体の比較だけをユーザに行わせるという対話型差分進化に関して実装し，今回のタスクに有効であ

るかを調べたい．

アンケートに関しては 11 人という少ない人数であるため，信憑性を高めるためにサンプルサイズの拡大はもちろん，今回行ってもらった際の世代数に関しても 5 世代未満という少ないものであるため，曲線調整が完成する段階までの使用感を測定する必要がある．インタフェース上に曲線を視覚的に表示したことの効果は IEC システム上で経験を積んで初めて出てくるものであると考えられるので，ある程度の世代数を経る実験において視覚情報のみで評価を行った割合を測定することを考えている．

今回歌声の表情付けとして周波数曲線の調整を扱ったが，歌声の強弱に関しても同様な最適化を行うことで，より豊かな表現が可能になる．よって今後，強弱に関してもモデルの検討を行い，調整可能にしたい．

第8章

結論

本研究では VOCALOID ユーザの負担を減らすために，IEC を用いた周波数曲線自動最適化システムを実装した．そして以下の点において本システムが優れていることを確認した．

周波数モデル

過去用いていたモデルは自由度の少なさから人間の歌声のパターンを表現しきれていなかった．そのため自由度を増やしたいくつかのモデルを提案し，この中からシステムで用いる周波数モデルを選択するため, 決められた世代数でどの程度人間の歌声の周波数曲線を再現できるかを測定し，3 次多項式エンベロープモデルが優れていることを明らかにした．

探索手法

ユーザへの個体の提示の方法として，GA によって探索したサンプルを与える手法の他に，単純に選んだものの近傍を与える局所探索手法が考えられるが，収束スピードを調べることで，GA 探索がより本システムに適切であり，発想支援という点においても優れていることを確認した．

インタフェース

インタフェースの評価をアンケートによって行い，既存の人手による調整よりも本システムが使いやすく，また満足度の高い調整を可能にしていることが分かった．コメント等から，発想支援の役割も果たしていることを確認した．

以上の結果から本システムは歌声の表情付けにおけるユーザ負担を軽減出来ていると言える．さらに IEC の人間による評価におけるノイズを考慮した実験で，本システムへのノイズの影響を見た．この影響を小さくするような GA の改良が今後の課題である．

謝辞

本研究を進めるにあたっては，多くの方々からご助言，ご指導をいただき，様々な形でのご支援をいただきました．

指導教員である伊庭斉志教授には，研究の方針に関するアドバイスや，研究の問題点の指摘など，様々な形で助けて頂きました．また研究環境に関しても，私が不自由無く実験等を行えるよう配慮して下さりました．時には先の見えないこともあった私の研究を進めることが出来たのは伊庭先生の指導のおかげであり，非常に感謝しています．

また，研究室の先輩方には，研究環境を整備して頂くとともに，研究に対する態度や考え方など多くの面で参考にさせて頂きました．柳瀬利彦氏には私の知識不足をフォローして頂き，また研究の主に基礎的な部分についてたくさんのアドバイスを頂きました．Claus Aranha 氏には論文や発表等での英語のチェックや，明るい研究室の雰囲気作りをして下さいました．丹治信氏には，サーバー管理など研究室の環境整備をして頂くと共に，同じグループの先輩として音楽の知識を全く持っていない私にも丁寧に指導して頂きました．

同期である石渡裕之氏，熊谷基樹氏，藤原健太氏，渡辺幹生氏には，良き話相手になって頂くことで，研究を進める上での励みになり，時には議論をする事が私の研究の助けとなりました．

後輩である Hettiarachchi Dhammika Suresh 氏，Vatanutanon Jiradej 氏，Badarch Tserenchimed 氏，今川哲矢氏には，楽しい研究室の雰囲気を作って頂き，また英語の添削の際にもお世話になりました．

最後に経済的な援助と共に，常に私の健康について心配して下さった両親に感謝いたします．

参考文献

[1] H Kenmochi, H Ohshita. *Singing synthesis system 'VOCALOID'*. 2007-MUS-72-(5), pp. 25-28.

[2] David Cope. *band in a box*. http://www.cameo.co.jp/PG/win/

[3] J Sundberg. Research on the singing voice in retrospect. TMH-QPSR, vol.45-1, 2003, pp. 11-22.

[4] T Nakano, M Goto. VocaListener: An Automatic Parameter Estimation System for Singing Synthesis by Mimicking User's Singing. 2008-MUS-75, No.50, pp. 49-56.

[5] H Takagi, T Unemi, T Terano. Perspective on Interactive Evolutionary Computing. Proceedings of the Annual Conference of JSAI, September,1998.

[6] M Sugawara, M Mitsunori, T Hiroyasu. Interactive genetic algorithm using generates initial individual based on a favorite color image. JSAI 2008, 2B1-2.

[7] J R Koza, R Poll. GENETIC PROGRAMMING.

[8] I Ono, M Yamamura, H Kita. Real-Coded Genetic Algorithms and Their Applications. Proceedings of the Annual Conference of JSAI, 2000.

[9] D Ando, P Dahlsted, M Nordahl, H Iba. CACIE: Computer Aided Composition System by Interactive Evolutionary Computation. 2005-MUS-59-(10), pp. 55-60.

[10] T Saitou, M Unoki, M Akagi. A study on a control method of vibrato modulation frequency for synthesizing natural singing-voice. IEICE technical report, 2005.

[11] T Saitou, M Goto, M Unoki, M Akagi. SingBySpeaking: Singing Voice Conversion System from Speaking Voice By Controlling Acoustic Features Affecting Singing Voice Perception. IPSJ SIG Notes 2008(12), pp.25-32, 2008.

[12] M Tanji, D Ando, H Iba. Improving Metrical Grammar with Grammar Expansion. AI 2008: Advances in Artificial Intelligence, pp. 180-191, 2008.

[13] 高木 英行，Denis Pallez. 対比較ベース対話型差分進化. 進化計算シンポジウム 2009.

[14] S F Wang, X F Wang, J Xue. An Improved Interactive Genetic Algorithm Incorporating Relevant Feedback. Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21 August 2005, pp. 2996-3001.

[15] M Unehara, T Onisawa. Music Composition System based on Subjective Evaluation. 0-7803-7952-7/03/S17. 00 0 2003 IEEE, pp. 980-986.

[16] S B Cho. Emotional Image and Musical Information Retrieval With Interactive Genetic Algorithm. Proceedings of the IEEE, Vol. 92, No. 4, April 2004, pp. 702-711.

発表文献

[1] Watanabe, A., Iba, H. An IEC System with Frequency Model for Creating Singing Vocal Expression. In *Genetic and Evolutionary Computation Conference*, 2010.(投稿中)

[2] Watanabe, A., Tanji, M., Iba, H. Creating Singing Vocal Expressions by means of Interactive Evolutionary Computation. In *The 5th International Workshop on Computational Intelligence and Applications*, 2009.

[3] 渡辺 晃生, 伊庭 斉志. IEC を用いたコンピュータ歌唱の表情付け. 第 2 回進化計算フロンティア研究会, 2009.

[4] 渡辺 晃生, 安藤 大地, 稲田 雅彦, 丹治 信, 伊庭 斉志. IEC を用いた歌声パラメータの最適化システム. 情報処理学会音楽情報科学研究会 2009-MUS-79.

[5] 渡辺 晃生, 安藤 大地, 丹治 信, 稲田 雅彦, 伊庭 斉志. 対話型進化論的計算を用いた歌声合成パラメータの探索. IPSJ SIG Technical Reports, 2008.