

IEC を用いた歌声パラメータの最適化システム

渡辺晃生、安藤大地、丹治信、稲田雅彦、伊庭斉志

akio@iba.k.u-tokyo.ac.jp

東京大学工学部工学系研究科電気系工学専攻 伊庭研究室

昨今、コンピュータに歌を歌わせることのできる VOCALOID というアプリケーションが注目されている。このアプリケーションにおいてはメロディラインや歌詞だけでなく、歌声のパラメータ調整を行う事によって様々な表現が可能であるが、歌声パラメータの調整は知識のないユーザには非常に大きな負担となっていた。今回は対話型進化手法 (IEC) と呼ばれる GA の一手法を用いて、ユーザはコンピュータによって提示される歌声パラメータから好みの物を選ぶという比較的簡単な操作によってパラメータの最適化を行うシステムを実装した。またシステムの評価として手動で調整した歌声パラメータにどの程度のユーザの負担で近づけるかを調べるための実験を行った。

The system for optimizing parameter of singing voice using IEC

Akio Watanabe, Daichi Ando, Makoto Tanji, Masahiko Inada, Hitoshi Iba

Department of Electrical Engineering and Information Systems, Graduate School of Engineering, The university of Tokyo

VOCALOID is an application that realize singing by computer. VOCALOID enables users to create songs sung by computer only with inputting of melody and lyrics. However, for higher quality of song, complex optimization of voice quality parameters is required. For optimizing problems, genetic algorithm is commonly used. In this paper, we introduce a prototype of an application for optimizing parameters of singing voice easily by using Interactive Evolutionary Computation(IEC), which adopt perceptual evaluation of human for evaluation functions. Besides, to examine the number of times of evaluation by human, we made an experiment to recreate the parameter of some music.

1 はじめに

近年、コンピュータによる歌唱を実現した VOCALOID*2 というアプリケーションが注目を集めている。このアプリケーションにおいてユーザ

は、歌詞とメロディラインを入力することで、コンピュータによる歌唱を実現することができる [1]。

このアプリケーションにおいて、より人間らしく、人間にとって快適な歌声を目指す場合、ユーザは声の抑揚や細かい高低など、声質に関わる複数パラメータを入力することになる。この時、手作業で各パラメータを調整し、何度も音楽を聞きながら最

*2 <http://www.crypton.co.jp/mp/pages/prod/vocaloid/>

適化していくことはユーザにとって大きな負担であった。

上記のような声質に関わるパラメータ（以下歌声パラメータ）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。歌声パラメータの探索空間の複雑性から、遺伝的アルゴリズム (Genetic Algorithm : GA) によるパラメータ最適化はその一つの解法として適切であると考えられる。しかし、今回のタスクにおいては、目的関数の設計自体も曖昧性を有している。その為、GA の手法をそのまま用いることは必ずしも適当とはいえない。そこで今回は対話型進化計算 (Interactive Evolutionary Computation : IEC) という GA の一手法について検討する。IEC は人間の主観的評価を GA の評価部分に利用することで、人間の感性をシステムに組み込んだ進化計算を可能とする [2]。これにより、IEC では人間の持つ印象や好みなどの数式化の困難な目的関数に対して、解探索が可能となる [3]。本研究では、IEC によって歌声パラメータを自動調整するシステムについて、そのプロトタイプを実装した。

一方、IEC によるシステムは人間による評価フェーズが存在するため、ユーザの負担を考慮する必要がある。この時、進化計算における解探索の効率は、ユーザの負担と直結する。すなわち、できるだけ少ない試行回数で最適解を求める必要がある。以下ではまず、本システムに用いた VOCALOID について解説したのち、IEC についてその枠組みを述べる。その後、実装したプロトタイプを紹介する。

そして、ユーザの負担を評価する為、本システムで用いた遺伝的プログラミング (Genetic Programming: GP) に基づく解探索について、その評価結果を示す。

合成音声による歌唱アプリケーション

VOCALOID とは、コンピュータによる歌唱を実現したアプリケーションであり、メロディラインと

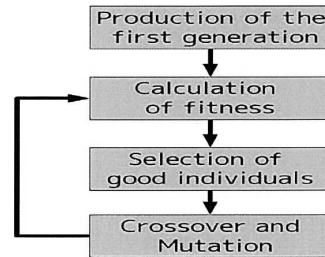


Fig. 1 Image of GA

歌詞を入力するとコンピュータが歌ってくれるというものである。今までに作曲支援のアプリケーションはいくつかあったが [4][5]、歌詞を入力することが出来る点で表現の幅が広がり、今日では多くのユーザが様々な楽曲を作成し、ウェブ上で評価しあうという現象が起きており、非常に注目されている。

対話型進化計算

1.1 概要

数値化の困難なタスクに対して、近年、対話型進化計算 (Interactive Evolutionary Computation : IEC) とよばれる GA が検討されている。Interactive evolutionary computing (IEC) [2] とは遺伝的アルゴリズムの枠組みにおいて、その評価部分と個体生成をユーザとコンピュータの相互のやりとりとみなし、進化計算を行う。これにより、評価部分に人間の主観的評価を導入でき、人間の持つ印象や好みを探索に反映することができる。

まず通常の GA の流れを説明する。Fig. 1 が GA のフローチャートである。

まず初期個体をいくつか生成し、与えられた評価関数を用いて各個体の評価を行う。その後、良個体を親として交叉や突然変異を行い、次の世代の個体をいくつか生成してさらに評価を行う。以上を繰り返すことで次第に個体を良くしていくという手法である。

通常の GA においてはある評価関数が存在し、各個体の評価が数値として与えられるため、コン

ユーザは他の情報無しに良個体を選択することが出来る。

本研究で用いた IEC について説明する。IEC とは、コンピュータが生み出した作品をユーザが評価し、その評価をもとにしてより (少なくともそのユーザから見て) 良い個体を生みだし、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の高い作品を完成させるというシステムである。わかりやすく言えば、通常の GA における評価関数を人間の感覚に置き換えた計算方法とすることができる。

1.2 IEC の特徴

IEC の特徴として前項でも述べたとおりまず第一に、このシステムは評価関数を必要としないことがあげられる。そのため、システムを作る人間は、たとえば作曲をする場合において、どんな曲が良く、どんな曲は良くないのかということについての知識を必要としない。そのため比較的簡単にユーザの満足度の高い作品を生み出すシステムを作ることが可能である。

次に、デメリットとして、ユーザの疲労度があげられる。このシステムにおいてユーザは多数回の評価を強いられ、またある程度の時間を要する。完全にコンピュータで全て解決できる問題と違い、ユーザに負担がかかってしまうため、この負担をなるべく軽減できるようなシステムが望ましい。

1.3 IEC の具体的な流れ

以下、IEC による解探索について述べる。まずコンピュータに、ランダムなパラメータをもったいくつかの個体を生成させる。各個体のもつパラメータはとくに進化計算において、遺伝子と呼ばれる。この時点では遺伝子はユーザにとって好ましいものではない。

その個体をみてユーザは評価を与える。評価の方法はインタフェースによって様々であるが、ユーザの情報を詳しく抽出しようとするほど、ユー

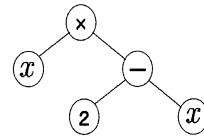


Fig. 2 Image of GPtree

ザ疲労度は高くなる。

この評価が反映され、それを元にして子供の個体が生成される。これを交叉と呼ぶ。また、探索空間を広げるために、親とは全く異なったパラメータを低確率で子供の個体に持たせることがある。これを突然変異と呼ぶ。

交叉、突然変異の後に生まれた子供の個体に対し、さらにユーザが評価を与える。このプロセスを繰り返して、最終的にはユーザの評価を高くするためにパラメータの最適化された個体を生成する。

遺伝的プログラミング

本研究で提案するシステムにおいて歌声パラメータは時間に対する関数としてあらわす。すなわち、進化計算によって生成される個体は関数として表現されることががのぞましい。遺伝的プログラミング (Genetic Programming : GP)[6] は、木構造を用いた GA の事であり、上記の要求をうまく満たすことができる。

Fig. 2 が GP の木の例である。この例では演算子 \times の下に x と $(2-x)$ がぶら下がっているため、全体として $x*(2-x)$ という式を表している。GP の交叉は、木の一部分を他の木のものと取り換えることによって行われる。GP の突然変異は、木の一部分をランダムに入れ換えることによって行われる。

2 提案するシステム

先のセクションで説明した IEC に基づき、歌声パラメータ自動最適化システムについて、そのプロトタイプを実装した。

2.1 システムの目的

本システムを、一種の作曲システムとみなした時、その目的は大きく二つある。

一つは、人間らしい歌声を作り上げることである。この試みはこれまででもなされており、ここで先行研究を紹介する。

中野らは、人間の歌唱をコピーしてビブラートなどを再現する VocaListener を開発した [7]。この研究では、合成された歌唱の音高・音量を目標（ユーザ歌唱）と比較しながら、歌声パラメータを反復更新していく。このシステムで調整された作品は非常に人間らしいと、ウェブ上で高い評価を得ている。

もう一つの目的はコンピュータが歌唱するという特性をいかした、コンピュータにしか出せない歌声を探求することにある。人間らしい歌い方に関してはデータや研究が存在するが、コンピュータにしか出せない方法でなおかつ人間が快適に感じる声質に関してはあまり研究がなされていないため、この目的を達成することが IEC を用いた本システムのオリジナリティである。

2.2 仕様

本システムの仕様を説明する。

はじめに GP によってランダムな 8 個体を作成し、作成された関数の変数 x を 0.1 刻みで変化させ、配列にする。その配列を用いて、VOCALOID におけるメロディラインと歌詞の入ったファイルを編集し、ピッチバンドパラメータの時系列のデータにする。この際、得られた曲線データは各音符の持つ歌声パラメータに反映され、音符が変わるたび、再帰的に反映される。次にこの編集したファイルを Wave ファイルに書き出せば、8 個の音楽ファイルができる。それをユーザに聞かせ、評価する。ユーザはこの際画面に、曲線データも視覚的に見ることが出来る。

Fig. 3 がこのシステムのスクリーンショットである。ユーザは 8 個の個体を曲線データの右のバーで

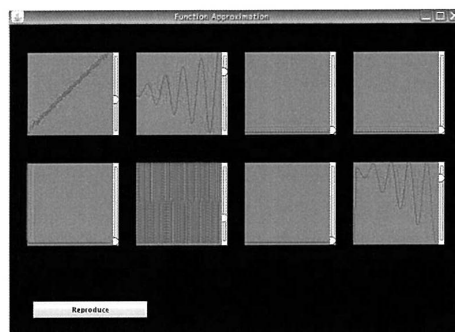


Fig. 3 Image of this application

評価し、その情報が次の世代を生み出す際に利用される。

3 ユーザ疲労度に関する検討

3.1 システムの評価法について

本システムは手動でのパラメータ調整に比べてユーザの疲労度を減らすことが一つの目的であるので、必要な疲労度に関する評価が必要である。ユーザの疲労度は進化計算における解探索の効率と直結する。そこで今回は、正解（目的関数）が既知の状態では GP によるパラメータ生成を行い、一世代の個体数、世代数と目的関数の関連を調べた。

評価の為に、手動で調整した VSQ ファイルを用意した。本システムにおいて GP は四則演算、正弦関数、対数関数を用いて表現される全ての関数を与える。この GP を適用し、VSQ ファイルの歌声パラメータとの二乗誤差の平均を評価関数として GA による計算を行い、どの程度の世代数を用いれば十分に元のパラメータに近づけることが出来るかを調べる。IEC の場合は個体数と世代数の積がユーザの評価回数になるため、この実験により本システムにおけるユーザの疲労度を推定することができる。今回は個体数 8, 16, 32 で 200 世代の計算を行い、最良個体の曲線データ、世代ごとの二乗誤差平均の推移のグラフを得た。Table 1 が今回用いた GP のパラメータである。

Table 1 GP Parameters

Population size	8, 16, 32
Selection	tournament
Maximum depth	5
Minimum depth	3
Crossover rate	0.8
Mutation rate	0.2
Elite size	1

Table 2 result

個体数	8 個体	16 個体	32 個体
	1209	1085	1029
	1044	1625	802
	1038	1021	1217
	1533	1430	1635
	880	1118	669
平均	1199.6	1255.8	1070.4

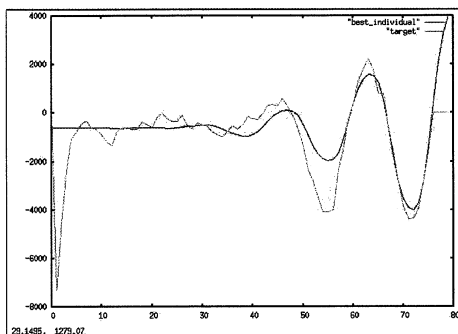


Fig. 4 Best Individual

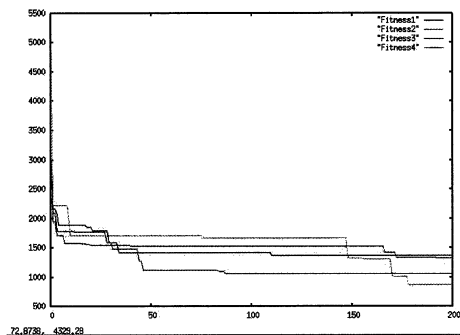


Fig. 5 Fitness

3.2 結果・考察

Table 2 は、得られた最良個体の二乗誤差平均である(常に0の時二乗誤差平均は3130である)。それぞれ5回実験を行い、その平均値を出した。個体数8と個体数16ではあまり平均が変わらず、今回は個体数8の方が良い結果となった。個体数32で

は平均が比較的良くなった。また、各個体数においてベストの個体を見てみると、個体数32での669が最良であるが、個体数8でも880と、かなり元のパラメータデータに近いものが得られている。以上のことから、システムの提示する個体数は8でよく、より平均的に良いものを作るために初期世代に工夫が必要であることが考えられる。Fig. 4は個体数32における最良個体の曲線データを示している。緑色の曲線が手本としたデータ、赤色の曲線が最良個体のデータである。細かい波が取れていない部分があるが、大まかには形をとることができている。また、Fig. 5は個体数8における世代ごとの最良個体の二乗誤差平均の推移である。4回分の実験データを示す。本システムにおいては50世代までに収束することを目標としているが、十分に近づくには150世代程度必要であることが分かった。

4 まとめ

VOCALOIDにおける歌声パラメータの最適化を手動で行う際のユーザの負担を軽減するため、IECを用いて自動で最適化を行うシステムを実装した。ユーザの負担の評価として、収束までにどの程度の評価回数が必要であるかを測定する目的で、調整済みのVSQファイルのピッチバンドパラメータの再現を行い、世代数と再現度の関係を調べた。

5 今後の課題

5.1 GP の改良

今後、さらに少ない世代で質の良いパラメータ調整を行うために GP を特化していくことが不可欠である。いくつかの改良案を述べる。

まず、前の時点のパラメータ情報を利用して、漸化式を解くようなアプローチで関数を作り上げることが考えられる。通常この手法は株価予測など時系列予測に用いられており、今回の声質調整の場合においても直前にどのような歌いかたをしていたかという情報を利用することで滑らかな歌の流れを作ることが出来るし、逆にそこから外すことで独特な歌わせかたをすることもできる。

次に、Estimation of Distribution Algorithms(EDA) という手法 [9] を用いることが考えられる。これは通常の GA の交叉をより効率良く行うためのアルゴリズムである。具体的には優良個体の情報からその分布を推定し、一回の世代交替でより最適解に近いものを産む。この手法は世代数はあまりかからないが、分布を推定するという部分で計算時間がかかるため、トータルの計算時間としては GA と比べて良くなるという保証はない。ところが今回のシステムは IEC を用いたものであるため、計算時間はユーザの評価時間と比べて非常に小さくなっており、1 世代ごとの計算時間を減らすよりも世代数を減らす方がユーザの負担が減る。

以上のような改良を GP に加え、ユーザの負担の軽減に繋げていきたいと考えている。

5.2 インタフェースについて

IEC においてはインタフェースの構成もユーザの負担に大きく影響すると考えられる。

IEC においてはまず、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。全体的なイメージを変えたい、一部分を変えたいといったように進化させる範囲を変えたいとできれば、よりユーザの高い満足度を得られると考えられる。

さらに、見た目に関して、ユーザは現在の個体の持つパラメータがどのようなものであるかが直感的に分かる工夫が必要である。どの部分のパラメータが気に入らなくて、これからどのようなことを考えて進化させていけばいいのかを理解出来れば、よりスムーズに進化プロセスを進めることが可能である。

参考文献

- [1] Hideki Kenmochi, Hayato Ohshita, "Singing synthesis system 'VOCALOID'", 2007-MUS-72-(5), pp. 25-28
- [2] Hideyuki Takagi, Tatsuo Unemi, Takao Teran, "Perspective on Interactive Evolutionary Computing", 人工知能学会誌, September, 1998
- [3] Maiko Sugawara, Miki Mitsunori, Tomoyuki Hiroyasu, "Interactive genetic algorithm using generates initial individual based on a favorite color image", JSAI 2008, 2B1-2
- [4] 安藤 大地, Palle Dahlsted, Mats Nordahl, 伊庭 斉志, "対話型進化論的計算によるクラシック音楽作曲支援システム:CACIE", 2005-MUS-59-(10), pp. 55-60
- [5] "band in a box", <http://www.cameo.co.jp/PG/win/>
- [6] John R. Koza, Riccardo Poll, "GENETIC PROGRAMMING" <http://www.springerlink.com/content/164g618r24182v26/>
- [7] Tomoyasu Nakano, Masataka Goto, "VocaListener: An Automatic Parameter Estimation System for Singing Synthesis by Using User's Singing and Its Lyrics", September, 2008
- [8] Larranaga. P, and Lozano. J. A, "Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation" Kluwer Academic Publishers, 2001