

MFCCを距離尺度に用いたGAによるエレキギターの音色再現システムの開発

有山 大地^{†1,a)} 安藤 大地^{†1} 串山 久美子^{†1}

概要: エレキギターの音作りは一般的にアナログのアンプ、エフェクタなどを用いて行われ、その組み合わせは膨大な数に上り、演奏者やエンジニアが求めている音色を得るために大きな労力を必要とする。例えば、憧れのギタリストの音色を再現したいという場合、まず同一の機材の入手、加えてステージセッティングの画像などからエフェクタのツマミの位置を推測する必要があり、手動での再現のプロセスは困難なものである。そこで音色再現の問題を解決するため、機械学習により手持ちの機材から様々なエフェクタ・パラメータの組み合わせを自動的に評価することで、同一機材・同一設定を用いずに音色を再現するシステムの開発を行っている。また、システムの開発を通じた手法の研究を行っている。本稿ではギターエフェクトの一つである「ディストーションエフェクト」を中心に、開発したシステムによって再現された音色の再現度について行ったアンケート調査をもとに、開発中の再現手法について考察を行う。

1. はじめに

エレキギターという楽器は、弦の振動をボディを通した空気振動経路で発音するのではなく、マグネティックピックアップの電磁誘導を使い直接電気信号に変換するという特性がある。そのため「エレキギター」と「ギターアンプ」の間に「エフェクタ」と呼ばれる波形を加工する装置を繋ぎ、音に様々な変化をもたらすことが可能である。

エレキギターの音を決定する大きな要素としては、「パーツ」「ギターアンプ」「ピックアップ」「エフェクタ」の4つを挙げることができる。そして自分が目指す音を再現する作業を「音作り」と呼び、音作りはこの4要素の組み合わせで行われる。この4要素はそれぞれ膨大な数が存在し、また、「ギターアンプ」と「エフェクタ」はそれぞれ無段階で調節できる複雑な複数のパラメータを持つ。エレキギターの音作りは一般的に「エフェクタ」を複数台接続して行われ、演奏者やエンジニアが求めている音色を得るためには非常に労力がかかる。

1.1 背景と問題点

音作りの指標として、一般的に好まれているのは、「プロギタリスト」と呼ばれる人々の音作りである。多くのユーザは、自らエフェクタを用意し、パラメータを調節することで、自分の目指すギタリストの音色を模索することにな

る。しかし、パラメータの組み合わせや、エフェクタの組み合わせは近現代の音楽シーンにおいて限りなく巨大化している。

そのため、例えば憧れのギタリストの音色を再現したいという場合、機材の情報やアンプ、エフェクタ接続順、つまみの位置などの写真によるパラメータ推定が必須であり、これらの情報なしには再現が困難である。多くのギタリストは自ら、あるいは熱心な研究者の推測によって、「アンプ」「ピックアップ」「エフェクタ」の組み合わせを公開している。しかし、実際に公開された組み合わせを試しても、多くの場合期待した結果は得られない。というのも、ギターの音色とは演奏者の癖や、使用している機器の個体差によって微妙な違いが生じてくるためである。また、プロのギタリストの使用する機材は非常に高価な品・ワンオフ品・希少品である可能性もあり、そもそも手に入れることが困難な場合も容易に想像できる。

1.2 目的と手法

そこで本研究では、エレキギターの音作りを手助けする為、「パーツ」「アンプ」「ピックアップ」「エフェクタ」等の組み合わせが一切不明である状況においても理想とするギターの音色を再現できるシステム、つまり機材や環境に左右されない音色再現のシステムをすることを目指している。

また、本研究で想定する状況では、前述の「パラメータの組み合わせが限りなく多様化する」という大きな問題が浮上する。音色再現システムを開発するに当たり、このパ

^{†1} 現在、首都大学東京システムデザイン研究科

^{a)} scarcement0623@gmail.com

ラメータの多様化に対応することも大きな課題の一つである。そこで、開発するシステムは可能な限り自動化されることが望ましい。自動化の手法としては、現在「進化論的計算」を採用している。理想とする音色を目標音源として入力し、これを正解としてパラメータを変化させ、自動的な探索を行うという手法である。

2. 音色の類似度の測定

再現手法について、まずはどのような形に録音データを変換し、どのような尺度で類似度を測るかについて述べる。

音色の類似度を測定する為には、目標となる録音と自身の所有するギターの録音の2つのデータを入力としてとり、何らかの形で比較することになる。一般的な再生/録音環境を想定すると、録音データは非圧縮のPCM音源であるWAV形式のデータを利用することになる。

しかしながらWAV形式の音声データは非常にデータ量が多く、例えば音楽制作で多く使われる解像度では1秒間に1,411,200 (bit) のデータを持つ。ここで、再現したいのは波形そのものではなく、その音色「らしさ」であることに注意する。したがって「らしさ」を表す重要な音響特徴量を抽出し、その類似度により機械学習を行う必要がある。

そこで、本研究では音色の類似度に特化した音響特徴量として、Mel-Frequency Cepstrum Coefficients (MFCC) を選択した。MFCCとは主に人間の喋り声を対象に開発されたので、音声から声道特徴量を強調して取り出し、音声を人間の知覚に特化した形で圧縮できるという特徴がある。通常1つの短い音声を対象に変換が行われ、その特徴量を12次元の実数ベクトルで表すことができる。

しかし、ギターの音源のような長さ不明のデータ全体をまとめて12次元に圧縮した場合、計算時間や処理するデータ量を推定することが難しくなる。また、もともとMFCCは声道特徴量を抽出するという意図から窓サイズが大きい音声を圧縮することは想定されていない。そこで窓サイズを小さくし、ベクトルの遷移として表現する必要があるため、音源データを4096個ずつ(約0.1秒分に相当する)のグループに分け、入力音源の中心の163,840データ(約4秒)分MFCCの計算を行った。つまり一つの音源データからMFCCとして12次元のベクトルが40本抽出される。これでどのような長さの音声データを入力しても480個の浮動小数点数でMFCCが抽出されるようになった。

3. 距離尺度の検討

特徴量として音声データを12個の浮動小数点数からなる40セットのデータに圧縮したが、このベクトル特徴量を用いて、教師データとの類似度(距離)を適切に算出する必要がある。様々な手法を試した結果、抽出したMFCC

間距離を算出する方法としてはEarth Movers Distance (EMD)を採用することとした。

ユークリッド空間におけるベクトル間の距離の測り方としてはユークリッド距離が挙げられるが、ここで注意すべき点は、ユークリッド距離は単純にベクトル間の距離を1対1で測る点である。抽出したMFCCは40本のベクトルが並んでいる状態になっているため、目標となる音声の40本と、自身の持つギターの音声の40本のデータのサンプル箇所が一致していない場合、正しく距離を測る事が出来ない。音源に僅かでもずれが生じた場合正しく距離を測れない方法は、実際の使用シーンを考えると合理的ではない。

そこでベクトル並びの集合の類似度を測る手法としてEMDを採用した。EMDはRubnerらによって提唱された手法[1]で、ある点からある点への輸送量という視点で集合間の類似度を測る事が出来る。

4. 機械学習によるパラメータの自動推定

3節で検討した距離尺度を用い、その有効性を検証する為、これを評価関数として機械学習によるパラメータ推定システムを構築した。機械学習の手法については、本研究では遺伝アルゴリズムを採用した。遺伝アルゴリズムは多点同時探索であるため、エレキギターのエフェクトパラメータのような多峰性が予測される問題に適していると考えられる。以下にその手順について述べる。

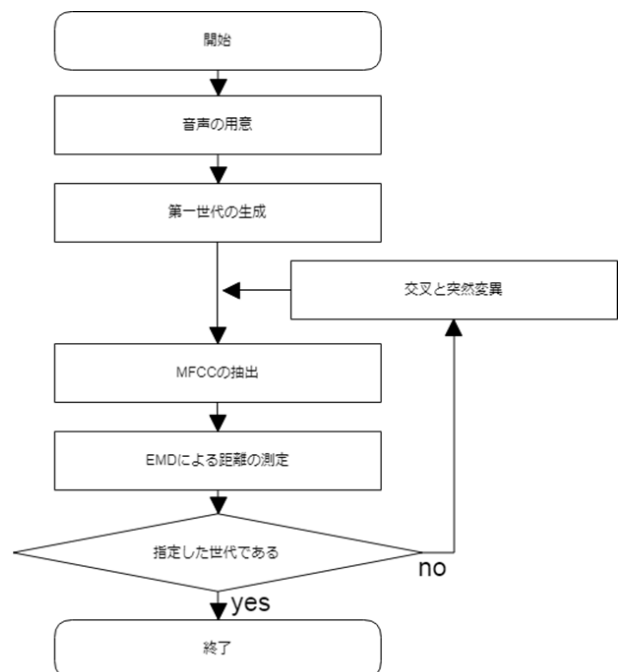


図1 遺伝アルゴリズムによるパラメータ推定の流れ

まず、「目標とする音源」と「自身の所有するギターの音源」の2つの音源を用意する。前者を評価の基準とし、後者にランダムなエフェクトをかけて類似度の比較を行うためである。用意する音源は計算領域を揃える為、所謂「CD

音質」と呼ばれるサンプルレート 44,100Hz の wav 形式に限る。また、エレキギターの音声はその性質上モノラル 1ch であるため、目標とする音声もこれに準拠するものとする。音声の処理には青木による C 言語のライブラリ [2] を利用した。

検証を行うために自由にパラメータを変更できるディストーションエフェクタを C++ にて実装した。実装したエフェクタはハードクリッピングによる歪み、レベルに加えて 10 帯域のイコライザ、コンプレッサーをパラメータとして持つ。イコライザはピーキングフィルタを使用し、中心周波数はギターの周波数特性を考慮し、31.25Hz, 61.5Hz, 125Hz, 250Hz, 500Hz, 1kHz, 2kHz, 4kHz, 8kHz, 16kHz に設定した。強度は -12.0dB から 12.0dB の間を浮動小数点数でとる。

続いて「自身の所有するギターの音源」に対してランダムなパラメータのディストーションを適用して第一世代を生成する。1 世代を n 個の個体と設定した場合、 n セットのパラメータが生成され、自動的にナンバリングされ、wav 形式で書きだされる。用語の混乱を避ける為、ここで生成されるパラメータ 1 つ 1 つを遺伝子、1 音声分のパラメータの集まりを染色体、染色体とそれを用いて出力される音響をまとめて個体と呼ぶ。

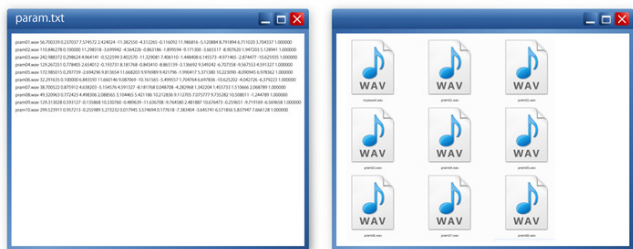


図 2 生成されるパラメータと書き出される wav ファイルのイメージ

続いて書き出された wav ファイルを読み込み、2 節で説明した方法で MFCC を抽出する。MFCC は 1 つの wav ファイルにつき 40 本の 12 次元ベクトルで表され、同一の txt ファイルにまとめて書き出される。txt ファイルは 40 本ごとに改行され、wav ファイルごとのデータの区切りを持つ。ここで同時に「目標とする音源」についても MFCC を抽出し、比較用のデータの準備を行う。

表 1 抽出される MFCC の一例

	1	2	3	4	5	6	7	8	9	10	11	12
param01.wav	-29.3574	6.935212	-6.68934	3.478735	-3.39752	0.989509	-0.00398	-2.44473	-0.12749	-1.32881		

これらの抽出された MFCC について EMD による距離の測定を行い、類似度を求める。求めた類似度はテキストデータとして書きだされ、第二世代以降はこれを基に世代の交配を行う。実装は伊庭による解説 [3] を参考に、予備実験の結果も考慮して行った。次世代に残す個体はエリート選択方式とルーレット選択方式の両方を用いて選択する。

次世代のうち 1 つは最も類似度の高かった個体をそのまま引き継ぐ。続いて全体の 8 割の個体は類似度を基に算出したスコアを基準として、ルーレット選択方式で選んでいく。N 番目のスコア $s(n)$ を求める式は $femd$ を EMD によって導かれた距離として、以下の通り設定した。

$$s(n) = 100 / femd + 1 \quad (1)$$

上式よりスコアの最大値は 100 となり、類似度が低いほどスコアは小さくなる。また、スコアの合計を $fsum$ とした場合に、各個体がルーレット盤面に占める割合 $f(n)$ は次式で求められる。

$$fsum = \sum_{i=1}^n s(i) \quad (2)$$

$$f(n) = s(n) / fsum \quad (3)$$

ここで、0 から $fsum$ までの数字をランダムに生成し、上式をもとに対応する染色体の選択を行う。染色体の交叉は 2 点交叉を用いて行った。ルーレット選択によって個体を順に 2 つずつ選び、これを親として交配を行った。染色体を交叉する点数は 2 点で固定であるが、交叉する範囲はランダムに決定される。

個体は歪み、レベル、10 帯域のイコライザ、コンプレッサーの合計 14 個の遺伝子を持つ。つまり交叉する点の候補は 13 か所から選ぶことができる。そこで 1 から 13 までのランダムな数値を 2 つ生成し、その数値を基に 2 点交叉を行う区切りを決定した。これを個体の数が第一世代と同じになるまで行う。エリート選択した 1 つの個体を除き、残るすべての個体に対して突然変異の判定を行う。突然変異は各個体の遺伝子の合計からランダムに任意の数の遺伝子を変異させる。この突然変異の確率 $fmutation$ は大向による研究 [4] を参考に、遺伝子のビット長を L として次式で求められる。

$$fmutation = 1 / L \quad (4)$$

また局所的最適解からの脱出を狙い、全体の約 2 割を新規の個体と置き換える。突然変異を終えたところで世代の生成は完了とし、再度類似度の計算を行う。これを繰り返して最適解を導く。

実際の運用シーンでは、エフェクタの個々のパラメータはそこまで精度が高い実数である必要はない。またギターのエフェクタの場合、一つのエフェクタを駆動するパラメータの数は少なくパラメータ間の関係性にそこまで気を使う必要はない。そこで低精度の組合せ最適化として計算コストが一番低い遺伝的アルゴリズムを今回は採用した。

5. 自動推定の実験

設定した距離尺度が適切であるか検証する為、実際の使用シーンを想定した状況で 3 種類の実験を行った。

5.1 実験1：同一のギターを用いた再現実験

まず同一のギターを用いて、正しくパラメータの推定が行えるか実験を行った。1世代の個体数を100個体に設定し、100世代まで交配を行った。結果を表2に示す。

表2 同一ギターを用いた自動再現の結果

	歪み	レベル	31.25Hz	62.5Hz	125Hz	250Hz	500Hz	1,000Hz	2,000Hz	4,000Hz	8,000Hz	16,000Hz	スレシヨルド	レシオ	距離
目標音源	79.7	0.1	-3.5	-11.2	-8.3	8.0	2.4	-8.9	-11.8	-9.3	-1.2	-1.1	0.4	0.7	0.0
元の音源	1.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	191.4
再現音源	65.1	0.1	-4.7	-1.6	-8.4	10.7	5.5	-9.2	-5.4	-5.0	-3.7	1.9	0.7	0.7	3.9

表2における「目標音源」は自作のディストーションエフェクトを適用した音源、「元の音源」はエフェクトを適用する前の音源、「再現音源」は自動推定を行い100世代目で最も類似度が高いと判定された音源である。「元の音源」と「目標音源」の距離は191.4と大きな数値になっていることがわかる。これに遺伝アルゴリズムによって自動的に推定されたパラメータを適用することにより、距離を3.9と大きく縮めることに成功している。求められたパラメータも「目標音源」と近いものが推定できている。

続いて各音源のパワースペクトルの類似度をコリレーション分析によって求めた結果を表3示す。

表3 各音源のパワースペクトルの類似度をコリレーション分析によって求めた結果

	目標音源	元の音源	再現音源
目標音源	1		
元の音源	0.938896	1	
再現音源	0.996622	0.937204	1

コリレーション分析における類似度は1で完全一致であり、「再現音源」と「目標音源」の類似度は非常に高い数値を示していることがわかる。これにより波形的な側面からもMFCCとEMDを距離尺度に用いた自動推定の有効性を確認できた。

5.2 実験2：異なる特性を持つギターでの再現実験

続いて低音に特徴のある中空構造のギター*1（以下A）と、高音に特徴のあるソリッドギター*2（以下B）の2つ異なるギターを用いても再現が可能であるか実験を行った。

1世代の個体数を100個体に設定し、100世代まで交配を行った。結果を表4に示す。

表4 異なるギターを用いた自動再現の結果

	歪み	レベル	31.25Hz	62.5Hz	125Hz	250Hz	500Hz	1,000Hz	2,000Hz	4,000Hz	8,000Hz	16,000Hz	スレシヨルド	レシオ	距離
目標音源	1.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
元の音源	1.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	15.5
再現音源	3.6	0.2	11.5	-3.3	-8.6	-6.4	2.9	9.6	9.7	7.0	4.9	-2.4	0.0	0.7	9.5

表4における「目標音源」はBのクリーン音源、「元の音源」はAのクリーン音源、「再現音源」は自動推定を行い100世代目で最も類似度が高いと判定された音源である。高音に特徴のあるBの音色に合わせ、Aの1k-4kHzが強調されるようにイコライザが適用され、距離が縮まってい

*1 Gibson ES-335

*2 ELK Deluxe

ることが見て取れる。

この実験により、遺伝アルゴリズムによるパラメータ推定が、異なる特徴を持つギター間の音色の再現においても一定の効果を持つことが確認された。

5.3 実験3：市販のエフェクターを用いた再現実験

最後に実際に市販され、広く使用されている歪みエフェクタ5種類を目標として、開発した自動再現システムを用いて音色の再現が可能か実験を行った。なお、エフェクタのパラメータは全てのつまみを12時とした。

全ての実験において1世代の個体数は100個体、100世代まで交配を行った。

5.3.1 BOSS:DS-1

BOSSのエフェクター「DS-1」*3を目標とし、再現を行った。結果を表5に示す。

表5 DS-1を目標とした自動再現

	歪み	レベル	31.25Hz	62.5Hz	125Hz	250Hz	500Hz	1,000Hz	2,000Hz	4,000Hz	8,000Hz	16,000Hz	スレシヨルド	レシオ	距離
目標音源	1.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	121.9
元の音源	1.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	121.9
再現音源	156	0.1	8.5	1.7	-1.9	-3.7	-6.1	-4.2	1.4	11.4	7.6	6	0.8	0.9	16.8

5.3.2 BOSS:MT-2

BOSSのエフェクター「MT-2」*4を目標とし、再現を行った。結果を表6に示す。

表6 MT-2を目標とした自動再現

	歪み	レベル	31.25Hz	62.5Hz	125Hz	250Hz	500Hz	1,000Hz	2,000Hz	4,000Hz	8,000Hz	16,000Hz	スレシヨルド	レシオ	距離
目標音源	1.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	136.8
元の音源	1.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	136.8
再現音源	83.7	0.9	10.8	3.4	8.3	4.8	-7.5	-8.6	-7.0	-10.8	-11.9	0.2	0.1	0.3	19.9

5.3.3 ELECTRO-HARMONIX:Big Muff

ELECTRO-HARMONIXのエフェクター「Big Muff」*5を目標とし、再現を行った。結果を表7に示す。

表7 Big Muffを目標とした自動再現

	歪み	レベル	31.25Hz	62.5Hz	125Hz	250Hz	500Hz	1,000Hz	2,000Hz	4,000Hz	8,000Hz	16,000Hz	スレシヨルド	レシオ	距離
目標音源	1.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	152.3
元の音源	1.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	152.3
再現音源	263	0.2	-8.8	3.5	10.8	-11	-0.9	-7.9	-9.4	9.1	10.4	-3.0	0.6	0.3	10.9

5.3.4 Pro Co:RAT

Pro Coのエフェクター「RAT」*6を目標とし、再現を行った。結果を表8に示す。

表8 RATを目標とした自動再現

	歪み	レベル	31.25Hz	62.5Hz	125Hz	250Hz	500Hz	1,000Hz	2,000Hz	4,000Hz	8,000Hz	16,000Hz	スレシヨルド	レシオ	距離
目標音源	1.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	93.6
元の音源	1.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	93.6
再現音源	47.7	0.6	3.2	8.3	3.2	-8.8	-10	-1.8	1.0	11.2	2.7	-12.0	0.5	0.6	21.5

5.3.5 IBANEZ:TS808

IBANEZのエフェクター「TS808」*7を目標とし、再現を行った。結果を表9に示す。

表9 TS808を目標とした自動再現

	歪み	レベル	31.25Hz	62.5Hz	125Hz	250Hz	500Hz	1,000Hz	2,000Hz	4,000Hz	8,000Hz	16,000Hz	スレシヨルド	レシオ	距離
目標音源	1.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	34.7
元の音源	1.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	34.7
再現音源	3.9	0.8	6.5	-9.9	-11.3	-11	1.4	11.2	11.6	8.2	11.2	-2.1	0.1	0.9	21.4

*3 <https://www.boss.info/jp/products/ds-1/>

*4 <https://www.boss.info/jp/products/mt-2/>

*5 <https://kcmusic.jp/ehx/big-muff-pi.html>

*6 <https://www.ratdistortion.com/>

*7 https://www.ibanez.com/jp/products/detail/ts808_99.html

6. 考察

実験3について、すべての機種について多少の差はあるものの、開発した手法で距離を縮め、同一機材、同一設定を用いずに類似度を高めることができることを示した。実験の成果について中学校・高等学校で軽音楽部に所属する（或いはしていた）複数人にアンケートでの聞き取り調査を行ったところ、概ね再現できているとの回答を得た。

しかし、比較的歪みの量が少ない「TS808」について、71.5%の人が似ていない、或いはどちらかといえば似ていないと回答した。このほかにも歪みの量が少ない場合に再現が不安定になるケースが確認できており、今回改めて特徴量について見直しを行った。表3はクリーンの音源とディストーションエフェクトを掛けた音源の2つについて、抽出したケプストラムの0次から19次までの20次元分を比較したものである。

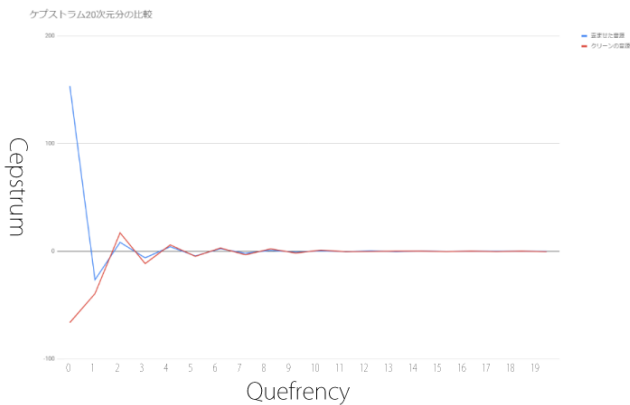


図3 20次元分のケプストラムの比較

音声認識の分野では通常抽出したケプストラムのうちの1次項から12次項までの値を特徴量として用いるため、本研究でもそれに従い同じ条件で比較した。表3から確認できる通り、13次項以降の数値は双方殆ど変動もなく、ギターの音源においても13次項以降は無視できるように思われる。

一方で一番左の0次項は大きく変動している。0次項は「各周波数帯域の平均エネルギーの集合」と相関があるとされ、信頼性の低さから通常は除外される。この0次項については音声認識の分野では様々な議論がある。清華大学のチェンらによる研究[5]では、パワー係数を考慮することで中国語の音節の誤りを約10%低減できるとしている。ギターの音色についても歪みの量が多いほど音圧が上昇し、この音圧と「各周波数帯域の平均エネルギーの集合」には強い相関があることが分かった。そこで実験3の「TS808」について、MFCCの0次項も加えた状態で再現を行った。その結果を表10に示す。

比較するベクトルの次元が増えた為、0次項を考慮する

表10 TS808を目標とした0次項を加えた自動再現

	歪み	レベル	31	25Hz	62	5Hz	125Hz	250Hz	500Hz	1000Hz	2000Hz	4000Hz	8000Hz	16000Hz	スレショルド	レシオ	距離
目標音源																	0.0
元の音源	1.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	42.9
再現音源	2.8	0.3	9.9	-2.5	5.7	-2.3	-5.9	5.2	5.8	9.7	5.2	-2.7	0.1	0.7	27.7		

前と比べ、距離は相対的に離れており、パラメータにも差が出ていることがわかる。0次項を考慮した音源について再び聞き取り調査を行ったところ、0次項を加えていないものと比べ、やや類似度が高まったとのフィードバックを受けた。このことから0次項に関してギター音色の再現において有用である可能性が示された。MFCCの0次項の取り扱いについては今後さらなる検証を行う予定である。

7. DAWプラグインへの移植

Steinberg^{*8}の提供するSDKである「VST SDK3.6」を元にコーディングを行った。「VST SDK3.6」はC++で記述されるため、4節までで開発したシステムをほぼそのまま統合することができた。

大きな違いは入出力デバイスの管理やデータサイズの設定をホストとなるアプリケーションに依存する事である。これまではwavデータのサイズ分のバッファをとり、すべての処理を終えた段階で音声を出力していたが、DAWプラグインとして実装することで、音声は最低限の作業バッファを通過する過程で処理され、入力とほぼ同時に出力される。これによってリアルタイムでのエフェクトの適用を実現している。先読みなどの処理が必要なエフェクトの場合、この入力の処理を変更する必要があるが、今回は4節で実装したものと同一の計算で問題ない。

また、「VST SDK3.6」にはデフォルトでGUIの機能が備わっている。移植に伴い、推定後のパラメータをこのGUIで微調整できるようにした。実装したGUIを用いることで、DAW上で計算に必要な音源の書き出しから、染色体・個体・世代数の設定をシームレスに行うことができる。

8. おわりに

本研究ではPCM音源からMFCCを抽出し、EMDを用いて類似度を測定することでエフェクトの掛けられたエレキギターの音色の類似度を測る手法を提案した。また、遺伝アルゴリズムを手法として選択することで、音色の再現に必要な要素を統合的なパラメータとして設定し、目標とするパラメータの自動推定を行うシステムの開発を通して新たな再現手法を提案した。

現在はシステムにハードクリッピングによるディストーションエフェクトを中心としたエフェクトのみを実装しているため、再現には限界があるが、今後扱うエフェクトを増やすことでより正確な再現が可能になると考える。ギターの音色という物理的でありながら個人の印象に左右さ

^{*8} ドイツのソフトウェア会社、オーディオドライバに関する規格のASIOや、音楽ソフトウェアに関するプラグイン規格のVSTを策定している



図 4 DAW のプラグインとして実装された本システムの GUI

れるものを再現するにあたり，本研究で提案した手法は，組み合わせの多様化や環境などの制約にとらわれない新たな音色再現の手法として提案できる可能性を持つと考える．直近の課題としては，エフェクトやその他の要素，例えばユーザの聴覚印象フィードバックを受けた音色の感情的な側面などについても組み合わせ，複合的な実験を重ねていく予定である．

参考文献

- [1] Goossens, Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas.: *A Metric for Distributions with Applications to Image Databases*, Proceedings of the 1998 IEEE International Conference on Computer Vision, pp.59-66 (1998).
- [2] 青木直史: C 言語ではじめる音のプログラミング, オーム社 (2008).
- [3] 伊庭齊志: C による探索プログラミング 基礎から遺伝的アルゴリズムまで, オーム社 (2008).
- [4] 大向一輝: 分散遺伝的アルゴリズムにおける最適な交叉率および突然変異率 (online), 入手先 http://mikilab.doshisha.ac.jp/dia/monthly/monthly99/19990830/ohmukai_21.pdf (1999.08).
- [5] Fang Zheng, Guoliang Zhang and Zhanjiang Song.: *COMPARISON OF DIFFERENT IMPLEMENTATIONS OF MFCC*, J. Computer Science and Technology.(2001).
- [6] 中西正洋, 門田暁人, 松本健一, 井上克郎: ギターの演奏情報の抽出と分析, 情報処理学会第 59 回全国大会 (1999).