

深層学習を用いたアナログエフェクターの非線形モデリング

柴田 幸奈

Yukina Shibata

法政大学情報科学部デジタルメディア学科

yukina.shibata.6k@stu.hosei.ac.jp

概要

A method for estimating Tone and Drive parameters from audio signals of a physical analog guitar effect pedal is proposed. To capture nonlinear characteristics of the device (BOSS OD-3), an efficient data collection method is introduced, utilizing sustain sounds synthesized from clean guitar signals and recording audio while continuously manipulating control knobs. A CNN model is constructed using noise-filtered clean and processed signals as two-channel inputs, formulating the estimation as an 11-class classification problem to enhance robustness against variability inherent in analog circuits. In evaluation experiments, mean Absolute Error (MAE) was 3.8 for Drive and 4.4 for Tone, demonstrating performance comparable to estimation accuracy of digital simulations in prior research. In conclusion, the results show that an 11-class classification approach is effective for parameter estimation of analog guitar effect pedals. Future work includes applying the proposed method to real guitar performances and further optimizing the network architecture to stabilize estimation accuracy.

1 序論

1.1 研究背景

ポピュラー音楽やロックにおいてギターは中心的な楽器である。また音楽制作の現場、特に録音やミックスの過程では、ギターから出力されたそのままの音（クリーンサウンド）だけではなく、その音色を意図的に変化させることが広く行われている。この音色の加工に用いられる機材がエフェクターである。エフェクターは、入力された音声信号に対して音を意図的に歪ませて力強いサウンドを生み出す歪みや、音の減衰を遅らせて長く響かせるサステインなどの処理を施す。これらのエフェクター処理は、アナログ回路やデジタル処理によって実装されている。

プロのアーティストやバンドが発表した楽曲のサウンドは、多くのギタリスト、特にコピーバンドなどで活動する人々にとって再現の対象となる。その際、多くのギタリストは実際にエフェクターを操作し、試行錯誤を繰り返しながら目標の音

を探求する。

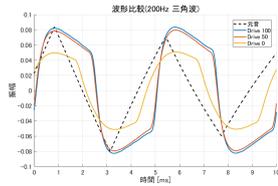
しかしこの試行錯誤のプロセスは、大きな壁となる。この難しさの要因は、エフェクターの内部回路が持つ二重の複雑さに起因する。第一に、回路の心臓部であるオペアンプとダイオードによるクリッピング回路は、音色そのものを非線形に変化させる物理的な要因である。入力された信号はオペアンプで増幅され、その振幅がダイオードのしきい値を超えた部分が強制的に潰されることで、特有の歪みが生まれる。第二に、その効果を調整する Drive つまみ自体も、操作性を考慮した非線形なカーブを持つことが多く、ギタリストの感覚と実際の音の変化が一致しない要因となっている。これら回路の物理的な非線形性と、操作上の非線形性が組み合わさることで、パラメータの特定を非常に困難にしている。

さらに、これらの非線形性が音声信号に与える影響は、入力される波形の複雑さに依存して増大する。図 1a と図 1b は、単純な周期波形と複雑なギター音における、エフェクト付与による波形変化の違いを示したものである。図 1a は 200Hz の三角波、図 1b はギター演奏音を入力とし、Drive という歪みに関わるパラメータを 0, 50, 100 の 3 段階で変化させた際の出力波形の一部である。三角波のような単純な周期波形では、歪み量の増大に伴う波形変形は比較的規則的であり、元の波形の概形が維持されている様子が確認できる。しかし、ギター演奏音のように倍音成分を豊富に含んだ複雑な波形を入力とした場合、設定値ごとに波形の概形が劇的に、かつ非一律に変化する。このように、入力信号が複雑になるほど回路内での非線形な干渉が強まり、出力信号のみから元の設定値を一意に特定することを困難にしている。

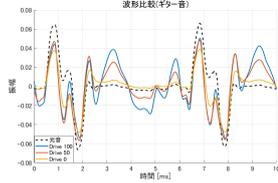
1.2 研究の目的

上記課題を解決するため、本研究では、エフェクター処理によって生み出されたギターの音響情報から、その元となったエフェクターのパラメータを自動的に推定する手法の確立を目的とする。具体的には、広く普及している実機のオーバードライブエフェクターである BOSS 社製の OD-3 を研究対象とする。この OD-3 は音量を調整するレベル、高周波成分をカットし、音の明るさを変えるトーン、歪みの量を操作するドライブという 3 つのパラメータがある。そのうち、直接音色の変化に関わるトーンとドライブの 2 つを推定対象とする。

本研究では、機械学習モデル、特に画像認識などで高い性能を示すことで知られる CNN（畳み込みニューラルネットワーク）を中心としたモデルを用いる。クリーンなギター音源と、それを OD-3 で加工した音源から抽出した音響特徴量をもと



(a) 200Hz 三角波



(b) ギター音

図 1: Drive パラメータ設定による出力波形の比較

に、パラメータ値を回帰的に予測するモデルを構築する。これにより、ギタリストが目標とするサウンドを手軽に再現するための支援を目指す。評価方法は演奏者の主観的な判断に頼るのではなく、客観的な数値指標に基づいて行う。

2 関連研究

2.1 エフェクトパラメータ推定に関する先行研究

ギターサウンドの音響情報からエフェクトの種類やそのパラメータ設定を推定する研究は、近年、特に機械学習を用いて進展している。Jürgens ら [1] の研究では、まずエフェクトの種類を特定し、その後に特定されたエフェクトに応じたアルゴリズムでパラメータを推定するという 2 段階のアプローチを提案している。エフェクトの分類には MFCC (メル周波数ケプストラム係数) などの音響特徴量を抽出し、サポートベクターマシン (SVM) が用いられた。パラメータ推定の段階では、回帰問題として浅いニューラルネットワークが利用されたが、この手法ではディストーション、ディレイ、トレモロといった各エフェクトに対して、それぞれに合わせた特徴量を選択する必要があった [1]。この課題に対し、Hinrichs ら [2] は、CNN を用いることで特徴量の自動抽出を可能にし、単独のエフェクトだけではなく、混合音源からの分類やパラメータ推定へ適用範囲を広げている。

さらに、Comunità ら [3] は、13 種類のオーバードライブ、ディストーション、ファズといった非線形エフェクトを対象に、畳み込みニューラルネットワーク (CNN) を用いた。入力にはメルスペクトログラムを用い、CNN が画像のようなデータから自動で特徴を学習する。この研究の特筆すべき点は、Conditional Network (SetNetCond) と呼ばれるモデルを導入したことである。このモデルは、エフェクトの種類を条件として与えることで、単一のネットワークアーキテクチャで、対象となる全てのエフェクトのパラメータ推定を可能にした。これにより、Jürgens らの研究 [1] で見られたエフェクト毎の個別な特徴量設計という課題を克服し、平均絶対誤差 0.05 以下という高い精度を達成している。

2.2 本研究の意義

先行研究においても、パラメータ推定の対象は一貫して PC 上で動作するデジタルエフェクター (オーディオプラグイン) であった。加えて、Comunità らの研究 [3] ではデジタルプラグインを対象としたため、著者自身が考察で述べているように、エイリアシングというプラグイン特有の雑音が推定結果に影響した可能性が指摘されている。

また、シミュレーションモデルの限界についても議論がある。Wright らおよび Damskäg らは、歪み回路を持つ複雑な非線形挙動やメモリ効果について、SPICE 等の回路シミュレーションでは完全な再現が困難であると指摘している。特に Damskäg らは、実機の入出力データを学習するブラックボックスモデリングの方が、回路図に依存する手法では再現しきれない、実機特有の複雑な非線形挙動や個体差を含んだ現実的な歪みを捉えるために不可欠であることを示唆している [4, 5]。

これまでに述べた先行研究では、主にデジタルエフェクターを対象としたパラメータ推定が検討されてきた。一方で、多くのギタリストが実際に使用し、その音の再現を目指しているのは物理的なペダルである。この実用上の需要にも関わらず、実機を対象とした研究はこれまで進んでいなかった。これは、アナログ回路を持つ電源の状態や電子部品の個体差による揺らぎなど、データを精密に扱う上での技術的な困難さが伴うことが、その一因であると考えられる。また、Juvola ら [6] が示唆するように、物理的なつまみ操作を含む実機の制御空間を体系的にデータ化すること自体が、研究遂行上の大きな障壁となっていた。

上記課題を解決するため、本研究では、エフェクト処理によって生み出されたギター音響情報から、その元となったエフェクターのパラメータを自動的に推定する手法の確立を目的とする。

3 実機を対象としたパラメータ推定手法

本研究における提案手法の全体的な処理フローを図 2 に示す。本手法は、実機からのデータ収集、前処理、特徴量抽出、そして CNN によるパラメータ推定の 4 つのフェーズで構成される。

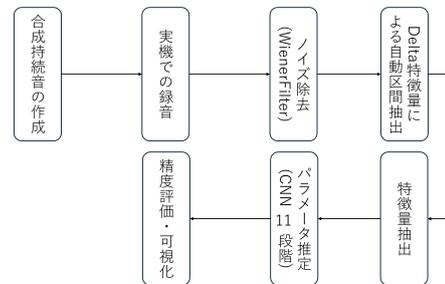


図 2: 提案手法

3.1 データセットの構築

本研究におけるデータ収集のための実験は、外部の音響的な影響を排した環境で行った。具体的には、アンプやスピー

カー、マイクを使用せずに、オーディオインターフェイスを介したライン接続により、純粋な電気信号としてデータを取得した。

本研究では、ギター音の自然な減衰が特徴量に与える影響を排除し、純粋なエフェクターの特性のみを観測するため、入力信号の品質が担保された研究用データセットである、Guezinらが公開した「EGFXSET」[7]に含まれるクリーンなギター音源から、基準として1弦開放弦の音高(E4, 約330Hz)の信号を選択し音が安定しているアタック直後などの一部分を切り出した。切り出した部分を用い、音量が減衰しない持続音(サステイン音)を人工的に合成し、これを実験用の入力音源として使用した。ただし、この合成持続音は単一波形の繰り返しであるため、アナログ回路が持つ可能性のある長期的なメモリ効果を完全には捉えきれないという課題がある。本研究ではまず単純化した条件で検証を行い、今後の検討課題とした。

次に、この合成したギター音源をPCから再生し、音声信号をオーディオインターフェイスを介して出力し、実機のオーバードライブエフェクターに入力した。エフェクターで加工された音声信号は、ライン出力からレコーダーに送られ、録音された。

機械学習モデルの学習のためのデータセットを作成した。対象とするアナログエフェクターは、デジタルエフェクターのような数値入力ではなく物理的なつまみ操作を行うため、目盛りが存在せず、特定の設定値を正確に再現することが困難である。加えて、Juvelaら[6]が指摘するように、実機の制御空間を網羅的にデータ化すること自体も困難な課題である。

これらの問題に対応するため、本研究では、つまみを最小値から最大値まで連続的に操作した音源を録音する手法を採用した。加えて各操作の開始前と終了後に、数秒間の静止区間を設けることで、パラメータが0%と100%の状態が安定している区間を明確化した。録音は、Toneパラメータのみを操作した音源(Driveは50に固定)を20本、Driveパラメータのみを操作した音源(Toneは50に固定)を20本、合計40本の音声ファイルを約25秒かけて録音した。録音された全ての音声データに対し、入出力の音量差による推定精度への影響を排除するため、振幅の最大値を基準とした一律の正規化を行った。

次に、収録した連続音源を離散的なデータセットへ変換するため、音声信号を0.2秒の短い静的なフレームにスライスする手法を採用した。録音データ全体の長さは約25秒であるが、前後の静止区間を除いた実質的なつまみ操作時間は約20秒である。そのため、0.2秒という時間は、操作範囲(0から100とする)を移動する際の1単位あたりの変化時間($20/100 = 0.2$)に相当する。各フレームに対しては、音色のスペクトル包絡を効率的に表現可能な音響特徴量であるメル周波数ケプストラム係数(MFCC)を抽出した。

この処理により、各パラメータの設定値に対応する合計440個のMFCC特徴量データ(以下、MFCC画像)を生成し、これを学習用データセットとした。本手法により、実機の制御空間を網羅的かつ効率的にデータ化することが可能となった。

収録した音声データには、アナログ回路に起因する定常的な背景雑音に加え、録音機材や電源環境などの外部的な録音環境に起因する定常雑音が含まれている。この雑音は、機械学習モデルが音色変化ではなくノイズ特性そのものを誤学習するリスクを高めるほか、正確な特徴量抽出や区間検出を阻害する要因となる。

そこで本研究では、データセットの品質を向上させるための新たな前処理手法を設計した。第一に、時間反転と固定プロファイルを用いたウィーナーフィルタ(Wiener Filter)を導入した。Wiener Filterは、観測された信号から本来の信号成分を推定する際、その統計的な誤差(平均二乗誤差)が最小になるよう設計されたフィルタであり、各周波数帯域の信号対雑音比(SNR)に基づいてノイズ成分を効果的に抑制する。

本実験のデータセットは録音終了付近(パラメータ100%)においてノイズ成分が最も顕著となる。そこで、波形を時間軸上で反転させ、最もノイズレベルが高い区間の信号からノイズプロファイルを推定することで、エフェクターのノイズを除去した。フィルタ適用後、波形を再反転させ元の時間軸へ戻している。

また、一般的なWiener Filterでは、VAD(Voice Activity Detection)を用いてノイズ情報を動的に更新する手法がとられる。しかし、本実験でこの処理を行うと、ギターの音色成分そのものが定常ノイズとして誤認され、目的信号まで過剰に除去されてしまう問題が生じる。そのため、本研究ではノイズプロファイルの学習を非演奏区間かつ最大ノイズ発生時のみに限定し、算出された統計情報を全区間に固定して適用する設計とした。図3はDrive100%時におけるノイズ除去前後の周波数スペクトルを比較したものである。

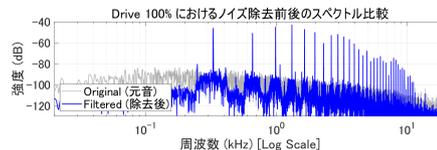


図3: Drive100 地点のスペクトル

図3を見ると、ギター特有の急峻な倍音ピーク(青色の線)は元音の構造を維持したまま保持されていることがわかる。一方で、倍音間の谷間にあたる帯域や、特に1kHz以下の低周波数領域において、背景雑音(灰色の線)が効果的に抑制されていることが分かる。この処理により、後述の区間抽出および学習モデルへの入力において、特徴量抽出への影響を抑えつつ、純粋な音色変化のみへの評価を可能にした。

ノイズ除去後の信号から、手動操作に伴う時間的な誤差を排して正確な正解ラベル(パラメータ値)を付与するため、スペクトル重心とその時間変化量(Delta)を計算し、変動に基づいてつまみを操作している区間と、設定値を固定している区間(0%および100%)を自動抽出するアルゴリズムを実装した。本手法による信号処理および区間抽出のプロセスを図4に示す。図4の上段はスペクトル重心の推移、下段はつまみの回転検知に用いたDelta特徴量を示している。このグラフ上に重ねて表示した緑色の帯はパラメータ0%の静止区間を、

青色の帯はパラメータ 100% の静止区間をそれぞれ自動抽出した結果である。Delta 値が安定している区間からこれらの定常状態を特定することで、手動操作に伴う時間的な誤差を排した正確なラベル付与が可能となった。

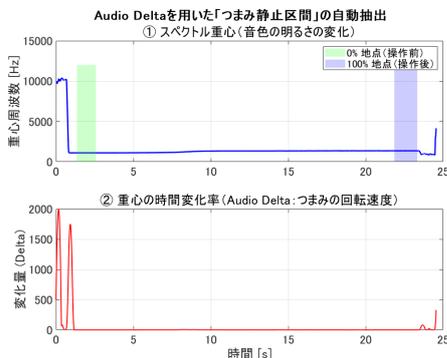


図 4: Delta 特徴量を用いた区間抽出

音声データからパラメータ推定に有効な情報を抽出するため、音響特徴量への変換を行う。MFCC 主要な入力特徴量として採用した。先行研究 [3] 等で用いられる特徴量を参考に、本タスクにおいて音色の変化を捉えるのに適した CNN モデルを設計した。

また現在データセット作成に用いている特定の音程 (E4) と異なる音域の入力に対しても推定を行うため、サンプリングレート変換 (リサンプリング) を用いたデータ拡張手法を導入した。拡張にあたっては、学習データ (E4) に対し、リサンプリングを用いて 1 オクターブ下の音程 (E3) を擬似的に生成し、これを元の学習データセット各 Tone ファイルから 1 つずつ追加して再学習を行った。アナログフィルタ特性 (Tone 回路) に与える影響を比較するため、Tone0 に加え Tone50 の音源にもリサンプリングを行い、拡張の元データとして使用するデータセットも構築した。

3.2 学習モデルの構造

本研究で構築した学習モデルは、先行研究 [3] と同様に CNN を基本とした。図 5 にモデルの全体構造を示す。本モデルは、左端の入力層から右端の出力層に向かって信号が伝播する構成となっている。主要な処理単位として、図中の畳み込み層ブロックは、畳み込み層、正規化層 (Batch Normalization)、活性化層 (ReLU)、プーリング層の 4 つの要素で構成されている。本研究では、入力層から順にフィルタ数が増加する合計 3 層の畳み込み層ブロックと、それに続く 3 層の全結合層を配置した。また、学習の安定化のため、全ての全結合層には過学習抑制のための Dropout 層を導入している。

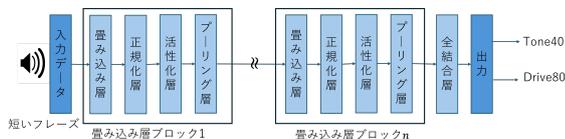


図 5: 学習モデル

モデルの入力特徴量に関しては、音色の変化成分をより直

接的に評価するため、クリーン音源とエフェクト音源双方から抽出した MFCC を入力特徴量として用いた。具体的には、クリーン音とエフェクト音から抽出した MFCC 特徴量を 2 チャンネルの入力データとして統合し、モデルに与えた。モデルの出力は、Tone と Drive の各パラメータに対して個別に学習・推定を行う手法を導入した。

4 実験結果と評価

4.1 モデルの学習精度評価

構築した CNN モデルが、学習した信号の特徴をどの程度正確に抽出できているかを検証するため、学習時と同一の音程である未知のテストデータを用いて評価を行った。本実験におけるモデルの具体的な学習条件およびパラメータの設定について述べる。出力形式に関しては、本研究の予備実験において回帰モデルでは誤差が大きくなる傾向が見られたため、推定精度を向上させる目的で、パラメータ値を 0 から 100 まで 10 刻みで分類する 11 段階 (0, 10, ..., 100) の分類問題として構成した。

実験にあたっては、構築したデータセットを学習用データ、検証データ、テストデータの 3 つに分割して実施した。モデルは学習データのみを用いて学習し、検証データを用いてハイパーパラメータの調整を行った。畳み込み層のブロックを $n = 3$ とした構成において学習を行い、最終的な推定精度は学習に一切使用していない未知のテストデータに対する性能で評価した。

構築したモデルの主要な評価指標としては、正解率 (Accuracy) を用いた。あわせて 11 段階分類の評価にあたっては、平均絶対誤差 (MAE) および二乗平均平方根誤差 (RMSE) を算出した。

Tone0 のリサンプリングで拡張したデータセットを用いて各パラメータの最適化を行い学習を行った結果、テストデータに対する正解率は Drive で 64%、Tone で 57% に達した。

分類結果を数値に換算して算出した MAE (平均絶対誤差) は、Drive で 3.9、Tone で 4.3 であった。これは、正解のクラスから平均して目盛りの約 4~5% 程度の誤差に収まっていることを示している。また、先行研究 [3] において Comunità らは、回帰モデルを用いて MAE 0.05 (0~100 換算で 5.0 相当) という精度を達成している。本研究による分類モデルの MAE は、先行研究と同等の推定精度を実現したと言える。

4.2 未知の音程に対する評価実験

学習に使用した音高以外のデータに対するモデルの汎用性を検証するため、E4 を含む合計 15 種類 (C3 から C5 までの 2 オクターブ分) の音程を用いた推定実験を実施した。本実験では、各音程のテストデータから学習データと同様に分割・抽出したフレームの推定結果を平均し、その性能を評価した。Tone0 と Tone50 の設定で録音された音源に対してもリサンプリングを行い、拡張の元データとして使用したデータセット (Tone0 および Tone50 拡張) での検証を行った。この条件における推定結果の比較を表 1 に示す。また、Drive50 (Tone 0) における音程ごとの時系列推定結果を図 6 に示す。

表 1 を確認すると、推定結果は正解値である 50 付近で推移

表 1: Tone0 及び Tone50 の音源からデータ拡張したパラメータ推定値の平均と誤差

設定 Tone	設定 Drive	Tone 平均 (誤差)	Drive 平均 (誤差)
0	50	22.0 (+22.0)	49.6 (-0.4)
100	50	22.5 (-77.5)	49.4 (-0.6)
50	0	19.8 (-30.2)	49.1 (+49.1)
50	50	22.8 (-27.2)	49.4 (-0.6)
50	100	28.9 (-21.1)	50.7 (-49.3)

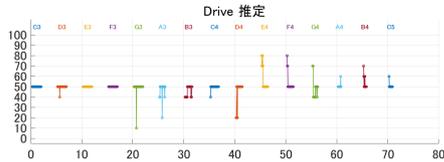


図 6: Tone0 及び Tone50 でデータセット拡張を行った際の Drive の推定結果 (Drive50 Tone0)

していることが確認できる。また数値を見ると、正解値の上昇に伴い推定値も増加しており、パラメータへの追従性が確認できる。また図 6 では、全体の推定値のばらつきが抑えられていることが分かる。この結果より、Drive パラメータの推定に関しては、リサンプリングによって音程の情報を増やすことが有効であると言える。

一方で、Tone パラメータの推定に関しては、推定ができていない。表 1 では、正解値が 50 であるにもかかわらず、推定値は平均 19.8 である。また Tone の正解値が 0, 50, 100 と変化した場合、推定値は 22.0, 22.5, 19.8 となっており、変化の傾向も追えていない。この原因は、リサンプリング処理に伴うフィルタ特性の物理的な変質にあると考えられる。アナログ回路の Tone のように、フィルタ特性を持つ音声をデジタル処理でピッチダウンさせると、音程とともにフィルタの形もずれてしまう。その結果として推定値が低下したと推察される。

5 考察

本実験で提案したデータ拡張手法の有効性を明確にするため、まず拡張を行わない単一音程 (E4) のみで学習させたモデルが、どのような挙動を示したかについて考察する。

単一音程のみのデータセットで学習を行った結果、テストデータに対する正解率は Drive で 69%, Tone で 65% であった。また分類結果を数値に換算して算出した MAE (平均絶対誤差) は、Drive で 3.8, Tone で 4.4 であった。これは、拡張後のデータセットと同様に、正解のクラスから平均して目盛りの約 4~5% 程度の誤差に収まっており、先行研究 [3] 先行研究と同等の推定精度を実現したと言える。

このモデルを用いて、E4 を含む合計 15 種類 (C3 から C5 までの 2 オクターブ分) の音程を用いた推定実験を実施した。結果を以下の表 2 に示す。

Tone パラメータに着目すると、実機の設定値を 0, 50, 100 と増加させるに従い、推定値は 27.5, 28.5, 33.6 と上昇しており、音色変化の方向性自体は概ね捉えられていることが確認できる。しかし、その推定範囲は 20 から 40 の間に留まっており、中央付近へと収束する傾向が見られた。

表 2: 複数音程 (2 オクターブ) におけるパラメータ推定値の平均と誤差

設定 Tone	設定 Drive	Tone 平均 (誤差)	Drive 平均
0	50	27.5 (+27.5)	56.0 (+6.0)
100	50	33.6 (-66.4)	58.7 (+8.7)
50	0	29.9 (-20.1)	59.7 (+59.7)
50	50	28.5 (-21.5)	51.7 (+1.7)
50	100	37.2 (-12.8)	58.7 (-41.3)

また Drive パラメータにおいては、設定値が 0 から 100 へと変化しているにもかかわらず、推定値はいずれも 50 から 60 の範囲に集中した。特に、設定値 0 における推定値が 59.7, 設定値 100 における推定値が 58.7 となるなど、パラメータの増減と推定値の間に正の相関が認められなかった。さらに、この推定精度の偏りを詳細に分析するため、Drive 設定 50 (Tone0) の条件において、入力された音程ごとの時系列推定結果を可視化したものを図 7 に示す。

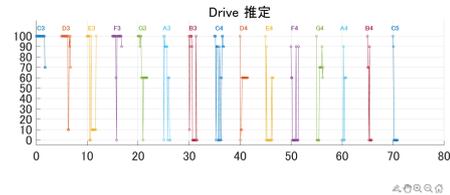


図 7: 各音程ごとの Drive の推定結果 (Drive50 Tone0)

図 7 の左側、低音域 (C3 ~ G3 付近) に着目すると、本来の設定値が 50 であるにもかかわらず、推定値は最大値である 100 に張り付いている。対照的に、図の右側へ進み音程が高いと (A4 ~ C5 付近) 推定値は低下し、最小値である 0 付近で推移している様子が確認できる。表の平均値が 50 付近であったのは、これらの極端な値が平均化された結果であることがわかる。

以上のことから、単純なりサンプリングによるデータ拡張は、Drive 推定 (非線形歪み) に対しては物理的整合性を保ち有効に機能していることが分かる。しかし、Tone 推定 (線形フィルタ) に対しては単純なりサンプリングでは Tone 回路のフィルタ特性が物理的に変質し、推定精度が悪化することが明らかとなった。

また図 8 は Drive が 50, Tone が 50 の際の各音程の推定結果である。C4 から G4 の推定結果を確認すると、20 から 40 に推定されている部分はあるものの、学習データ (E4) を中心として概ね ± 5 半音程度の音域においては、パラメータの変化傾向を他音程に比べて追従できている。よって、1 つの学習データがカバーできる有効な推定範囲は概ね ± 5 半音程度であると推測される。この有効範囲を考慮して学習データを拡張することで、効率的に全音域を網羅できる可能性が高い。

6 結論

本研究では、実機のアナログギターエフェクター (BOSS OD-3) を対象としたパラメータ推定手法を提案した。提案手法では、連続的なつまみ操作録音による効率的なデータセッ

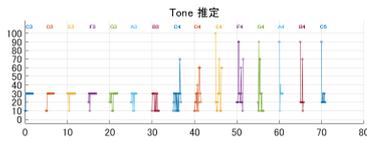


図 8: Tone の推定結果 (Drive50 Tone50)

ト構築, Wiener Filter による定常雑音の抑制, および音響特徴量としての MFCC の採用により, 実機特有の不安定な信号特性の克服を試みた. 評価実験の結果, 学習時と同一の音高 (E4) における推定精度は, 11 段階分類において平均絶対誤差 (MAE) が Drive で 3.8, Tone で 4.4 となり, 実機の物理的な設定値を実用的な精度で推定可能であることを示した. これは, 先行研究におけるデジタルシミュレーションを対象とした推定精度と同等以上であり, 出力を離散化した分類モデルが, アナログ回路の微小な変動を吸収するフィルタリングとして有効に機能したと言える.

しかし, 学習データに含まれていない, 未知の演奏条件に対する汎用性において課題が残された. これらを踏まえ, 今後の展望を以下に示す.

考察で述べた通り, 本モデルは学習データの音程に対し, 5 半音程度の範囲内であれば推定値が正しく推移する傾向も確認されている. この知見に基づき, 今後はモデルの有効範囲が重複するような 5 半音程度の間隔で選定した離散的な複数音程によるデータセットを構築する. これにより, 実機収録のコストを最小限に抑えつつ, ギターの全音域において音程に依存しない頑健なモデルの実現を目指す.

実機収録に加え, 信号処理によるデータ拡張も検討する. 考察において, 単純なりサンプリングでは Tone 回路のフィルタ特性が物理的に変質し, 推定精度が悪化することが明らかとなった. この問題を解決するため, 逆フィルタ (Inverse Filter) を用いたソース・フィルタ分離手法の導入を検討する. 具体的には, 回路の伝達関数に基づいて Tone のフィルタ特性を一度除去し, ピッチ変換後に正しいフィルタ特性を再度畳み込むことで, 物理的な整合性を保ったまま学習データを拡張する手法を検討する.

また本研究の評価実験においては, 純粋なパラメータ推定性能を検証するため, 入出力信号のピークレベルを正規化し, 音量差による影響を意図的に排除した条件下で評価を行った. しかし, 実環境におけるギター演奏では, ピッキングの強弱やギターのボリューム操作によって入力レベルが常に変動する. 予備検証の結果, ギター側のボリューム操作やピッキングの強弱による入力レベルの違いが, 歪み回路による倍音構成に変化を与えることが確認された. 図 9 は, エフェクターの設定 (Drive, Tone) を固定した状態で, ギターの入力レベルを変化させた際の周波数スペクトルを比較したものである. a 赤線がレベル最大のもの, 青線がレベル半分程度のものとなっている. 図 9 を確認すると, 入力レベルの違いにより, 特に 1kHz 以上の高周波数帯域における倍音成分の現れ方に顕著な差異が生じていることがわかる. Comunità ら [3] をはじめと

する先行研究においても, 入力信号は正規化された一定の条件を前提として議論されており, こうした入力ダイナミクスの変化に伴う非線形な音色変化への対応については, 未検討の課題として残されている. このようにレベルの低下は歪み量の変化に直結するため, 今後は入力音量の異なるデータを含めた学習やその方法の検討が必要となる.

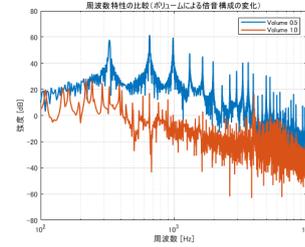


図 9: 入力音量差による歪み方の差

参考文献

- [1] Henrik Jürgens, Reemt Hinrichs, and Jörn Ostermann. Recognizing guitar effects and their parameter settings. In *Proceedings of the 23rd International Conference on Digital Audio Effects (DAFx2020)*, pp. 310–317, 2020.
- [2] Johanna Hinrichs, Giso Grimm, Matthias Kretschmer, and Christian Dittmar. Convolutional neural networks for the classification of guitar effects and extraction of the parameter settings of single and multi-guitar effects from instrument mixes. *EURASIP Journal on Audio, Speech, and Music Processing*, Vol. 2022, No. 1, pp. 1–18, 2022.
- [3] Marco Comunità, Dan Stowell, and Joshua D. Reiss. Guitar effects recognition and parameter estimation with convolutional neural networks. *Journal of the Audio Engineering Society*, Vol. 69, No. 7/8, pp. 594–604, 2021.
- [4] Alec Wright, Eero-Pekka Damskägg, and Vesa Välimäki. Real-time guitar amplifier emulation with deep learning. *Applied Sciences*, Vol. 10, No. 3, p. 766, 2020.
- [5] Eero-Pekka Damskägg, Juusopetteri Silén, and Vesa Välimäki. Automatic black-box modeling of distortion circuits with wavenet. In *Proceedings of the 22nd International Conference on Digital Audio Effects (DAFx-19)*, pp. 1–8, 2019.
- [6] Lauri Juvela, et al. End-to-end amp modeling: From data to controllable guitar amplifier models. arXiv preprint arXiv:2403.08559, 2024.
- [7] Maxime Guezin and Geoffroy Peeters. Egfxset: Electric guitar tones processed through real effects of distortion, modulation, delay and reverb. In *Proceedings of the 23rd International Society for Music Information Retrieval Conference (ISMIR)*, pp. 662–668, 2022.