理工学研究科

2026 年 3 月

修士論文

深層学習に基づく予測と探索アルゴリズムによるオーディオエフェクト推定

47024810　沖田 陽一

（人間システム工学専攻）

要旨

オーディオエフェクトは様々な信号処理技術によって音響信号を加工するツールであり，放送・映画・音楽・ゲームなどにおけるサウンドデザインに広く用いられ重要な役割を担っている．単一のエフェクトの構成は，その種類を大別したタイプと，タイプごとにその作用を調節するパラメータから決定される．多くの場合，複数のエフェクトが直列や並列に接続して適用され，その全体としての構成は各エフェクトのタイプ・パラメータおよびそれらのルーティングから決定される．接続が直列のみの場合はオーディオエフェクトチェインと呼ばれ，そのルーティングはエフェクトの順序によって決定される．

本研究では，オーディオエフェクト推定と呼ばれるタスクに取り組む．これは，エフェクト適用後の音響信号であるウェットシグナルからエフェクトの構成を推定するタスクである．従来，複雑で多様なオーディオエフェクトを駆使して所望のサウンドデザインを実現するためには，技術面と芸術面の双方で高い専門性が要求されてきた．そこで，本研究で取り組むような自動のオーディオエフェクト推定が実現されることにより，サウンドデザインの初学者から専門家までが，既存の制作物からそこに含まれるサウンドデザインの手法を効率的に学習したり再利用したりすることが可能となる．

オーディオエフェクト推定に取り組む既存の研究の多くは予測的アプローチをとってきた．ここでの予測的アプローチとは，DNN などの機械学習モデルを事前に学習させたうえで，未知のウェットシグナルに対して推論させることでエフェクト構成を予測するというものである．このような予測モデルの学習は，ドライシグナルにエフェクトを適用することで得られるウェットシグナル・エフェクト構成の教師データとの誤差を目的関数として行われる．また，ドライシグナルとエフェクト構成の同時推定という新たなタスクも導入されている．この研究では，エフェクトチェイン中で最後に適用された単一のエフェクトについて，そのエフェクトへの入力信号とエフェクト構成を予測するモデルを構築し，それをウェットシグナルに繰り返し適用することでチェイン全体についての推定を行う手法が提案された．この手法により，ウェットシグナルのみからエフェクトチェインの構成を順序も含めて推定することが可能となったが，反復的な推論による誤差の累積は課題として残されていた．

一方，予測的アプローチとは異なる探索的アプローチも考えられる．ここでの探索的アプローチとは，ドライシグナルにエフェクトを適用してウェットシグナルの再構成を行いながら，それと目標のウェットシグナルとの類似度である再構成類似度を目的関数としてエフェクト構成を最適化するというものである．これまで，エフェクトのスタイル転移などにおいてこれに近いアプローチがとられてきた．

本研究では，予測的アプローチと探索的アプローチを組み合わせた新たなアプローチを提案する．このアプローチでは，まず，DNN でドライシグナルとエフェクト構成の全体または一部の予測を行い，その後，それらを用いてウェットシグナル再構成類似度に基づく探索を行う．予測段階でドライシグナルの推定を行うことで再構成類似度を評価することが可能となり，探索段階でこれを目的関数として予測の補完や改善を行う．

提案アプローチの有効性を確認するための評価実験をギターによって演奏された音楽的抜粋とオーディオエ

フェクトチェインについて行った．また，このような二段階のアプローチに基づいて手法を設計するとき，それらの間でタスクの分担は本質的に不可欠な選択である．そこで，提案アプローチに基づく手法設計の指針を得るため，予測と探索の間のタスク分担を複数検討し，全体としての推定性能の比較を行った．その結果，提案アプローチに基づく全手法が予測のみによる手法を上回る性能を示し，提案アプローチの有効性が示された．また，タスクの分担については，予測段階においてエフェクトタイプの組み合わせのみを予測し探索段階においてその順序を推定する手法が，多くの指標において最適であることが示された．さらに，推定対象エフェクトチェインの長さごとの各性能指標の傾向も調査した．その結果，多くの指標において，長いチェインほど推定が困難であることが示された．また，推定対象がそのような長いチェインであるほど探索段階においてエフェクトタイプの順序を推定する手法が高い性能を示す傾向にあり，順序推定における探索手法の有効性が示唆された．

目次

図目次

表目次

第 1 章

序論

1.1 背景と目的

オーディオエフェクト（本稿の本文中では単にエフェクト，図表中では AFx とも書く.）は様々な信号処理技術によって音響信号を加工するツールであり [1, 2]，放送・映画・音楽・ゲームなどにおけるサウンドデザインに広く用いられ重要な役割を担っている [3]．また，エフェクトに関する研究も，構成推定・除去・スタイル転移・モデリングなどの様々な側面から取り組まれている [4]．単一のエフェクトの構成は，その種類を大別したカテゴリカルな変数であるタイプと，タイプごとにその作用を調節するパラメータから決定される．多くのエフェクトは古典的な信号処理の組み合わせからなり，タイプやパラメータは基本的に解釈可能であり，その原理や典型的な設定が知られている．また，多くの場合，複数のエフェクトが直列や並列に接続して適用され，その全体としての構成は各エフェクトのタイプ・パラメータおよびそれらのルーティングから決定される．接続が直列のみの場合はオーディオエフェクトチェインと呼ばれ，そのルーティングはエフェクトの順序によって決定される．

本研究では，オーディオエフェクト推定と呼ばれるタスクに取り組む．これは，エフェクト適用後の音響信号であるウェットシグナルから適用されたエフェクトの構成を推定するタスクである．従来，複雑で多様なオーディオエフェクトを駆使して所望のサウンドデザインを実現するためには，技術面と芸術面の双方で高い専門性が要求されてきた．そこで，本研究で取り組むような自動のオーディオエフェクト推定が実現されることにより，サウンドデザインの初学者から専門家までが，既存の制作物からそこに含まれるサウンドデザインの手法を効率的に学習したり再利用したりすることが可能となる．

オーディオエフェクト推定に取り組む既存の研究の多くは予測的アプローチをとってきた [5, 6, 7, 8, 9, 10, 11, 12, 13, 14]．ここでの予測的アプローチとは，深層ニューラルネットワーク（Deep Neural Network; DNN）などの機械学習モデルを事前に学習させたうえで，未知のウェットシグナルに対して推論させることでエフェクト構成を予測するというものである．このような予測モデルの学習は，ドライシグナルにエフェクトを適用することで得られるウェットシグナル・エフェクト構成の教師データとの誤差を目的関数として行われる．また，ドライシグナルの推定を行うオーディオエフェクト除去も，同様の DNN を用いた予測的アプローチによって取り組まれている．さらに，ドライシグナルとエフェクト構成の同時推定という新たなタスクも導入されている [15]．この研究では，エフェクトチェイン中で最後に適用された単一のエフェクトについて，そのエフェクトへの入力信号とエフェクト構成を予測するモデルを構築し，それをウェットシグナルに繰り返し適用することでチェイン全体についての推定を行う手法が提案された．この手法により，ウェットシグナルのみから任意の長さのエフェクトチェインの構成を順序も含めて推定することが可能となったが，反復的

な推論による誤差の累積は課題として残されていた．

一方，予測的アプローチとは異なる探索的アプローチも考えられる．ここでの探索的アプローチとは，ドライシグナルにエフェクトを適用してウェットシグナルの再構成を行いながら，それと目標のウェットシグナルとの類似度である再構成類似度を目的関数としてエフェクト構成を最適化するというものである．これまで，エフェクト [16, 17]，ミキシング [18] やマスタリング [19] のスタイル転移において，これに近いアプローチがとられてきた．このウェットシグナル再構成誤差を，予測的アプローチにおけるモデルの学習に用いた手法 [12] も存在するが，エフェクトは微分可能なもの [20] に限定され，また，推論時には再構成誤差は利用されない．

本研究では，予測的アプローチと探索的アプローチを組み合わせた新たなアプローチを提案する．このアプローチでは，まず，DNN でドライシグナルとエフェクト構成の全体または一部の予測を行い，その後，それらを用いてウェットシグナル再構成類似度に基づく探索を行う．予測段階でドライシグナルの推定を行うことで再構成類似度を評価することが可能となり，探索段階でこれを目的関数として予測の補完や改善を行う．熟練したミュージシャンや音響エンジニアが既存のサウンドデザインの模倣を行うときも，まず，経験的に得た知識や感覚からエフェクトを大まかに予測し，その後，ウェットシグナルの再構成を行いながらエフェクトの順序やパラメータなどを探索的に調整するのが典型的であり，提案手法はこれに近いアプローチといえる．

提案アプローチの有効性を確認するための評価実験をギターによって演奏された音楽的抜粋とオーディオエフェクトチェインについて行う．また，このような二段階のアプローチに基づいて手法を設計するとき，それらの間でタスクの分担は本質的に不可欠な選択である．そこで，提案アプローチに基づく手法設計の指針を得るため，予測と探索の間のタスク分担を複数検討し，全体としての推定性能の比較を行った．さらに，この実験において得られた推定結果の事例はウェブサイトで公開している[*1]．

1.2　本研究の貢献

本研究の主要な貢献は次のように整理できる．

- オーディオエフェクト推定において，従来の予測的アプローチと探索的アプローチを組み合わせた新たなアプローチを提案した．
- 提案アプローチに基づく手法の有効性を，予測的アプローチのみに基づくベースラインとの比較実験によって示した．
- 提案アプローチにおいて，予測段階でエフェクトタイプの組み合わせのみを予測し探索段階でその順序を推定するというタスク分担が多くの指標において最良であり，特に長いエフェクトチェインほど有利であるという傾向を評価実験から示した．

1.3　本論文の構成

本論文の構成を整理すると次のようになる．

第 1 章　本研究の背景と目的を説明し，貢献を整理した．
第 2 章　オーディオエフェクトの適用とオーディオエフェクト推定の定式化を行う．

[*1] `https://okitayouichi.github.io/afx-pred-sch-demo`

第 3 章　オーディオエフェクト推定とそれに関連したタスクについての既存研究を整理する．

第 4 章　オーディオエフェクト推定に対するアプローチとして，予測的アプローチと探索的アプローチを組み合わせた新たなものを提案する．まず，第 3 章で整理した既存手法の中で，このアプローチを位置づける．次に，このアプローチに基づく複数の提案手法を設定し，その定式化や手法の詳細な説明を行う．そして，それらの手法について行った評価実験の手法と結果を説明し，結果に対する考察を述べる．

第 5 章　本研究において残された課題とそれらについての今後の展望を述べる．

第 6 章　本研究で提案されたアプローチ，評価実験と今後の展望を総括して述べる．

第 2 章

オーディオエフェクト推定の定式化

本章では，まず，オーディオエフェクトチェインの適用の定式化を行う．そして，それに基づいて，その逆問題であり本研究で取り組む問題であるオーディオエフェクトチェイン推定の定式化を行う．なお，ここで定義する用語は，基本的にサウンドデザインや音響信号処理の分野で標準的なものから選んだが，一部はこれまで明確に呼称が定義されていない対象について独自に定義したものである．

2.1 オーディオエフェクトの適用

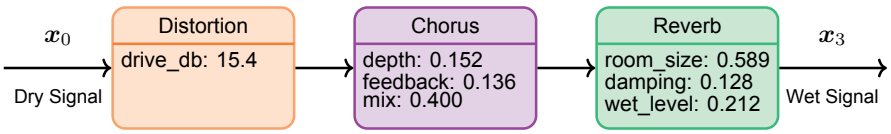


図 2.1: 長さ 3 のオーディオエフェクトチェインの適用の例．それぞれのエフェクトは上部に記載したようなタイプと下部に記載したようなパラメータによって設定される．本研究では，各パラメータの値はこの図に示したようなものから単位区間に正規化されて扱われる．

まず，音響信号への単一のオーディオエフェクトの適用は

$$\boldsymbol{x}_1 = f_{c,\boldsymbol{p}}(\boldsymbol{x}_0) \tag{2.1}$$

と表せる．ここで，$\boldsymbol{x}_0 \in \mathbb{R}^T$ はドライシグナル，$\boldsymbol{x}_1 \in \mathbb{R}^T$ はウェットシグナルである．これらはどちらもサンプル長 T のディジタル音響信号であり，本研究ではモノラルのものを考える．$f_{c,\boldsymbol{p}} \colon \mathbb{R}^T \to \mathbb{R}^T$ はオーディオエフェクト，$c \in \mathcal{C}$ はそのタイプ，$\boldsymbol{p} \in [0,1]^{d_c}$ は c に対応したパラメータである．タイプ c はカテゴリカルな変数であり $\mathcal{C} = \{\mathrm{Chorus}, \mathrm{Distortion}, \mathrm{Reverb}, \ldots\}$ のような集合の要素である．パラメータ $\boldsymbol{p}$ はその次元 d_c がタイプ c によって異なり，また，各成分は単位区間 $[0,1]$ に正規化されている連続変数であるとする．また，単一のエフェクトの構成とはタイプとパラメータの順序組 $(c, \boldsymbol{p})$ を指すとする．

次に，オーディオエフェクトチェインを考える．オーディオエフェクトチェインは，図 2.1 のように，複数のエフェクトが直列に接続されたものである．チェイン中の n 番目のエフェクトの適用は $\boldsymbol{x}_n = f_{c_n,\boldsymbol{p}_n}(\boldsymbol{x}_{n-1})$ と表せて，このとき長さ N のチェイン全体の適用は

$$\boldsymbol{x}_N = f_{c_N,\boldsymbol{p}_N} \circ \cdots \circ f_{c_2,\boldsymbol{p}_2} \circ f_{c_1,\boldsymbol{p}_1}(\boldsymbol{x}_0) = F_{C,P}(\boldsymbol{x}_0) \tag{2.2}$$

と表せる．ここでは，ドライシグナルは $\boldsymbol{x}_0$，ウェットシグナルは $\boldsymbol{x}_N$ である．タイプとパラメータの順序付きの列とそれらによって設定されたチェイン全体を表す記号，それぞれ，$C = (c_1, \ldots, c_N), P = (\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N), F_{C,P}$ を導入した．また，チェイン中のある段階のウェットシグナル $\boldsymbol{x}_n$ に対して，最後に適用された単一のエフェクトの適用前の音響信号 $\boldsymbol{x}_{n-1}$ をバイパスシグナルと呼ぶ．

2.2　オーディオエフェクトの推定

本研究で取り組むオーディオエフェクトチェイン推定は

$$\left(\hat{C}, \hat{P}, \hat{\boldsymbol{x}}_0\right) = E\left(\boldsymbol{x}_N\right) \tag{2.3}$$

と表せる．ここで，E は推定器であり，この処理はウェットシグナル $\boldsymbol{x}_N$ からエフェクト構成 $\left(\hat{C}, \hat{P}\right)$ とドライシグナル $\hat{\boldsymbol{x}}_0$ の推定を行うというものである．エフェクトチェインの構成は各エフェクトの構成 $(c_n, \boldsymbol{p}_n)$ の順序付きの列である．エフェクト除去もあわせて行いドライシグナルの推定を行う．$\left(\hat{C}, \hat{P}\right)$ の推定にあたってチェインの長さ $\hat{N}$ の推定も行われる．本稿では，なんらかの変数 a があるとき，その推定値を $\hat{a}$ と表すこととする．

第 3 章

関連研究

本章では，本研究で取り組むオーディオエフェクト推定とそれと関連したタスクについての既存研究を整理する．タスクごとに既存研究を挙げ，手法や適用可能範囲からそれらを分類し，また，課題を整理する．

3.1 オーディオエフェクト推定

本研究で取り組むオーディオエフェクト推定（Audio Effect Estimation）は，ウェットシグナルから適用されたエフェクトの構成の全体または一部を推定するタスクである．このタスクに取り組む既存の研究の多くは予測的アプローチをとってきた．ここでの予測的アプローチとは，DNN などの機械学習モデルを事前に学習させたうえで，未知のウェットシグナルに対して推論させることでエフェクト構成を予測するというものである．このような予測モデルの学習には，ドライシグナルの集合にエフェクトを適用することで得られるウェットシグナルとエフェクト構成の組からなる教師データの集合が用いられる．このデータセットからサンプリングしたウェットシグナルをモデルに入力し，その出力であるエフェクト構成の推定値とデータセットから得られるエフェクト構成の真値との誤差を損失関数としてモデルを学習させるのである．

オーディオエフェクト推定の研究の初期においては，DNN ではなくサポートベクタマシン [21, 22, 23] やランダムフォレスト [24] といった古典的な機械学習モデルが分類や回帰に用いられていた．このような古典的なモデルを用いる場合，その限られた特徴抽出性能により，入力する特徴量を人手で設計し選択する必要があった．そこで，近年では，高度な特徴抽出能力を自動的に獲得可能な DNN が特徴抽出から分類や回帰にまで用いられるようになっている．

このように，予測的アプローチによって，単一のエフェクト [21, 23, 24, 5, 6, 7, 8] や複数のエフェクト [22, 9, 10, 11, 12, 13, 14] が適用された音響信号からのオーディオエフェクト推定が取り組まれてきた．しかし，複数のエフェクトの推定に関しては，順序を考慮しない多ラベル分類としてのタイプ分類 [22, 10, 13]，タイプの組み合わせとルーティングが既知である状況下でのパラメータ回帰 [9, 12] のように，ほとんどが限定的な条件下での手法であった．このような制約なしに複数のエフェクトの構成の全体を推定する手法の一つとして，エフェクトのグラフ構造を Transformer[25] で扱うものが提案されている [11].

一方，予測的アプローチとは異なる探索的アプローチも考えられる．ここでの探索的アプローチとは，ウェットシグナルの再構成に基づく類似度または誤差を目的関数として，エフェクト構成を推定時に動的に最適化するというものである．このウェットシグナル再構成類似度は，ドライシグナルに推定された構成のエフェクトを適用してウェットシグナルの再構成を行い，それと目標のウェットシグナルとの類似度を評価したものである．再構成類似度の評価にはエフェクト構成の真値は不要である一方でドライシグナルが必要であ

る．探索的アプローチによって，エフェクト [26, 27] やミキシング [28, 29] の推定が取り組まれてきたが，これらの手法では対応するドライシグナルが必要である．なお，このようなドライシグナルとウェットシグナルの両方が与えられる問題設定に対して，ウェットシグナルのみからのエフェクト推定は特にブラインドオーディオエフェクト推定と呼ばれることもある．

ウェットシグナル再構成誤差を予測的アプローチにおけるモデルの学習に用いた手法 [12] も存在するが，深層学習のアルゴリズムの要請によりエフェクトは微分可能なもの [20] に限定され，また，ドライシグナルが利用できない推論時には再構成誤差は評価されない．

3.2　オーディオエフェクト除去

オーディオエフェクト推定に関連したタスクにオーディオエフェクト除去（Audio Effect Removal）がある．これはウェットシグナルからドライシグナルを推定するタスクである．

このタスクにおいても既存の研究の多くは予測的アプローチをとってきた．ここでの予測的アプローチとは，ドライシグナルとウェットシグナルのペアデータで学習を行った機械学習モデルを，未知のウェットシグナルに対して推論させることでドライシグナルを予測するというものである．このアプローチによって，単一のエフェクト [30, 31, 32] やエフェクトチェイン [33] の除去が取り組まれてきた．

また，ウェットシグナルのみからのエフェクトのシステム同定のためにエフェクト除去を行う研究 [34] も存在する．

3.3　オーディオエフェクト構成とドライシグナルの同時推定

オーディオエフェクト構成とドライシグナルの同時推定という新たなタスクも導入されている [15]．この研究では，エフェクトチェイン中で最後に適用された単一のエフェクトについて，その構成とバイパスシグナルを予測する DNN が構築された．そして，このモデルの出力であるバイパスシグナルの推定値を再びそのモデル自身に入力するという過程を繰り返して予測を行う．このような反復適用をウェットシグナルから開始することで，チェイン全体の構成とドライシグナルを推定する．この手法により，ウェットシグナルのみから未知の長さのエフェクトチェインの構成を順序も含めて推定することが可能となったが，反復的な推論による誤差の累積は課題として残されていた．

3.4　オーディオ制作スタイル転移

オーディオ制作スタイル転移（Audio Production Style Transfer）と呼ばれるタスクも存在する．オーディオ制作スタイル転移とは，参照となるウェットシグナルとそれとは異なるコンテンツを持つドライシグナルが与えられたもとで，ドライシグナルが参照と近いスタイルを持つようなエフェクト・ミキシング・マスタリングなどの構成を推定するタスクである．ここでのコンテンツとはエフェクトなどによって変化しない旋律や和声といった楽譜的な内容，スタイルとはエフェクトなどによって決定される特徴を指すものである．

オーディオ制作スタイル転移では，ドライシグナルの入力が本質的に不可欠であり，このことを利用した探索的アプローチがとられることが多い．スタイル転移における探索的アプローチとは，参照ウェットシグナルと再構成ウェットシグナルのそれぞれからスタイルにあたる特徴を抽出するモデルを構築し，それらのスタイル間の距離や類似度を目的関数としてエフェクトなどの構成の最適化を行うものである．このアプローチに

よって，エフェクト [16, 17]，ミキシング [18] やマスタリング [19] のスタイル転移が取り組まれてきた．

なお，このようなスタイル転移においてスタイルとは何であるかの定義には文献によってぶれがある．エフェクトスタイル転移の場合，スタイルとはエフェクト構成そのものであり，エフェクト構成が等しければスタイルが一致すると考える立場 [17] がある．この立場では，ドライシグナルが不要なエフェクト推定の方がスタイル転移よりも少ない入力で同等の目的を達成可能であり，その意味ではより効率的な枠組みだといえる．一方，スタイルとはエフェクト構成とは異なる抽象的な特徴であり，参照のエフェクト構成と理想的に転移されたエフェクト構成は近いが必ずしも等しくはないと考える立場 [16] がある．この立場では，スタイル転移には参照ウェットシグナルとドライシグナルの入力の両方が必要であり，ウェットシグナルのみからのエフェクト推定では同等の目的は達成不可能であるといえる．

第 4 章

深層学習モデルに基づく予測と探索アルゴリズムによるオーディオエフェクト推定

4.1 提案アプローチの位置づけ

ここでは，第 3 章で整理した既存手法の中で，本研究で提案するアプローチを位置づける．本研究では，オーディオエフェクト推定で典型的にとられてきた予測的アプローチとオーディオ制作スタイル転移で典型的にとられてきた探索的アプローチを組み合わせた新たなアプローチを提案する．このアプローチでは，まず，DNN でドライシグナルとエフェクト構成の全体または一部の予測を行い，その後，それらを用いてウェットシグナル再構成類似度に基づく探索を行う．

前段の予測段階は，オーディオエフェクト構成とドライシグナルの同時推定を行うもので，同様のタスクに取り組んだ先行研究 [15] の手法をベースとする．提案アプローチでは，予測段階のタスクの一部を探索段階に担わせたり探索段階で予測の改善を行ったりすることで，全体としてより精緻な推定を行うことを目指す．

提案アプローチは，ウェットシグナルのみからオーディオエフェクトチェインの構成の全体とドライシグナルを推定可能なものであり，理論上の適用可能範囲は多くの既存手法よりも広い．適用可能範囲が提案アプローチと同等な手法としては提案手法の予測段階のベースとした手法 [15] が挙げられる．そして，提案アプローチよりも適用可能範囲の観点で優位にある手法は，並列接続を含むグラフ構造を推定可能な先行研究 [11] のものである．これと比較すると，提案アプローチには，予測と探索を組み合わせた新たなアプローチを導入した点に新規性があり，また，エフェクト除去も同時に行う点で優位性がある．

ウェットシグナル再構成類似度に基づく探索的アプローチは，一部のエフェクト推定 [26, 27] や多くのエフェクトスタイル転移 [16, 17] でとられてきた．しかし，これらの手法は，スタイル転移においては本質的に不可欠であるものの，推定時にドライシグナルを与える必要があった．これらに対し，提案アプローチは，ウェットシグナルのみが与えられた状況下において，予測段階でドライシグナルの推定を行うことで再構成類似度の評価を行う点で新規性がある．

4.2 推定手法

本研究では，式 (2.3) で表されるタスクを，図 4.1 のように，DNN による予測とウェットシグナル再構成に基づく探索の二段階で解決するアプローチを提案する．そして，このような二段階のアプローチに基づいて手法を設計するとき，それらの間でタスクの分担は本質的に不可欠な選択である．よって，全体として解決しよ

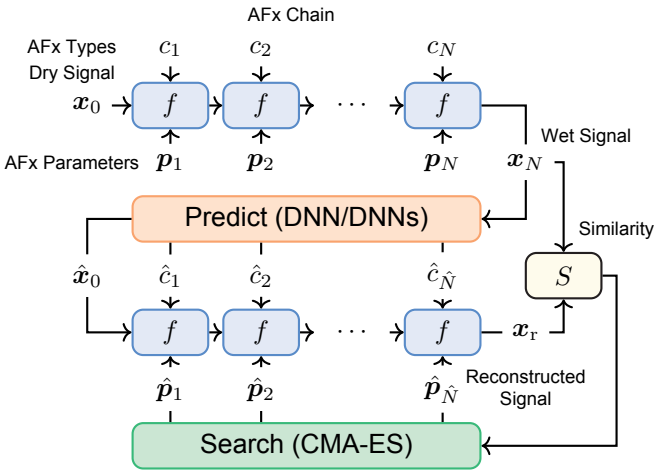


図 4.1: 深層学習モデルに基づく予測と探索アルゴリズムによるオーディオエフェクト推定のアプローチ．まず，DNN でドライシグナルとエフェクト構成の全体または一部の予測を行い，その後，それらを用いてウェットシグナル再構成類似度に基づく探索を行う．この図は，提案するアプローチに基づく手法のうち，予測段階でエフェクトタイプの列を推定し，探索段階でエフェクトパラメータを推定するような手法を表している．その他に，予測段階ではタイプの組み合わせのみを推定する手法や，予測段階でパラメータまで暫定的に推定する手法も検討する．

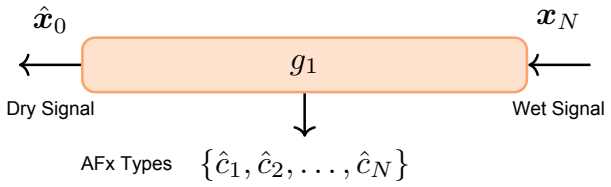
うとするのは全て同一のこのタスクだが，予測段階と探索段階でのタスクの分担が異なる手法を複数検討し，比較を行った．

4.2.1　深層学習モデルによる予測

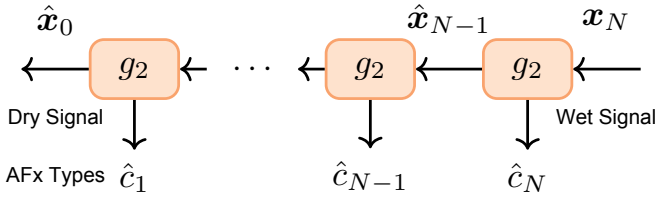
問題設定と予測手法

前段の予測段階が担うタスクとして以下の 3 つを考える．いずれもドライシグナルとエフェクトタイプの順序無しの組み合わせの予測は行う点は共通しており，それに加えてのエフェクトタイプの順序やエフェクトパラメータの予測の有無に差異がある．残ったタスクは探索段階が担うものとした．

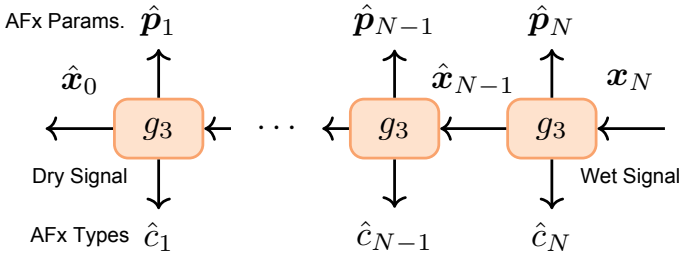
- **Dry-Type-Direct**
  DNN 予測器 g_1 は，図 4.2a のように，チェイン全体に対して，
  $$\left(\left\{\hat{c}_1, \ldots, \hat{c}_{\hat{N}}\right\}, \hat{\boldsymbol{x}}_0\right) = g_1\left(\boldsymbol{x}_N\right) \tag{4.1}$$
  のように，ドライシグナル $\hat{\boldsymbol{x}}_0$ とエフェクトタイプの順序無しの組み合わせ $\left\{\hat{c}_1, \ldots, \hat{c}_{\hat{N}}\right\}$ を一度に予測する．タイプの予測は多ラベル分類問題にあたる．このタスクにおいては，空チェイン，すなわち，何もエフェクトが適用されていない場合は推定対象としない．
- **Bypass-Type-Iter**
  DNN 予測器 g_2 は，図 4.2b のように，チェイン中で最後に適用された単一のエフェクトについて，
  $$\left(\hat{c}_n, \hat{\boldsymbol{x}}_{n-1}\right) = g_2\left(\boldsymbol{x}_n\right) \tag{4.2}$$
  のように，エフェクトタイプ $\hat{c}_n$ とバイパスシグナル $\hat{\boldsymbol{x}}_{n-1}$ を予測する．タイプの予測は単一ラベル分



(a) Dry-Type-Direct の予測段階．ドライシグナルとエフェクトタイプの組み合わせを一度に予測する．



(b) Bypass-Type-Iter の予測段階．バイパスシグナルとエフェクトタイプの予測を反復的に行うことで，音響信号とエフェクトタイプの列を予測する．



(c) Bypass-Config-Iter の予測段階．バイパスシグナルとエフェクト構成の予測を反復的に行うことで，音響信号とエフェクト構成の列を予測する．

図 4.2: オーディオエフェクト推定の予測段階

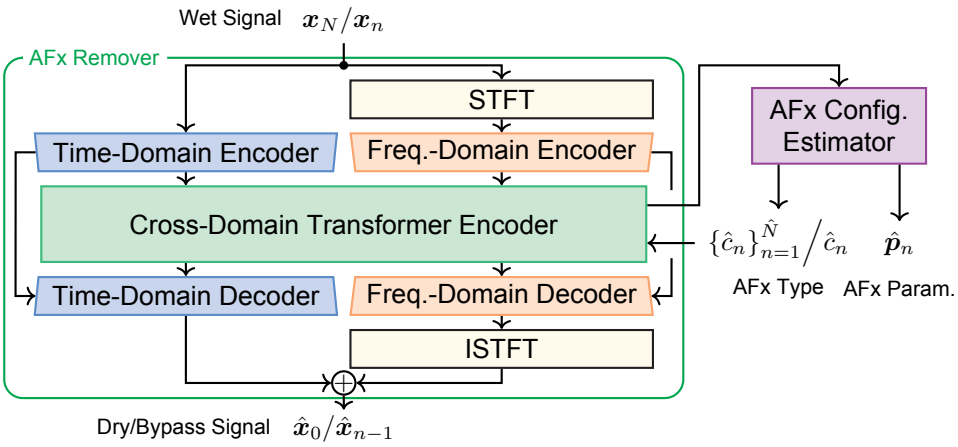
類問題にあたる．このタスクでは，空チェインも推定対象とし，その場合は特殊クラス “None” として予測する．この予測器は，その出力 $\hat{\boldsymbol{x}}_{n-1}$ を再び g_2 自身に入力するという過程を，$\hat{c}_n = \text{None}$ となるまで繰り返して予測を行う．このような反復適用をウェットシグナル $\boldsymbol{x}_N$ から開始することで，音響信号の列 $\left(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{\hat{N}-1}\right)$ とタイプの列 $\hat{C}$ を予測する．

- **Bypass-Config-Iter**

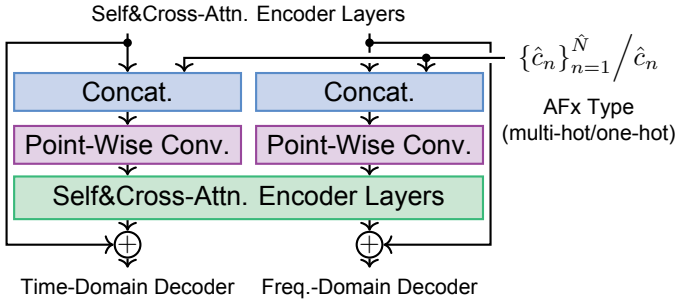
  DNN 予測器 g_3 は，図 4.2c のように，チェイン中で最後に適用された単一のエフェクトについて，

$$(\hat{c}_n, \hat{\boldsymbol{p}}_n, \hat{\boldsymbol{x}}_{n-1}) = g_3(\boldsymbol{x}_n) \tag{4.3}$$

  のように，エフェクト構成 $(\hat{c}_n, \hat{\boldsymbol{p}}_n)$ とバイパスシグナル $\hat{\boldsymbol{x}}_{n-1}$ を予測する．タイプの予測は Bypass-Type-Iter と同様に，単一ラベル分類問題にあたり，空チェインの特殊クラスも予測する．Bypass-Type-Iter と同様の反復適用をウェットシグナル $\boldsymbol{x}_N$ から開始することで，音響信号の列 $\left(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{\hat{N}-1}\right)$ とエフェクト構成の列 $\left(\hat{C}, \hat{P}\right)$ を予測する．このタスクは SunAFXiNet[15] が取り組んだものとほぼ同



(a) 予測モデルの全体図．時間・周波数各ドメインの信号をそれぞれ U-Net ベースのネットワークで処理するエフェクト除去器と，そのボトルネック部のクロスドメインエンコーダから分岐したエフェクト構成推定器からなる．



(b) クロスドメインエンコーダ内におけるエフェクトタイプによる条件付け．これによってエフェクトタイプに応じたエフェクト除去が行われることが期待される．

図 4.3: オーディオエフェクト推定の予測モデルのアーキテクチャ．

等のものである．

以下では，Bypass-Type-Iter と Bypass-Config-Iter を合わせて **Bypass-*-Iter** と書くこともある．

ネットワークアーキテクチャ

予測を行うモデル g_1, g_2, g_3 は概ね同様のアーキテクチャの DNN とした．このアーキテクチャは SunAFXiNet[15] に倣ったものであり，多くのハイパーパラメータもこのモデルと同様のものを採用した．モデルはいずれも，図 4.3a のように，時間・周波数各ドメインの信号をそれぞれ U-Net ベースのネットワークで処理するエフェクト除去器と，そのボトルネック部のクロスドメインエンコーダから分岐したエフェクト構成推定器からなる．エフェクト除去器は，先行研究 [34, 15] と同じく，音源分離で高い性能を発揮した Hybrid Transformer Demucs（HTDemucs）[35] をベースとしたものである．具体的には次のような処理を行う．

- **時間ドメインエンコーダ**
  モデルへの入力はモノラル音響信号であるウェットシグナルである．この時間ドメインの信号について，時間方向の畳み込み（カーネルサイズ 8，ストライド 4）などの処理を行う．各層は HTDemucs と同様の構造を持ち，層の数は 6 とした．最初の層の出力のチャネル数は 32 とし，後続の層では入力のチャネル数の 2 倍とした．
- **周波数ドメインエンコーダ**
  各層は HTDemucs と同様の構造を持つ．まず，時間ドメインの入力信号は，短時間フーリエ変換（Short-Time Fourier Transform; STFT）によって，周波数ドメインに変換される．このとき，HTDemucs と同様に，FFT サイズは $2F = 8192$，ホップ長は $H = F/2$ とし，また，出力の実部と虚部をチャネル方向に展開する．その後，この周波数ドメインの信号について，周波数方向の畳み込み（カーネルサイズ 8，ストライド 4）などの処理を行う．各層は HTDemucs と同様の構造を持ち，層の数は 6 とした．この構成により，周波数ドメインの信号の周波数方向の長さは，STFT の直後は F であり，全エンコーダの処理後は $F/4^6 = 1$ となる．
- **クロスドメインエンコーダ**
  HTDemucs と同様の Transformer[25] ベースの構造を持つ．まず，前処理として，各ドメインの信号に正規化と位置符号化が行われる．周波数ドメインの信号はエンコーダによって周波数方向の長さが 1 になっているため，そのまま系列として扱うことができる．その後，自己注意・交差注意エンコーダ層が 5 層にわたって交互に処理を行う．その第 3 層の後からエフェクト構成推定器が分岐し，第 4 層の前にその出力のひとつであるエフェクトタイプによる条件付けが行われる．
- **エフェクトパラメータ推定器**
  クロスドメインエンコーダの第 3 層が出力する時間・周波数両ドメインの信号を入力とし，エフェクトタイプとエフェクトパラメータの推定値を出力する．まず，各ドメインの信号をそれぞれ 3 層の独立した畳み込みブロックによって処理する．畳み込みブロックは，時間方向の 1 次元畳み込み層（カーネルサイズ 4，ストライド 2）・バッチ正規化・ReLU からなる．出力のチャネル数は入力のチャネル数と等しいとした．畳み込みブロックの次は，各ドメインで時間方向の平均値と最大値を加算する大域プーリングを行い，それらをチャネル方向に結合する．そして，その信号を 3 層のタイプ分類用の全結合ブロックによって処理する．最終ブロック以外の全結合ブロックは，全結合層・バッチ正規化・ReLU・ドロップアウト（確率 0.2）からなる．また出力のチャネル数は，入力のチャネル数の 0.5 倍とした．最終ブロックは全結合層と手法によって異なる活性化関数からなる．Dry-Type-Direct では sigmoid 関数を適用し，この出力を閾値 0.5 で multi-hot 表現に変換する．Bypass-*-Iter では softmax 関数を適用し，argmax 演算によって one-hot 表現に変換する．また，Bypass-Config-Iter では，これに加え，タイプ分類用の全結合ブロックの開始箇所から，パラメータ回帰用の全結合ブロックの分岐を設ける．この分岐の最終ブロック以外の全結合ブロックは，全結合層・バッチ正規化・ReLU・ドロップアウト（確率 0.05）からなる．層の数とチャネル数はタイプ分類用と同様とした．最終ブロックは全結合層のみからなり，タイプ分類結果に関係なく想定する全タイプのパラメータの推定値を出力する．これにより，タイプ分類結果の正誤に関係なく，パラメータ回帰の学習を行うことが可能となる．さらに，タイプの推定値は，クロスドメインエンコーダの第 4 層への条件付けに用いられる．これは，図 4.3b のように，まず，エフェクトタイプの表現を全時刻に複製し，次に，各ドメインでチャネル方向に結合し，その後，チャネル方向の点単位畳み込みにより各ドメインのチャネル数を元のチャネル数に戻すことによって実現される．

- **時間ドメインデコーダ**
  エンコーダからのスキップ接続をもち，時間方向の転置畳み込み（カーネルサイズ 8，ストライド 4）などの処理を行う．各層は HTDemucs と同様の構造を持ち，層の数とチャネル数は時間ドメインエンコーダと対称である．
- **周波数ドメインデコーダ**
  エンコーダからのスキップ接続をもち，周波数方向の転置畳み込み（カーネルサイズ 8，ストライド 4）などの処理を行う．各層は HTDemucs と同様の構造を持ち，層の数とチャネル数は周波数ドメインエンコーダと対称である．最後のデコーダの後は，逆短時間フーリエ変換（Inverse Short-Time Fourier Transform; ISTFT）を行い，時間ドメインの信号と同一の形状としたのちに加算し，最終的なドライシグナルまたはバイパスシグナルの推定値を出力する．

学習処理

モデルの学習も先行研究 [15] に倣った二段階で行う．

第一段階では，エフェクト除去器のみの学習を行う．損失関数として，Dry-Type-Direct では

$$L_{11} = L_{\mathrm{mae}}\left(\hat{\boldsymbol{x}}_0, \boldsymbol{x}_0\right) + \alpha L_{\mathrm{mrstft}}\left(\hat{\boldsymbol{x}}_0, \boldsymbol{x}_0\right) \tag{4.4}$$

を，Bypass-*-Iter では，共通して，

$$L_{12} = L_{\mathrm{mae}}\left(\hat{\boldsymbol{x}}_{n-1}, \boldsymbol{x}_{n-1}\right) + \alpha L_{\mathrm{mrstft}}\left(\hat{\boldsymbol{x}}_{n-1}, \boldsymbol{x}_{n-1}\right) \tag{4.5}$$

用いる．ここで, $L_{\mathrm{mae}}, L_{\mathrm{mrstft}}$ は, それぞれ, 平均絶対誤差（Mean Absolute Error; MAE）と Multi-Resolution STFT loss（MR-STFT）[36] である．Bypass-*-Iter の学習において，推定対象が空チェインである場合は，入力信号自身を真のバイパスシグナルとする．また，条件付けのためのエフェクトタイプには真値を用いる．Bypass-*-Iter の両モデルのエフェクト除去器はそのパラメータも含め同一のものである．

第二段階では，エフェクト除去器のパラメータを固定し，エフェクト構成推定器のみの学習を行う．損失関数として，Dry-Type-Direct・Bypass-Type-Iter・Bypass-Config-Iter それぞれで，

$$L_{21} = L_{\mathrm{bce}}\left(\left\{\hat{c}_1, \ldots, \hat{c}_{\hat{N}}\right\}, \left\{c_1, \ldots, c_{\hat{N}}\right\}\right) \tag{4.6}$$

$$L_{22} = L_{\mathrm{ce}}\left(\hat{c}_n, c_n\right) \tag{4.7}$$

$$L_{23} = L_{\mathrm{ce}}\left(\hat{c}_n, c_n\right) + L_{\mathrm{mse}}\left(\hat{\boldsymbol{p}}_n, \boldsymbol{p}_n\right) \tag{4.8}$$

を用いる．ここで，$L_{\mathrm{bce}}, L_{\mathrm{ce}}, L_{\mathrm{mse}}$ は，それぞれ，二値交差エントロピー・交差エントロピー・平均二乗誤差である．なお，便宜上 $L_{\mathrm{bce}}, L_{\mathrm{ce}}$ は $\hat{c}_n$ などについて計算するような表記を行ったが，正確にはモデルの出力する確率分布について計算する．また，パラメータ回帰の学習時は，真のタイプに対応したパラメータのみを用い，他の出力からは誤差の逆伝播を行わない．

4.2.2　ウェットシグナル再構成に基づく探索

問題設定

探索段階では，基本的に，エフェクトパラメータに関して，再構成ウェットシグナル $\boldsymbol{x}_{\mathrm{r}} = F_{\hat{C},P}\left(\hat{\boldsymbol{x}}_0\right)$ と元のウェットシグナルの類似度を最大化する最適化問題

$$\hat{P} = \arg\max_{P} S\left(F_{\hat{C},P}\left(\hat{\boldsymbol{x}}_0\right), \boldsymbol{x}_N\right) \tag{4.9}$$

を解く．ここで，S は音響信号間の類似度であり，本研究では Scale-Invariant Signal-to-Distortion Ratio（SI-SDR）[37] を用いる．

ドライシグナル $\hat{\boldsymbol{x}}_0$ は予測段階で得た推定値を用いる．予測段階でエフェクトタイプ列 $\hat{C}$ の順序を推定しない Dry-Type-Direct では，この探索を二段階で行う．まず，第一段階では，予測されたエフェクトタイプの組み合わせからなる全順列についてこの探索を行い，類似度が最良のものを推定エフェクトタイプ列する．その後，そのエフェクトタイプ列について追加で探索を行い，さらなる精緻化を目指す．この際，第一段階で得た $\hat{P}$ を第二段階における初期解 P_0 として用いる．また，Bypass-Config-Iter でも，予測段階で得た $\hat{P}$ を探索の初期解 P_0 として用いる．

探索アルゴリズム

式 (4.9) の目的関数 S の中に現れるエフェクトチェイン $F_{\hat{C},P}$ は，多くの場合，その実装が不明であったり，微分不可能であったりする．よって，この問題にはブラックボックス最適化手法を適用する必要がある．最適化アルゴリズムとしては，基本的に，スタイル転移で同様の再構成類似度最適化を行う先行研究 [16] に倣い，共分散行列適応進化戦略（Covariance Matrix Adaptation Evolution Strategy; CMA-ES）[38] を用いた．探索パラメータ数が 1 であるときはその特性上 CMA-ES は効果的ではないため，代わりに Tree-structured Parzen Estimator（TPE）[39] を用いた．

探索の総試行回数を，本研究では，

$$M = \lfloor M_0 d^r \rfloor \tag{4.10}$$

の形式で設定する．複数の個体からなる世代の単位で最適化を行う CMA-ES については，M は全世代にわたる個体数の和である．M_0 と r は定数のパラメータである．d は探索次元，すなわち，推定されたタイプ列 $\hat{C}$ 内の各タイプに対応するパラメータ数の総和 $d = \sum_{\hat{c}\in\hat{C}} d_{\hat{c}}$ である．この問題の探索空間は $[0,1]^d$ である．よって，単純な線形探索で d に依存しない一定の精度を期待するためには，M は d の指数関数とするのが妥当である．しかし，指数関数は d が大きいときに時間計算量が膨大になることと，CMA-ES や TPE は線形探索よりも効率的なアルゴリズムであることを踏まえ，このように指数関数より増加が緩やかなべき乗関数に基づく形式を採用した．

4.3　評価実験

提案アプローチの有効性を確認し，また，それに基づく手法設計の指針を得るため，第 4.2 節で提案した手法群の評価実験を行った．本研究では，特にオーディオエフェクトが多用される領域の一つであるギター演奏におけるサウンドデザインを対象として評価を行った．

4.3.1　データセット

まず，既存のデータセットからドライシグナルを収集した．具体的には，IDMT-SMT-Guitar[40] の `dataset4`，GuitarSet[41] の `audio_mono-pickup_mix`，EGDB[42] の `audio_DI` と Guitar-TECHS[43] の `P3_music/audio/directinput` に含まれる音響信号を利用した．これらは，アコースティックギターまたはエレクトリックギターによる音楽的抜粋であり，エフェクトを適用していないものである．ここでの音楽的抜粋とは，単音や単発の和音ではなく，旋律や伴奏といった楽譜的内容がある程度複雑なものを指す．これらから 10.0 s のチャンクを合計 2232 件重複なく切り出した．この際，IDMT-SMT-Guitar 内の音響信号の冒頭に

表 4.1: データセット作成に用いた pedalboard のオーディオエフェクト.

Type	Variable Parameter	Range
Chorus	depth	0.1–0.3
	feedback	0.0–0.5
	mix	0.3–0.7
Distortion	drive_db	10.0–20.0
Reverb	room_size	0.1–0.7
	damping	0.1–0.9
	wet_level	0.1–0.4

含まれる無音区間とカウントインの区間を除去した．そして，得られたチャンク対する前処理として，チャネル数は 1，サンプリング周波数は 44.1 kHz に統一し，二乗平均平方根（Root Mean Square; RMS）を 0.1 とする音量正規化を適用した．

次に，これらにエフェクトを適用しウェットシグナルのデータセットを作成した．エフェクトにはライブラリ pedalboard[*1]を用いた．実用的なサウンドデザインと推定可能性を考え，表 4.1 に示すタイプ，パラメータとその範囲でエフェクトを選択した．これらの 3 タイプから各タイプが 0–1 回現れる全順列として，エフェクトチェイン $\sum_{n=1}^{3} {}_3P_n = 15$ 種類を構成した．そして，その中間の信号も含め，単一のドライシグナルあたり $\sum_{n=1}^{3} n_3P_n = 33$ 個のウェットシグナルを作成した．パラメータは表の範囲で一様分布に従ってランダムに決定し，記載のないものはライブラリのデフォルト値を用いた．学習と評価の際は，パラメータの値は表の範囲を $[0, 1]$ に正規化して用いた．また，エフェクトの適用によって音量も変化するが，音量変化よりもそれ以外のエフェクトの中心的な特性に注目するため，チェイン中の各エフェクトの適用後にも RMS を 0.1 とする音量正規化を適用した．さらに，ファイルを介した音響信号の扱いの再現性を確保するため，範囲 $[-1.0, 1.0]$ の外の値はクリッピングした．

これにより，データの総エントリ数は $2232 \times 33 = 73656$，総時間は $10.0\,\mathrm{s} \times 73656 = 204\,\mathrm{h}$ となった．空チェインについても予測を行う Bypass-*-Iter の学習と評価の際には，これに加えて空チェインもデータエントリに含めた．このとき，空チェインのエントリ数は，他のエフェクトタイプが最後に適用されたエントリ数と等しく，単一のドライシグナルあたり $33/3 = 11$ とした．これにより 11 件の同一のデータが含まれることになるが，タイプ分類問題の学習と評価にあたってクラス数の不均衡を避けるために，このような構成とした．

4.3.2　実験設定

データセットを 80%,15%,5% に，ドライシグナルが重複しないように分割し，それぞれ学習・検証・評価用とした．

式 (4.4) と式 (4.5) の MR-STFT の実装はライブラリ auraloss[44] のものを用い，その重みは $\alpha = 0.01$ とした．モデルの学習における最適化アルゴリズムは AdamW[45]，学習率は第一・第二段階でそれぞれ $1 \times 10^{-4}, 1 \times 10^{-5}$ とし，weight decay は 1×10^{-2} とし，勾配の L^2 ノルムを閾値 5.0 でクリッピングし

[*1] https://github.com/spotify/pedalboard

表 4.2: 単一のオーディオエフェクトについてのタイプ分類の性能評価．予測段階において反復推論を行う手法について，チェイン中の最後に適用された単一のエフェクトのタイプ分類の評価を行った．

Method	Macro F_1 (↑)
Bypass-Type-Iter	**0.919**
Bypass-Config-Iter	0.917

た．バッチサイズは 64 とし，エポック数は第一・第二段階でそれぞれ 170, 50 とした．検証データ上での評価指標としては，第一段階のエフェクト除去器の学習においては，ドライシグナルまたはバイパスシグナルの SI-SDR を用いた．第二段階のエフェクト構成推定器の学習においては，タイプの Macro F_1 を用いた．これらの検証データ上での評価指標の値が最良の状態のモデルを評価に用いた．

Bypass-Config-Iter のエフェクトパラメータ回帰においては，表 4.1 で定めた範囲の外の値が推定されることが有り得るが，学習時はこれをそのまま扱った．一方，後段の探索の初期解として用いる際や，ベースラインの評価時は，範囲外の値は表の範囲にクリッピングした．

探索アルゴリズムの実装はライブラリ Optuna[46] のものを用い，ここに特記しない設定はそのデフォルトに従った．式 (4.10) の総試行回数 M のパラメータについては，小規模な予備実験から，Dry-Type-Direct の第一段階では $M_0 = 5$，Dry-Type-Direct の第二段階と Bypass-*-Iter では $M_0 = 20$，そして，全手法で $r = 1.5$ とした．

CMA-ES の個体群サイズは文献 [47] で推奨されている $\lambda = 4 + \lfloor 3 \log d \rfloor$ とした．世代数は $\lfloor M/\lambda \rfloor$ となる．初期解 P_0 が与えられるときは，アルゴリズムとは独立に最初に P_0 を評価し，その後，初期分布の平均を P_0 と設定した．一方，P_0 が与えられないときは，初期分布の平均は探索空間の中心の値とした．

TPE において P_0 が与えられるときは，それをそのアルゴリズムの中で最初に必ず P_0 を評価する．一方，P_0 が与えられないときは，初期試行は一様分布に従ってランダムに選択される．

空チェインのエントリは Bypass-*-Iter の反復推論における停止条件の学習のみを目的として学習に用いた．よって，エフェクトチェイン除去（表 4.5）とウェットシグナル再構成（表 4.6）の評価においては，過大評価などを避けるため，真の空チェインおよび誤って空チェイン識別されたエントリはどちらも除外した．また，Bypass-*-Iter の反復推論の停止条件は，モデルが特殊タイプ “None” を予測するかチェインの長さが $\hat{N} = 3$ になることとした．

4.3.3　結果と考察

提案手法の評価用データセット上での性能を報告し，また，それらに対する考察を行う．評価は，オーディオエフェクト構成推定，ウェットシグナル再構成に必要となるオーディオエフェクト除去，エフェクト構成推定の別の評価方法であるウェットシグナル再構成の 3 つの側面から行った．

結果の数値は，特記しない限り，対象となるデータセット全体にわたる平均値を示す．結果を示す図表において，↑は値が大きいほど，↓は値が小さいほど性能が良いことを示す．また，表中の太字は各指標において最良の値であることを示す．

表 4.3: オーディオエフェクトチェインについてのタイプ分類の性能評価．これらの結果は Dry-Type-Direct は探索段階と組み合わせて，Bypass-*-Iter は予測段階のみで得られたものである．

Method	Macro F_1 (↑)	LD (↓)	EMA (↑)
Dry-Type-Direct + Search	**0.958**	**0.313**	**0.774**
Bypass-Type-Iter	0.949	0.369	0.723
Bypass-Config-Iter	0.942	0.408	0.702

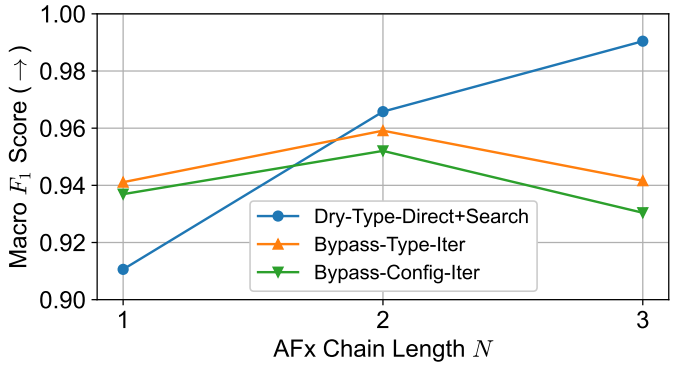


図 4.4: オーディオエフェクトタイプ分類におけるチェイン長ごとの Macro F_1 Score の傾向．

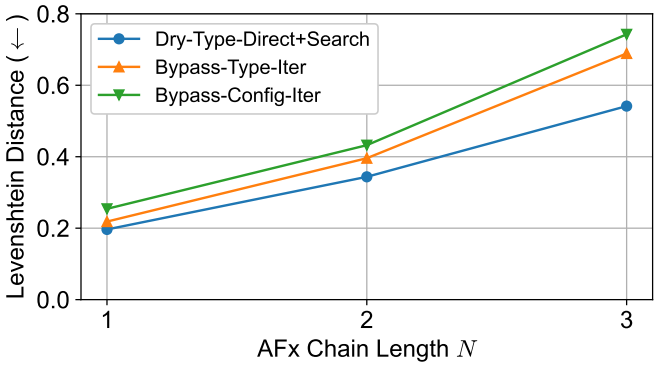


図 4.5: オーディオエフェクトタイプ分類におけるチェイン長ごとのレーベンシュタイン距離の傾向．

オーディオエフェクト構成推定

オーディオエフェクト構成推定の性能評価を行った．まず，予測段階において反復的な推論を行う Bypass-*-Iter の二手法について，チェイン中の最後に適用された単一のエフェクトのタイプ分類の評価を行った．タイプ c の単一ラベル分類の Macro F_1 Score による評価結果を表 4.2 に示す．なお，Macro F_1 Score は個別のサンプルからではなくデータセット全体があって初めて定義されるものであるから，サンプル全体にわたる平均値ではない．また，予測段階でエフェクトパラメータ回帰を行う Bypass-Config-Iter については，チェイン中の最後に適用された単一のエフェクトについてその評価を行うと，パラメータの MAE は 0.0885 となった．

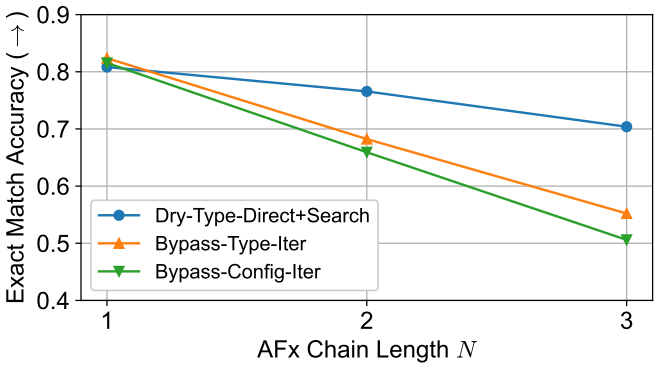


図 4.6: オーディオエフェクトタイプ分類におけるチェイン長ごとの厳密一致率の傾向.

表 4.4: 単一オーディオエフェクト除去（バイパスシグナル推定）の性能評価．予測段階において反復推論を行う手法について，チェイン中の最後に適用された単一のエフェクトの除去の評価を行った．

Method	SI-SDR (↑)	MR-STFT (↓)
Bypass-Type-Iter	**26.32**	**0.690**
Bypass-Config-Iter	26.30	0.691

次に，エフェクトチェイン全体のタイプの推定の評価を行った．タイプの組み合わせ $\{c_1, c_2, \ldots, c_N\}$ の多ラベル分類の Macro F_1 Score，タイプ列 C の推定値と真値の間のレーベンシュタイン距離（Levenshtein Distance; LD）および厳密一致率（Exact Match Accuracy; EMA）による評価結果を表 4.3 に示す．Macro F_1 Score は順序を考慮せずに組み合わせのみを評価する指標である．LD は編集距離（Edit Distance）とも呼ばれ，一方の記号列を 1 記号ずつ編集して他方の記号列に変形するのに必要な編集回数として定義されるため，部分的な一致も評価することが可能である．EMA は全体が厳密に一致しているかそうでないかの二値で評価する指標である．予測段階でタイプの組み合わせのみを推定する Dry-Type-Direct については，探索段階と組み合わせた “Dry-Type-Direct+Search” としての性能を示している．一方，Bypass-*-Iter については，予測段階で順序も含めてタイプ列を推定するため，ここでの結果は予測段階のみによるものである．

さらに，推定対象のチェインの長さ N ごとの傾向を，Macro F_1 Score，LD，EMA についてそれぞれ図 4.4，図 4.5，図 4.6 に示す．

表 4.3 においては，Dry-Type-Direct+Search がいずれの指標に関しても最良の性能を示している．チェイン長ごとの傾向を見ると，Macro F_1 Score 以外は，N の増大にともなって性能が低下する，すなわち，長いチェインほど推定が困難であるという直感と一致する結果となっている．そして，N が大きいほど Dry-Type-Direct+Search が優位な傾向がみられる．Macro F_1 Score に関しては，$N = 1$ では他の手法より劣っているにもかかわらず，$N = 3$ では他を上回る性能を示している．こうした Dry-Type-Direct+Search の優位性の背景には，Bypass-*-Iter で起こりうる反復推論による誤差蓄積を回避したことがあると考察される．また，Bypass-*-Iter の両手法の性能はいずれの指標においても差が小さく，パラメータの同時予測の有無はエフェクト推定性能に大きな影響を及ぼさないことが示唆されている．

表 4.5: オーディオエフェクトチェイン除去（ドライシグナル推定）の性能評価．これらの結果はいずれも予測段階のみで得られる結果である．

Method	SI-SDR (↑)	MR-STFT (↓)
Dry-Type-Direct	13.96	**0.813**
Bypass-Type-Iter	**14.95**	0.898
Bypass-Config-Iter	14.88	0.902

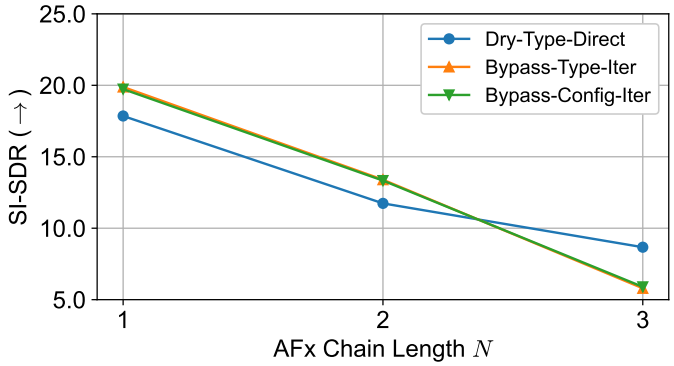


図 4.7: オーディオエフェクトチェイン除去におけるチェイン長ごとの SI-SDR の傾向．

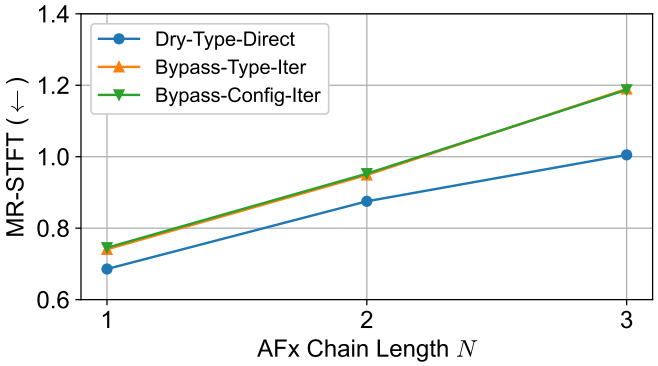


図 4.8: オーディオエフェクトチェイン除去におけるチェイン長ごとの MR-STFT の傾向．

オーディオエフェクト除去

オーディオエフェクト除去の性能評価を行った．エフェクト除去はいずれの手法においても予測段階のみによって行われるものである．まず，予測段階において反復的な推論を行うモデルについて，チェイン中の最後に適用された単一のエフェクトの除去によるバイパスシグナルの推定の評価を行った．バイパスシグナル $\boldsymbol{x}_{n-1}$ の推定値と真値の間の SI-SDR と MR-STFT による評価結果を表 4.4 に示す．

次に，エフェクトチェイン全体の除去後のドライシグナルの推定の評価を行った．ドライシグナル $\boldsymbol{x}_0$ の推定値と真値の間の SI-SDR と MR-STFT による評価結果を表 4.5 に示す．

表 4.6: ウェットシグナル再構成を通したオーディオエフェクト推定の総合的な性能評価．再構成に用いるドライシグナルとして推定値を用いた場合と真値を用いた場合の両方で評価を行った．エフェクト除去性能の影響を受けない純粋なエフェクト構成推定性能は後者に現れる．また，ベースラインとして予測段階のみによる手法についても評価を行った．

Method	SI-SDR (↑)	MR-STFT (↓)
reconstructed with estimated dry signal $\hat{\boldsymbol{x}}_0$		
Bypass-Config-Iter (Baseline)	16.80	0.807
Dry-Type-Direct + Search	18.57	**0.659**
Bypass-Type-Iter + Search	**21.23**	0.731
Bypass-Config-Iter + Search	21.04	0.735
reconstructed with ground-truth dry signal $\boldsymbol{x}_0$		
Bypass-Config-Iter (Baseline)	18.18	0.465
Dry-Type-Direct + Search	**23.07**	**0.340**
Bypass-Type-Iter + Search	22.68	0.361
Bypass-Config-Iter + Search	22.64	0.366

さらに，推定対象のチェインの長さ N ごとの傾向を，SI-SDR，MR-STFT についてそれぞれ図 4.7，図 4.8 に示す．

表 4.5 からは，SI-SDR では Bypass-Type-Iter が，MR-STFT では Dry-Type-Direct が最良の性能を示していることが分かる．チェイン長さごとの傾向を見ると，ここでも N の増大にともなって性能が低下するという直感と一致する結果となった．図 4.7 からは，SI-SDR では短いチェインでは Bypass-*-Iter が，長いチェインでは Dry-Type-Direct が高性能を示す傾向を確認できる．一方，図 4.8 からは，MR-STFT ではいずれのチェイン長においても Dry-Type-Direct が最良の性能を示していることが分かる．以上の結果では，時間領域での比較に基づく SI-SDR と時間周波数領域での比較に基づく MR-STFT による評価の差異が現れている．また，ここでも Bypass-*-Iter の両手法の性能の差は小さい．

ウェットシグナル再構成

ウェットシグナル再構成を通してオーディオエフェクト推定の総合的な性能評価を行った．ウェットシグナル再構成信号 $\boldsymbol{x}_{\mathrm{r}}$ と真のウェットシグナル $\boldsymbol{x}_N$ の間の SI-SDR と MR-STFT による評価結果を表 4.4 に示す．再構成に用いるドライシグナルとして推定値を用いた場合と真値を用いた場合の両方で評価を行った．前者の指標は探索段階で利用されている指標であり探索の信頼性の目安にはなるが，誤差を含んだ推定ドライシグナルに基づくものであり，これが良いほどエフェクト構成推定が正確であるとは限らない．提案手法の最も主要な目的であったエフェクト構成推定性能は後者の指標に現れる．また，Bypass-Config-Iter の予測段階のみによる手法をベースラインとして評価した．

さらに，推定対象のチェインの長さ N ごとの傾向を示す．推定ドライシグナルを用いた再構成における SI-SDR と MR-STFT についてはそれぞれ図 4.9 と図 4.10 に，真のドライシグナルを用いた再構成における SI-SDR と MR-STFT についてはそれぞれ図 4.11 と図 4.12 に示す．

表 4.6 においては，いずれの指標においても提案アプローチに基づく全手法がベースラインを上回る性能を

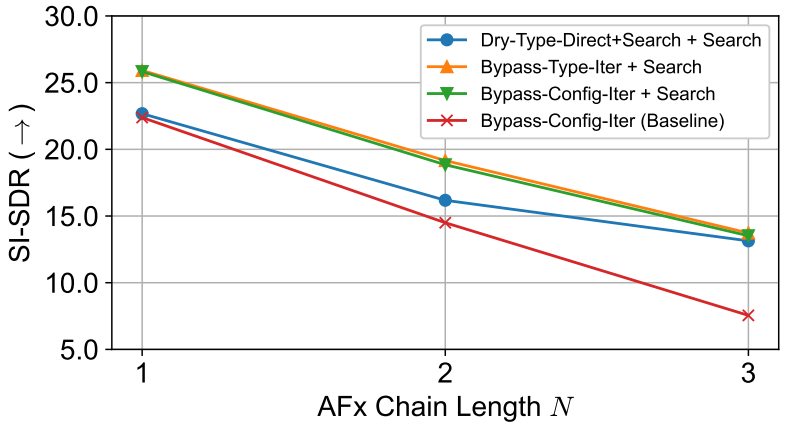


図 4.9: ウェットシグナル再構成におけるチェイン長ごとの SI-SDR の傾向．これは推定ドライシグナルを用いて再構成を行った場合の結果である．

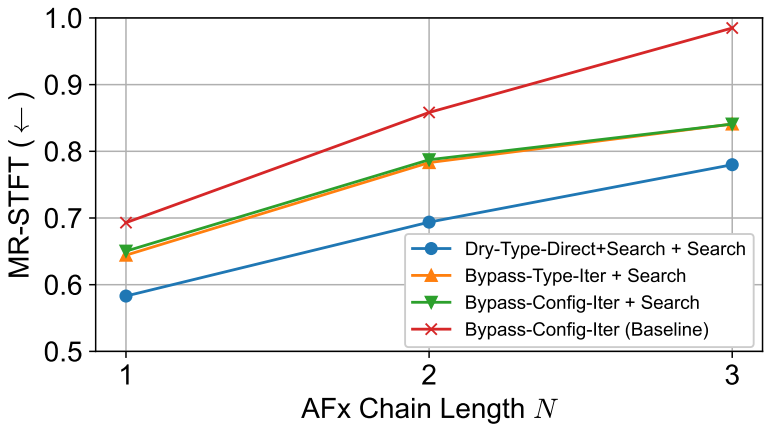


図 4.10: ウェットシグナル再構成におけるチェイン長ごとの MR-STFT の傾向．これは推定ドライシグナルを用いて再構成を行った場合の結果である．

示しており，提案アプローチの有効性が示されている．ドライシグナルの真値を用いた再構成を通した評価においては，いずれの指標においても Dry-Type-Direct+Search が最良の性能を示している．ドライシグナルの推定値を用いた再構成を通した評価においては，SI-SDR では Bypass-Type-Iter+Search が，MR-STFT では Dry-Type-Direct+Search が最良の性能を示している．これは，表 4.5 に現れていた性能差と対応している．図 4.11 と図 4.12 からは，いずれのチェイン長においても提案手法がベースラインを上回っていることが分かる．また，一方，図 4.8 からは，MR-STFT ではいずれのチェイン長においても Dry-Type-Direct が最良の性能を示していることが分かる．短いチェインでは提案手法間の性能差が少ないが，長いチェインでは Dry-Type-Direct+Search が高性能である傾向があることが分かる．また，ここでも Bypass-*-Iter+Search の両手法の性能の差は小さい．

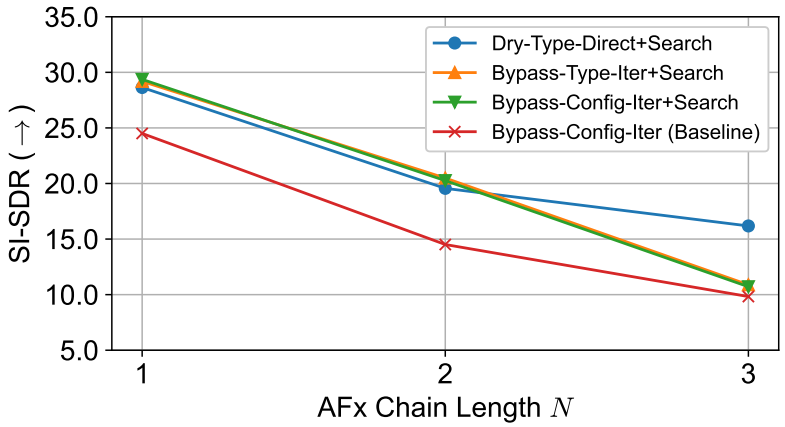


図 4.11: ウェットシグナル再構成におけるチェイン長ごとの SI-SDR の傾向．これは真のドライシグナルを用いて再構成を行った場合の結果である．

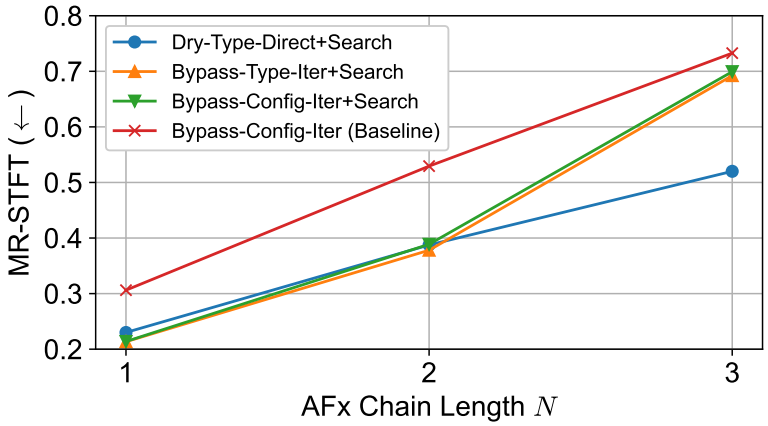


図 4.12: ウェットシグナル再構成におけるチェイン長ごとの MR-STFT の傾向．これは真のドライシグナルを用いて再構成を行った場合の結果である．

第 5 章

課題と展望

本研究において残された課題とそれらについての今後の展望は次のように整理できる．

- **性能分析**
  提案手法の性能の分析には余地も残されている．本稿で示した性能評価は，評価データ全体にわたる統計値とエフェクトチェインの長さごとの傾向であった．しかし，エフェクトタイプごとの傾向や，これらのばらつきの分析を行う余地が存在する．さらなる考察や手法の改善のためこのような分析を進めることは今後の課題である．
- **探索規模のチューニング**
  提案アプローチの探索段階では，推定されたドライシグナルを用いた最適化を行う．オーディオエフェクト推定においては，ドライシグナルの真値と推定されたエフェクト用いたウェットシグナル再構成類似度が最大になることが理想的な推定である．しかし，探索段階では真のドライシグナルは未知であるため，次善策として推定ドライシグナルを利用して再構成を行っている．よって，推定ドライシグナルは誤差を含むものであり，この最適化は過剰適合を起こす恐れがある．推定ドライシグナルを用いた再構成類似度が最大であるようなエフェクト構成は，真のドライシグナルを用いた再構成では最適ではない可能性があるのである．そこで，まずは，このような過剰適合の発生可否や挙動を観察する必要がある．そのうえで，過剰適合が起こる前の，真のドライシグナルを用いた再構成が最適となる段階まで探索を行うように，探索の規模をチューニングすることが考えられる．このようなチューニングは，式 (4.10) の試行回数のパラメータ M_0, r に対して行うことが想定される．第 4.3.3 節では，真のドライシグナルを用いた再構成性能に基づいて提案アプローチの有効性を示し，全体的としてはこのような過剰適合の傾向は小さいと言える．しかし，一部では過剰適合を起こした事例も観察されていた．そこで，小規模な観察にのみ基づいて決定されていた探索規模のパラメータをチューニングすることで，より高精度な推定が期待される．
- **エフェクトの多様性**
  本研究の課題として，最後に，扱ったエフェクトの多様性の限界を挙げる．本研究で推定対象としたエフェクトのタイプ，可変パラメータやその範囲，個数やルーティングなどは，実用的なサウンドデザインを考えると，限定的なものである．例えば，Chorus の rate は中心的な役割を果たす重要なパラメータだが，探索が困難であることが観察されたため，本研究では固定し推定対象としなかった．このパラメータに関しては，探索空間において真値の近傍で急激に再構成類似度が減少するような，急峻で最適化の難しい地形が観察された．このようにパラメータの小さな変化にともなってウェットシグナルの大

きな変化が起こるようなエフェクトやパラメータの推定が，本研究の手法のままでは困難だと予想される．また，今回扱わなかったタイプのエフェクトやより長いエフェクトチェインも現実では用いられる．例えば，Delay，Compressor が，本研究で扱わなかったがよく用いられるエフェクトとして挙げられる．このようなより多様なエフェクトについて有効な手法を探求することも今後の課題である．

第 6 章

結論

本研究では，オーディオエフェクト推定において，従来の予測的アプローチと探索的アプローチを組み合わせた新たなアプローチを提案した．このアプローチでは，まず，DNN でドライシグナルとエフェクト構成の全体または一部の予測を行い，その後，それらを用いてウェットシグナル再構成類似度に基づく探索を行う．予測段階でドライシグナルの推定を行うことで再構成類似度を評価することが可能となり，探索段階でこれを目的関数として予測の補完や改善を行う．

提案手法と予測のみによるベースラインについて，ギターによる音楽的抜粋とオーディオエフェクトチェインについての評価実験を行った．この際，いずれも提案アプローチに基づくが予測と探索の各段階間でのタスクの分担が異なる手法を比較した．その結果，提案アプローチに基づく全手法が予測のみによる手法を上回る性能を示し，提案アプローチの有効性が示された．また，タスクの分担については，予測段階でエフェクトタイプの組み合わせのみを予測し探索段階でその順序を推定する手法が多くの指標において最適であることが示された．この背景には，他の手法で起こり得る反復推論による誤差蓄積の回避による影響があると考察される．また，タスクの分担については，予測段階においてエフェクトタイプの組み合わせのみを予測し探索段階においてその順序を推定する手法が，多くの指標において最適であることが示された．さらに，推定対象エフェクトチェインの長さごとの各性能指標の傾向も調査した．その結果，多くの指標において，長いチェインほど推定が困難であることが示された．また，推定対象がそのような長いチェインであるほど探索段階においてエフェクトタイプの順序を推定する手法が高い性能を示す傾向にあり，順序推定における探索手法の有効性が示唆された．

残された課題としては，提案手法のさらなる性能分析，探索規模のチューニング，より多様なエフェクトへの拡張を挙げた．今後は，これらに対処することで，オーディオエフェクト推定に関するより多くの知見を獲得するとともに，より汎用的で高精度な推定を実現することが期待される．

謝辞

著者の指導教員である関西学院大学工学部教授の片寄晴弘先生には，本研究の着想から，実験の遂行，論文の執筆に至るまで，様々な観点からご指導をいただいた．ここに深く感謝する．著者が所属する関西学院大学工学部片寄研究室秘書の長澤育子様には，研究室の環境整備や研究会参加にあたっての事務手続きなどの面から，研究活動を支えていただいた．ここに深く感謝する．また，片寄研究室に同時期に在籍した学生たちとは研究内容について議論し，彼らから研究への取り組み方を見習い，彼らのおかげで研究室の日常も快適で楽しいものとなった．ここに深く感謝する．最後に，本研究の遂行は，家族や親族によって支えられた日常の中にあった．彼らにも深く感謝する．

2026 年 2 月 27 日

沖田 陽一

参考文献

[1] Udo Zölzer, editor. *DAFX - Digital Audio Effects (Second Edition)*. John Wiley & Sons, 2011.

[2] Joshua D. Reiss and Andrew P. McPherson. *Audio Effects Theory, Implementation and Application*. CRC Press, 2015.

[3] Thomas Wilmering, David Moffat, Alessia Milo, and Mark B. Sandler. A History of Audio Effects. *Applied Sciences*, Vol. 10, No. 3, p. 791, 2020.

[4] Marco Comunità and Joshua D. Reiss. AFxResearch: a repository and website of audio effects research. In *Proceedings of the DMRN+19: Digital Music Research Network One-day Workshop*, 2024.

[5] 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 (DAFx)*, pp. 310–316, 2020.

[6] 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.

[7] Christopher Mitcheltree, Christian J. Steinmetz, Marco Comunità, and Joshua D. Reiss. Modulation Extraction for LFO-driven Audio Effects. In *Proceedings of the 26th International Conference on Digital Audio Effects (DAFx)*, 2023.

[8] Côme Peladeau, Dominique Fourer, and Geoffroy Peeters. Audio Processor Parameters: Estimating Distributions Instead of Deterministic Values. In *Proceedings of the 28th International Conference on Digital Audio Effects (DAFx)*, pp. 275–282, 2025.

[9] Reemt Hinrichs, Kevin Gerkens, Alexander Lange, and Jörn Ostermann. 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, p. 28, 2022.

[10] Jinyue Guo and Brian McFee. Automatic Recognition of Cascaded Guitar Effects. In *Proceedings of the 26th International Conference on Digital Audio Effects (DAFx)*, 2023.

[11] Sungho Lee, Jaehyun Park, Seungryeol Paik, and Kyogu Lee. Blind Estimation of Audio Processing Graph. In *Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

[12] Côme Peladeau and Geoffroy Peeters. Blind Estimation of Audio Effects Using an Auto-Encoder Approach and Differentiable Digital Signal Processing. In *Proceedings of the 2024 IEEE International*

Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 856–860, 2024.

[13] Michele Rossi, Giovanni Iacca, and Luca Turchet. Automatic Classification of Chains of Guitar Effects Through Evolutionary Neural Architecture Search. In *Proceedings of the 28th International Conference on Digital Audio Effects (DAFx)*, pp. 350–357, 2025.

[14] Aogu Wada, Tomohiko Nakamura, and Hiroshi Saruwatari. Hyperbolic Embeddings for Order-Aware Classification of Audio Effect Chains. In *Proceedings of the 28th International Conference on Digital Audio Effects (DAFx)*, pp. 396–402, 2025.

[15] Osamu Take, Kento Watanabe, Takayuki Nakatsuka, Tian Cheng, Tomoyasu Nakano, Masataka Goto, Shinnosuke Takamichi, and Hiroshi Saruwatari. Audio Effect Chain Estimation and Dry Signal Recovery From Multi-Effect-Processed Musical Signals. In *Proceedings of the 27th International Conference on Digital Audio Effects (DAFx)*, pp. 1–8, 2024.

[16] Christian J. Steinmetz, Shubhr Singh, Marco Comunità, Ilias Ibnyahya, Shanxin Yuan, Emmanouil Benetos, and Joshua D. Reiss. ST-ITO: Controlling Audio Effects for Style Transfer With Inference-Time Optimization. In *Proceedings of the 25th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 661–668, 2024.

[17] Chin-Yun Yu, Marco A. Martínez-Ramírez, Junghyun Koo, Wei-Hsiang Liao, Yuki Mitsufuji, and György Fazekas. Improving Inference-Time Optimisation for Vocal Effects Style Transfer with a Gaussian Prior. In *Proceedings of the 2025 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2025.

[18] Soumya Sai Vanka, Christian Steinmetz, Jean-Baptiste Rolland, Joshua Reiss, and György Fazekas. Diff-MST: Differentiable Mixing Style Transfer. In *Proceedings of the 25th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 563–570, 2024.

[19] Junghyun Koo, Marco A. Martínez-Ramírez, Wei-Hsiang Liao, Giorgio Fabbro, Michele Mancusi, and Yuki Mitsufuji. ITO-Master: Inference-Time Optimization for Audio Effects Modeling of Music Mastering Processors. In *Proceedings of the 26th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 134–141, 2025.

[20] Jesse Engel, Lamtharn Hantrakul, Chenjie Gu, and Adam Roberts. DDSP: Differentiable Digital Signal Processing. In *Proceedings of the 8th International Conference on Learning Representations (ICLR)*, 2020.

[21] Michael Stein, Jakob Abeßer, Christian Dittmar, and Gerald Schuller. Automatic Detection of Audio Effects in Guitar and Bass Recordings. In *Proceedings of the 128th Audio Engineering Society Convention*, 2010.

[22] Michael Stein. Automatic Detection of Multiple, Cascaded Audio Effects in Guitar Recordings. In *Proceedings of the 13th International Conference on Digital Audio Effects (DAFx)*, 2010.

[23] Maximilian Schmitt and Björn Schuller. Recognising Guitar Effects – Which Acoustic Features Really Matter? In *Proceedings of the 47th Annual Conference of the German Informatics Society (INFORMATIK)*, pp. 177–190, 2017.

[24] Di Sheng and György Fazekas. Automatic Control of the Dynamic Range Compressor Using a Regression Model and a Reference Sound. In *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx)*, pp. 160–167, 2017.

[25] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 30, 2017.

[26] 有山 大地, 安藤 大地, 串山 久美子. ソフトウェアエフェクタを利用した同一機材を必要としない機械学習によるエレキギター音色の自動再現手法の検討. 情報処理学会論文誌, Vol. 61, No. 11, pp. 1729–1740, 2020.

[27] Chin-Yun Yu, Marco A. Martínez-Ramírez, Junghyun Koo, Ben Hayes, Wei-Hsiang Liao, György Fazekas, and Yuki Mitsufuji. DiffVox: A Differentiable Model for Capturing and Analysing Vocal Effects Distributions. In *Proceedings of the 28th International Conference on Digital Audio Effects (DAFx)*, pp. 334–341, 2025.

[28] Joseph T. Colonel and Joshua Reiss. Reverse Engineering a Nonlinear Mix of a Multitrack Recording. *Journal of the Audio Engineering Society*, Vol. 71, No. 9, pp. 586–595, 2023.

[29] Sungho Lee, Marco A. Martínez-Ramírez, Wei-Hsiang Liao, Stefan Uhlich, Giorgio Fabbro, Kyogu Lee, and Yuki Mitsufuji. Reverse Engineering of Music Mixing Graphs With Differentiable Processors and Iterative Pruning. *Journal of the Audio Engineering Society*, Vol. 73, No. 6, pp. 344–365, 2025.

[30] Johannes Imort, Giorgio Fabbro, Marco A. Martínez-Ramírez, Stefan Uhlich, Yuichiro Koyama, and Yuki Mitsufuji. Distortion Audio Effects: Learning How to Recover the Clean Signal. In *Proceedings of the 23rd International Society for Music Information Retrieval Conference (ISMIR)*, pp. 218–225, 2022.

[31] Chang-Bin Jeon and Kyogu Lee. Music De-Limiter Networks Via Sample-Wise Gain Inversion. In *Proceedings of the 2023 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2023.

[32] Ying-Shuo Lee, Yueh-Po Peng, Jui-Te Wu, Ming Cheng, Li Su, and Yi-Hsuan Yang. Distortion Recovery: A Two-Stage Method for Guitar Effect Removal. In *Proceedings of the 27th International Conference on Digital Audio Effects (DAFx)*, pp. 177–184, 2024.

[33] Matthew Rice, Christian J. Steinmetz, George Fazekas, and Joshua D. Reiss. General Purpose Audio Effect Removal. In *Proceedings of the 2023 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2023.

[34] Reemt Hinrichs, Kevin Gerkens, Alexander Lange, and Jörn Ostermann. Blind extraction of guitar effects through blind system inversion and neural guitar effect modeling. *EURASIP Journal on Audio, Speech, and Music Processing*, Vol. 2024, No. 1, p. 9, 2024.

[35] Simon Rouard, Francisco Massa, and Alexandre Défossez. Hybrid Transformers for Music Source Separation. In *Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

[36] Ryuichi Yamamoto, Eunwoo Song, and Jae-Min Kim. Parallel Wavegan: A Fast Waveform Generation Model Based on Generative Adversarial Networks with Multi-Resolution Spectrogram. In *Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6199–6203, 2020.

[37] Jonathan Le Roux, Scott Wisdom, Hakan Erdogan, and John R Hershey. SDR – half-baked or well done? In *Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal*

Processing (ICASSP), pp. 626–630, 2019.

[38] Nikolaus Hansen and Andreas Ostermeier. Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation. In *Proceedings of the 1996 IEEE International Conference on Evolutionary Computation*, pp. 312–317, 1996.

[39] James Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. Algorithms for Hyper-Parameter Optimization. In *Advances in Neural Information Processing Systems (NIPS)*, 2011.

[40] Christian Kehling, Jakob Abeßer, Christian Dittmar, and Gerald Schuller. Automatic Tablature Transcription of Electric Guitar Recordings by Estimation of Score- and Instrument-Related Parameters. In *Proceedings of the 17th International Conference on Digital Audio Effects (DAFx)*, pp. 219–226, 2014.

[41] Qingyang Xi, Rachel M. Bittner, Johan Pauwels, Xuzhou Ye, and Juan P. Bello. GuitarSet: A Dataset for Guitar Transcription. In *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 453–460, 2018.

[42] Yu-Hua Chen, Wen-Yi Hsiao, Tsu-Kuang Hsieh, Jyh-Shing Roger Jang, and Yi-Hsuan Yang. Towards Automatic Transcription of Polyphonic Electric Guitar Music: A New Dataset and a Multi-Loss Transformer Model. In *Proceedings of the 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 786–790, 2022.

[43] Hegel Pedroza, Wallace Abreu, Ryan M. Corey, and Iran R. Roman. Guitar-TECHS: An Electric Guitar Dataset Covering Techniques, Musical Excerpts, Chords and Scales Using a Diverse Array of Hardware. In *Proceedings of the 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025.

[44] Christian J. Steinmetz and Joshua D. Reiss. auraloss: Audio-focused loss functions in PyTorch. In *Proceedings of the DMRN+15: Digital music research network one-day workshop*, 2020.

[45] Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization. In *Proceedings of the 7th International Conference on Learning Representations (ICLR)*, 2019.

[46] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A Next-generation Hyperparameter Optimization Framework. In *KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2623–2631, 2019.

[47] Nikolaus Hansen. The CMA Evolution Strategy: A Tutorial. 2016. arXiv:1604.00772.

発表文献

本論文に関連する発表文献は，発表予定のものも含めると，以下の通りである．

[1] **沖田 陽一**, 片寄 晴弘. 目標演奏と異なるドライシグナル条件下でのオーディオエフェクト推定. 情報処理学会研究報告音楽情報科学（MUS）, Vol. 2025-MUS-142, No. 52, pp. 1–8, 2025.

[2] **沖田 陽一**, 片寄 晴弘. 深層学習に基づく予測と探索アルゴリズムによるオーディオエフェクト推定. 情報処理学会研究報告音楽情報科学（MUS）, Vol. 2025-MUS-145, No. 49, pp. 1–9, 2026. （予定）

[3] **Youichi Okita**, Haruhiro Katayose. Audio Effect Estimation with DNN-Based Prediction and Search Algorithm. In *Proceedings of the 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2026. （予定，採択済み）