Vol.2022-MUS-134 No.16
Vol.2022-SLP-142 No.16
2022/6/17

半教師あり深層異常検知手法を用いたクラシックギターにおける演奏ミス自動検出手法の提案

小川 健太[1]　澤田 隼[2]　桂田 浩一[2]　大村 英史[2]

概要：クラシックギターは気軽に始められる楽器であることから個人で練習することが多い．そのため，初心者が 1 音ずつ正しく弾けるようにサポートするシステムが求められている．エレキギターにおいては，適切に音が発せられたかを評価する練習支援システムが提案されている．このシステムでは，予め演奏ミスを分類し，演奏音のラベル付きデータから学習を行った分類器により演奏ミスの自動判定を行っている．しかしながら，演奏ミスをした音を大量に収集してラベル付けを行うことは困難であり，演奏ミスの分類も主観的な判断に大きく依存する問題が生じる．そこで本研究では，「適切に演奏された音」を正常データ，「演奏ミスをした音」を異常データとし，半教師ありの深層異常検知手法を用いた演奏ミスの自動検出手法を提案する．異常検知は，正常データ群から外れるものを一様に異常データとして検出するため，起こり得る様々な演奏ミスを検出することが期待できるうえに，データセットの構築も容易となる．評価実験を実際に演奏したデータに基づいて行ったところ，高い精度で演奏ミスを検出でき，提案手法の有効性が確認できた．

1. はじめに

ギターは，左手の指で弦を押さえながら右手の指で弦を弾くという複合的な動作により演奏する．そのため，どちらが適切にできていなかったり，両手がうまく連動しないと思った通りに音を奏でることができず，上達には多くの時間と努力を要する．特にクラシックギターは他のエレキギターやフォークギターと比べて弦高（指板から弦までの高さ）が高く，ネックが広い．そのため，初心者にとっては弦を正確に押さえるのが大変なことが多く，単音の基礎練習から丁寧に練習することが大切である．したがって本研究では，クラシックギターの単音演奏において，1 音ずつ正しく弾けるようサポートすることを目的とする．

ギター演奏の練習支援についての研究は数多く行われてきた．例えば，拡張現実（Augmented Reality：AR）表示技術を用いて正しい運指情報をユーザに提示するシステム [1][2][3]，ギターにセンサを取り付けて正しく押弦できているかを判定するシステム [4][5]，筋電計を用いて指の微細な動きや姿勢を検出するシステム [6] などが提案されている．しかし，これらのシステムでは音が正しく演奏されたかの評価は行っていない．演奏音の正確さの判定を組み込んだ研究として，カメラとセンサを利用したシステム [7] や，楽曲データの分析による練習曲としての重要性を考慮したシステム [8] が提案されているが，判定の際演奏するフレットは指定されており，判定制度の評価も行われていない．また，上記の研究は主にエレキギターやフォークギターのコード演奏についてしか議論されていない．

これに対し，下尾らは，エレキギター演奏を自動評価するための音響的特徴量を調査し [9]，それらを用いてエレキギターの単音演奏における押弦ミスの自動検出を行う練習支援システムを作成した [10]．このシステムでは，演奏音を予め「フレット上で弦を押さえている音 / 十分に弦を押さえられていない音 / 正しく押弦できている音」の合計 3 つに分類し，音響的特徴量によるクラス分類を行っている．下尾らの研究のようにミスを検出するシステムであれば，単音が適切に演奏されたかの評価が可能である．しかしながら，この手法では予めミスを分類する必要があり，それらの音を教師データとして用意しラベル付けを行う手間がかかる．

そこで本研究では，異常検知手法を用いてこれを解決することを試みる．具体的には，クラシックギターでの単音演奏において「適切に演奏された音」を正常データ，「演奏ミスをした音」を異常データとして半教師あり深層異常検知手法を用いることで，演奏ミスを自動検出する手法を提案する．異常検知（Anomaly Detection）とは，大多数のデータから逸脱した振る舞いを示すデータを識別するこ

[1] 東京理科大学大学院 理工学研究科
[2] 東京理科大学 理工学部

© 2022 Information Processing Society of Japan

と [11] である．通常，正常データと比べて異常データは手に入る数が少ないと仮定できるため，異常検知は基本的にラベルなしデータのみを用いる教師なし学習，もしくはさらに少量のラベル付きデータを用いる半教師あり学習の問題として扱われる [12]．本研究においては，ギター演奏における「適切に演奏された音」でデータ群を構成したとき，「演奏ミスをした音」はそのデータ群から外れると考えられるため，異常検知の一つである外れ値検出（Outlier Detection）が適用できる．これによりミスの分類を行うことなく，起こり得る様々な演奏ミスを検出することが期待できる．

2. クラシックギターにおける演奏ミス

ギター演奏において押弦や弾弦が不適切であると，思ったように音が奏でられない．一般的にこれはミスとされ，ギターを練習する際はなるべくミスをしないよう意識する必要がある．以下に，本手法で検出するクラシックギターの主な演奏ミスと，それによって生じる不適切な音について列挙した．

十分に弦を押さえられていない
弦を十分に押さえられていないと，フレットと弦が満足に接触せず，弦の振動が不規則になる．それにより，振動している弦がフレットなどに細く何度も接触を繰り返すことでエッジの効いたノイズが鳴ってしまうことがある．このノイズは「ビビリ」と呼ばれる．また，フレットと弦が全く接触していないと，弦の振動を直接指で止めている状態になり調波音がほとんど鳴らなくなってしまう（以降この音を「ミュート音」とする）．

フレットから遠くを押さえている
押弦はなるべくフレットのすぐ隣で行うのが望ましい．フレットから離れた位置を押さえると，フレットと弦が十分に接触しないことでビビリやミュート音が生じる場合がある．

フレットの上を押さえている
フレットの上を押さえてしまうと指が触れたまま弦が振動してしまい，弦の振動が急激に減衰し音がこもってしまう．

押弦が適切に継続できていない
弦をはじいて音が鳴っている間押弦は継続する必要があるが，押弦している指が緩むと，音が鳴っている途中でビビリが生じることがある．また，ポジション移動[*1]などで指を早く離した場合，音が不自然に切れて次の音と滑らかにつながらない[*2]．

弾弦が強すぎる
特に低音弦において，弾弦の力が強すぎると弦の振動が必要以上に大きくなってしまい，フレットと接触してビビリが発生する場合がある．

振動中の弦に指が接触する
弦が振動して音を鳴らしている最中に，押弦している指以外の指が接触することで急激に減衰したりビビリなどのノイズが発生することがある．

これら以外にも演奏ミスは起こり得るが，それら全てを取り上げて分類することは非常に困難である．これに対して異常検知手法を利用した本手法では，分類を必要とせずにこれらを含めたあらゆる演奏ミスを検出することが可能である．

3. 提案手法：異常検知手法を用いた演奏ミス自動検出

異常検知手法を用いた演奏ミス自動検出手法の構造を図 1 に示す．この手法では，まず入力された演奏音を 1 音ずつに分割して，それぞれを一定の長さに区切りメル周波数スペクトログラムに変換する．これらの前処理を施した演奏音データを演奏ミス検出モデルに入力し，それぞれの異常スコアを得る．そして，異常スコアが予め設定したしきい値を超えた演奏音をユーザに提示することで，演奏ミスを検出する．本研究では，Ruff らの提案した半教師あり深層異常検知モデルである Deep Semi-Supervised Anomaly Detection（Deep SAD）[13] を演奏ミス検出モデルとして用いる．

3.1 前処理

音響信号をニューラルネットワークに入力する際，より周波数成分の特徴を捉えやすいデータに変換されることが多い．本研究では，データ量を圧縮しつつ演奏音の本質的な特徴が表された形とするため，人間の音高知覚を考慮したメル周波数スペクトログラムに変換する．また，演奏音を一音ずつネットワークに入力するためにオンセット（発音時刻）検出を用いる．これらの前処理を以下の順序で行う．

(1) 入力された演奏データ（44100 Hz）から，オンセット（発音時刻）検出を用いて 1 音ずつに分割する．
(2) それぞれの音を 370 ms までに区切る（消音時のノイズ[*3]を入力音に含まないようにするため）．370 ms に満たない音には，足りない分ゼロパディングを適用する．
(3) それぞれの音をメル周波数スペクトログラムに変換し，正規化する．

[*1] 左手人差し指の位置（フレット）をポジションと呼び，今のポジションでは押さえられないフレットを押さえる必要がある時に，左手の位置を移動することをポジション移動と呼ぶ．
[*2] 意図的に音を短く切るスタッカートという奏法もある．
[*3] 弦をはじく直前に指が弦に触れる際，前の音を同一弦で鳴らしていたときに軽くノイズがなる場合がある．

© 2022 Information Processing Society of Japan

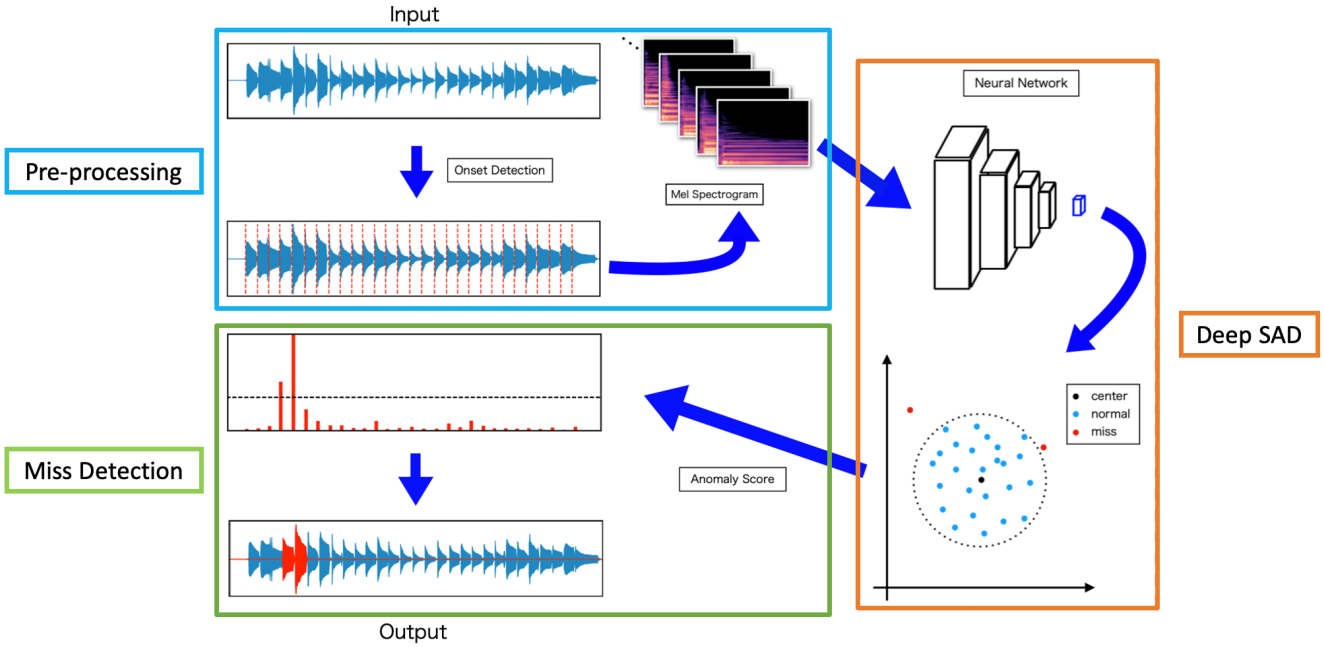


図 1: 半教師あり深層異常検知モデル Deep SAD を用いた演奏ミス自動検出手法．演奏データの音響信号をオンセット検出により分割し，それぞれをメル周波数スペクトログラムに変換する（赤い点線はオンセットを示す）．これを Deep SAD ネットワークに入力し，射影された特徴空間で超球の中心との距離を計算する．その距離を異常スコアとし，異常スコアがしきい値を超えた音を「演奏ミスをした音」としてユーザに提示する．

オンセット検出には，Python の librosa パッケージ [14] に含まれる onset モジュールの，onset_detect 関数を用いた．また，短時間フーリエ変換の窓関数をハニング窓，フレーム長を 512 samples，シフト長を 256 samples とし，メルフィルタバンクの数を 128 個，最大周波数を 22050 Hz とする．したがって，変換後のメル周波数スペクトログラムのサイズは 128×32 である．また，音圧はデシベルスケールに対数変換している．

3.2 異常検知手法：Deep SAD

Deep SAD は，外れ値検出の既存手法である Support Vector Data Description（SVDD）[15] をベースとした手法である．SVDD は，非線型カーネルにより射影された高次元特徴空間において正常データを包含する最小の識別超球を探索し，超球の内部に属さないデータを外れ値，すなわち異常データと判定する．Ruff らは，SVDD の非線型カーネルをニューラルネットワークに置き換えて高次元な入力データに対して外れ値検出を行う Deep Support Vector Data Description（Deep SVDD）[12] を提案した．Deep SVDD を含め，通常の異常検知手法はラベルなしの正常（と想定される）データのみで教師なし学習を行うが，現実では異常データが少数は手に入ることが期待される．それを踏まえ，Ruff らは Deep SVDD を半教師ありに拡張した Deep SAD を提案し，特に複雑な画像データに対して非常に高い有効性を示している [13]．本研究においても少数の「演奏ミスをした音」が手に入ると仮定できるため，「適切に演奏された音」を正常データ，「演奏ミスをした音」を異常データとし，演奏音を画像データに見立てて Deep SAD を適用する．

入力空間を $\mathcal{X} \subseteq \mathbb{R}^D$，出力空間を $\mathcal{F} \subseteq \mathbb{R}^d$ とする．$L \in \mathbb{N}$ の隠れ層で構成されるニューラルネットワークを $\phi(\cdot;\mathcal{W}) : \mathcal{X} \to \mathcal{F}$ とし，重みを $\mathcal{W} = \mathbf{W}^1, \ldots, \mathbf{W}^L$ に設定する．ここで，$\mathbf{W}^\ell$ は隠れ層 $\ell \in 1, \ldots, L$ の重みである．すなわち，$\phi(\mathbf{x}; W) \in \mathcal{F}$ は，パラメータ $\mathcal{W}$ のネットワーク ϕ により出力された $\mathbf{x} \in \mathcal{X}$ の特徴表現である．したがって Deep SAD の目的は，出力空間 $\mathcal{F}$ において正常データ全てを包含する中心 $\mathbf{c}$ の超球を最小にするようなパラメータ $\mathcal{W}$ を学習することである．データセットには，n 個のラベルなしデータ $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathcal{X}$, $\mathcal{X} \subseteq \mathbb{R}^D$ と，m 個のラベル付きデータ $(\tilde{x}_1, \tilde{y}_1, \ldots, \tilde{x}_m, \tilde{y}_m) \in \mathcal{X} \times \mathcal{Y}$, $\mathcal{Y} = \{-1, +1\}$ を用いる．ここで，$\tilde{y} = +1$ は既知の正常データ，$\tilde{y} = -1$ は既知の異常データを示すラベルであり，ラベルなしデータの大多数が正常データであると仮定する．このとき，Deep SAD の目的関数は以下の通りである：

$$\min_{\mathcal{W}} \frac{1}{n+m} \sum_{i=1}^{n} ||\phi(\mathbf{x}_i; \mathcal{W}) - \mathbf{c}||^2 + \frac{\eta}{n+m} \sum_{j=1}^{m} (||\phi(\tilde{\mathbf{x}}_j; \mathcal{W}) - \mathbf{c}||^2)^{\tilde{y}_j} + \frac{\lambda}{2} \sum_{\ell=1}^{L} ||\mathbf{W}^\ell||_F^2, \quad \lambda > 0.$$

ラベルなしデータには，射影された点と超球の中心との距

表 1: エンコーダの構造

Layer	Output Size	Stride	Padding	Activation
Input	(1,128,32)	–	–	–
Conv3x3	(256,64,16)	(2,2)	(1,1)	ReLU
Conv3x3	(128,32,8)	(2,2)	(1,1)	ReLU
Conv3x3	(64,16,4)	(2,2)	(1,1)	ReLU
Conv3x3	(32,8,4)	(2,2)	(1,1)	ReLU
Linear	(64)	–	–	–

表 2: デコーダの構造

Layer	Output Size	Stride	Padding	Activation
Input	(64)	–	–	–
Linear	(1024)	–	–	–
Deconv2x1	(64,16,4)	(2,1)	–	ReLU
Deconv2x2	(128,32,8)	(2,2)	–	ReLU
Deconv2x2	(256,64,16)	(2,2)	–	ReLU
Deconv2x2	(1,128,32)	(2,2)	–	Sigmoid

離, すなわち特徴表現と中心の平均二乗誤差（Mean Squared Error：MSE）を損失関数として与える．ラベル付き正常データ（$\tilde{y}=+1$）に対しては，特徴表現と中心の MSE にハイパーパラメータ $\eta>0$ を乗算したものを，ラベル付き異常データ（$\tilde{y}=-1$）にはその逆数をそれぞれ損失関数として与える．η はラベル付きデータとラベルなしデータのバランスを制御し，$\eta>1$ に設定した場合はラベル付きデータに，$\eta<1$ に設定した場合はラベルなしデータに重点を置く．以上によりネットワークは，正常データが集中するような潜在分布を学習するとともに，正常データが超球の中心のより近くへ，異常データが中心からより遠くへ射影されるよう学習する．最終項は重み正則化項であり，$||\cdot||_F$ はフロベニウスノルムを示す．

Deep SAD は，正常データのみでオートエンコーダの事前学習を行い，収束したエンコーダの重みでネットワーク ϕ のパラメータ $\mathcal{W}$ を初期化する．超球の中心 $\mathbf{c}$ は，初期化したネットワークに正常データを入力して得られた出力の平均値に設定する．テストデータ $\mathbf{x}\in\mathcal{X}$ に対しては，特徴表現と超球の中心の二乗誤差を測ることで異常スコア s を算出する：

$$s(\mathbf{x})=||\phi(\mathbf{x};\mathcal{W}^*)-\mathbf{c}||^2.$$

ここで，$\mathcal{W}^*$ は学習済みモデルのパラメータである．

3.3 演奏ミス検出モデルの学習

演奏ミス検出モデルは，まず特徴表現を抽出するために事前学習を行い，学習済みの重みで初期化した Deep SAD ネットワークの学習を行う．

本研究では，事前学習に深層畳み込みオートエンコーダ（Deep Convolutional AutoEncoder：DCAE）を用いる．DCAE は，画像処理分野で多く用いられる畳み込みニューラルネットワーク（Convolutional Neural Network：CNN）をベースとしたオートエンコーダであり，メル周波数スペクトログラムを 1 チャネルの画像データとして扱うことで特徴表現を抽出することができる．本研究で用いる DCAE の構造を表 1, 表 2 に示す．ConvXxY, DeconvXxY はそれぞれカーネルサイズ $X\times Y$ の 2 次元畳み込み層，2 次元転置畳み込み層を，Linear は全結合層を表す．エンコーダではメル周波数スペクトログラム（128×32）を 4 層の 2 次元畳み込み層でダウンサンプリングし，4 層目の出力を平坦化して全結合層で 64 次元の特徴量に圧縮している．全ての畳み込み層では出力にバッチ正規化（Batch Normalization）[16] と ReLU 関数を順に適用している．デコーダでは，エンコーダの処理を逆の順序で行うことで入力された演奏音のメル周波数スペクトログラムを復元する．ただし，畳み込み層を全て転置畳み込み層に置き換えている．また，1 層目の全結合層の出力値はサイズを (32,8,4) に変換し，最終層の出力にはシグモイド関数のみを，それ以外の層ではバッチ正規化と ReLU 関数を順に適用している．

本研究で用いる Deep SAD のネットワークの構造は，事前学習で用いた DCAE のエンコーダ（表 1）と同等である．初期化は収束したエンコーダの重みで行い，初期化したネットワークに正常データ（ラベルなしデータ及びラベル付き正常データ）を入力した際の出力平均ベクトルを超球の中心 $\mathbf{c}$ に設定する．ラベルなしデータは $\mathbf{c}$ との MSE, ラベル付き正常データは $\mathbf{c}$ との MSE にハイパーパラメータ η を乗算したもの，ラベル付き異常データはその逆数をそれぞれ誤差関数として学習する．

両者共に最適化には Adam[17], 誤差関数には MSE を使用し, 超球崩壊を防ぐためバイアスは用いていない [12][13]. 推論時は，ネットワークのデータ出力値と中心 $\mathbf{c}$ の二乗誤差を異常スコアとして算出する．

4. 評価実験

4.1 実験条件

提案した演奏ミス検出モデルの評価実験として, 第一著者が実際にクラシックギターを演奏，録音（16bit/44100Hz）してデータセットを用意し，Deep SAD ネットワークにおいてハイパーパラメータ η を変えながら学習を行いモデルの評価を行った．学習時は Early Stopping を適用してエポック数を自動決定する．具体的には，検証用データの loss が DACE で 15 エポック，Deep SAD ネットワークで 30 エポック改善しなかった場合学習を終了する．バッチサイズは共に 64 とした．また，全ての η のパターンでネットワークの初期値と中心 $\mathbf{c}$ は同一である．提案手法の評価指標には，PR 曲線下側面積（PR-AUC）と F 値（F-measure）を用いた．

PR 曲線（Precision-Recall curve）とは，再現率（Recall）に対する適合率（Precision）を曲線としてプロットしたも

© 2022 Information Processing Society of Japan

表 3: ラベルなしデータ

string	0 - 12 fret	13 - 14 fret	15 - 19 fret
1	i:25, m:25	i:10, m:10	i:10, m:10
2	i:25, m:25	i:10, m:10	–
3	i:25, m:25	i:10, m:10	–
4	i:20, m:20, p:10	–	–
5	i:20, m:20, p:10	–	–
6	p:50	–	–

のである．適合率と再現率は以下の式で算出される．

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

TP，FP，FN はそれぞれ真陽性，偽陽性，偽陰性を指す．PR 曲線は陽性判定の精度にのみ着目するため，異常検知などの偏りが大きいデータに対する分類精度を適切に評価することができる．

F 値は適合率と再現率の調和平均であり，以下の式で算出される．

$$\text{F-measure} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

本研究では，各しきい値での F 値を算出し，最も F 値が高くなるものをそのモデルのしきい値に設定する．以下，このときの F 値を最大 F 値（Max F-measure）と表記する．

4.2 データセット

データセットには学習用データ，検証用データ，テストデータを用意した．学習用データにはラベルなしデータとラベル付きの正常データ，異常データが含まれる．

学習用データのうち，ラベルなしデータはフレットごとに表 3 の通りに演奏し，オンセット検出により 1 音ずつに分割して用意した．0 フレットは開放弦（何も押さえていない状態）を示す．i，m，p は弦をはじいた指を表し，それぞれ人差し指，中指，親指を指す．

検証用データ及びテストデータは実際に基礎練習を演奏し，「適切に演奏された音」と「演奏ミスをした音」をそれぞれラベル付けして用意した．基礎練習には Andrés Segovia[*4]の「Diatonic Major And Minor Scales」[18] を用いた．これは長調と短調の各 12 調，合計 24 調からなる音階練習法であり，世界中で親しまれている基礎練習である．指板を幅広く使うため，各フレットの音名の記憶，スムーズなポジション移動，押弦や弾弦の技術向上に大きく役立つ．「Diatonic Major And Minor Scales」の楽譜の例を図 2 に示す．この音階練習を各調で 5 回ほど演奏し，それぞれの調で 2〜5 音ほど演奏ミスをしている録音データを採用した．それらを 1 音ずつに分割してラベル付けを行い，

[*4] 1893-1987．クラシックコンサート楽器としてのギターの地位を確立し，「現代クラシックギター奏法の父」とされている．

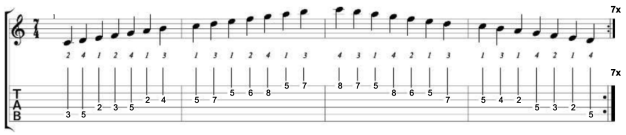


図 2: 「Diatonic Major And Minor Scales」[18] におけるハ長調（C メジャースケール）の楽譜

正常データ（「適切に演奏された音」）806 音，異常データ（「演奏ミスをした音」）105 音となった．これを正常データと異常データそれぞれでランダムにデータを 2 分割し，一方を検証用データ，もう一方をテストデータとした．

また，前述した基礎練習の録音データを利用して学習用のラベル付きデータを用意した．基礎練習の演奏データのうちテストデータ及び検証用データに採用しなかったものを全て 1 音ずつに分割し，ラベルなしデータの 1 割の数をランダムに抽出しラベル付けした．

以上より，学習用データがラベルなしデータ 4120 音，ラベル付き正常データ 295 音，ラベル付き異常データ 117 音の計 4532 音，検証用データが正常データ 403 音，異常データ 53 音の計 456 音，テストデータが正常データ 403 音，異常データ 52 音の計 455 音となった．また，演奏音は全て BPM120 の四分音符（1 音あたり 500ms）とし，アポヤンド奏法[*5]で演奏した．

4.3 実験結果と考察

実験結果を表 4 に示す．一番上の行はラベルなしデータのみを用いた教師なし学習の評価結果である．この結果から，ラベル付きデータを用いた半教師あり学習の有効性が明らかとなった．また，特に $\eta = 5.0$ 以上で PR-AUC，最大 F 値共に高くなっており，ラベル付きデータに重点を置くことでネットワークが効率良く学習できていると考えられる．η の値による変化をより詳しくみるため，各 η でのテストデータの「適切に演奏された音」の平均異常スコア（Normal MAS）と「演奏ミスをした音」の平均異常スコア（Miss MAS），及び最大 F 値のしきい値（Threshold）を表 5 に示す．η を大きくするほど，Miss MAS が大きくなっているのが分かる．これは，ラベル付き異常データが中心から遠くに射影されるほど，未知の「演奏ミスをした音」に対して敏感になるためと考えられる．今後，ハイパーパラメータのチューニングにより最適な η の値を探索していく必要がある．

最も最大 F 値の高い $\eta = 15.0$ のモデルについて，特徴空間の可視化を行った．64 次元の特徴量から 2 次元への変換には，t-SNE（t-Distributed Stochastic Neighbor Embedding）[19] という次元削減アルゴリズムを用いる．t-SNE は，高次元データの情報を保持したまま低次元データへ変

[*5] 弦を弾いた後，隣の弦に指が寄りかかる奏法を指す．弦を弾いた後に指が空中へ向かうアルアイレ奏法も存在する．

© 2022 Information Processing Society of Japan

表 4: 提案手法の評価実験結果

η	PR-AUC	Max F-measure	Precision	Recall
–	.712	.667	.727	.615
1.0	.914	.839	.783	.904
2.0	.961	906	.889	.923
5.0	**.979**	.940	.979	.904
10.0	**.979**	.950	1.00	.904
15.0	**.979**	**.951**	.980	.923
20.0	.977	.940	.979	.904

表 5: η の値による異常スコアの違い

η	Threshold	Normal MAS	Miss MAS
1.0	0.223	0.070	1.148
2.0	0.335	0.087	1.602
5.0	0.455	0.096	2.111
10.0	0.625	0.096	3.238
15.0	0.696	0.135	4.047
20.0	0.669	0.143	4.208

換する方法であり，元の空間での点同士の近さが，圧縮後の点同士の近さとできるだけ同じになるように次元を圧縮する．中心，真陰性（TN），真陽性（TP），偽陰性（FN），偽陽性（FP）にそれぞれ色分けし，t-SNE を適用した結果を図 3a に示す．定性的ではあるが，「適切に演奏された音」と「演奏ミスをした音」が超球により分離できていることが見て取れる．

また，テストデータの「演奏ミスをした音」を分類した場合の図を 3b に示す．分類は第一著者のみで行い，弾弦時にビビリの生じた音（buzz）20 音，押弦継続中にビビリの生じた音（mid-buzz）10 音，こもり音（muffled）19 音，ミュート音（mute）6 音となった．それぞれの例を最終ページの図 6，7，8，9，10 に示す．「弾弦時にビビリの生じた音」と「こもり音」の特徴量が集合しているのは，それぞれの特徴が一貫しているためと考えられる．一方で，「押弦継続中にビビリの生じた音」は特徴量が散らばっている．これは演奏ミスの度合いによってノイズの発生時刻や長さが異なるためと考えられる．また，「ミュート音」は「こもり音」の近くに多いが，これは調波音が急激に減衰するという類似点によるものと考えられる．これらのことから，提案した演奏ミス検出モデルが多様な「演奏ミスをした音」の特徴を適切に抽出して分離できることがわかり，様々な演奏ミスを検出できることを示した．

一方で偽陰性，偽陽性となった音も散見される．偽陰性となった「演奏ミスをした音」には，ビビリが 370ms 付近で発生しているものがあった（図 4）．これは，ポジション移動時などのフィンガーノイズ[*6]が 370ms 付近に鳴った音をラベル付き正常データに含めているのが原因と考えられる．偽陽性となった「適切に演奏された音」にも，押弦継続中にフィンガーノイズが鳴っている音があった（図 5）．具体的には，次に押さえるフレットのために開いた他の指につられ，押弦している指が動いてしまった際に発生したノイズである．フィンガーノイズが鳴った音を「適切に演奏された音」とするか「演奏ミスをした音」とするかは議論の余地がある．

[*6] 巻弦（4〜6 弦）上で指がフレットを移動する際に，指と弦が擦れて「キュッ」というような音が鳴るノイズ．「フレットノイズ」とも呼ばれる．

5. おわりに

本研究では，クラシックギターでの単音演奏において「適切に演奏された音」を正常データ，「演奏ミスをした音」を異常データとして，半教師あり深層異常検知手法の Deep SAD を用いた演奏ミスの自動検出手法を提案した．実験の結果，$\eta = 5.0$ 以上のモデルで，PR-AUC が .97 以上，F 値が .95 以上という高い精度で演奏ミスを検出することができた．また，特徴空間を二次元に圧縮して可視化することで，ネットワークが「演奏ミスをした音」の特徴を適切に抽出して異常検知を行っていることが分かり，演奏ミスの分類や大量のラベル付けの必要がない本手法の有効性を示した．

今後，本手法を用いた練習支援システムを作成し，ユーザ評価を行う必要がある．また，本研究でデータセットの録音に用いたクラシックギター及び演奏者は単一であるため，多様なクラシックギターや演奏者によるデータを集めて提案手法の汎用性を確認していきたい．さらに，本研究で扱ったのは BPM120 の四分音符のみであったが，入力を可変長にして様々な長さの音に対応できれば，より実用性の高いシステムになることが期待できる．

参考文献

[1] 元川洋一, 斎藤英雄. 拡張現実表示技術を用いたギターの演奏支援システム. 映像情報メディア学会誌, Vol. 61, No. 6, pp. 789–796, 2007.

[2] Markus Löchtefeld, Sven Gehring, Ralf Jung, and Antonio Krüger. guitar: supporting guitar learning through mobile projection. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pp. 1447–1452, 2011.

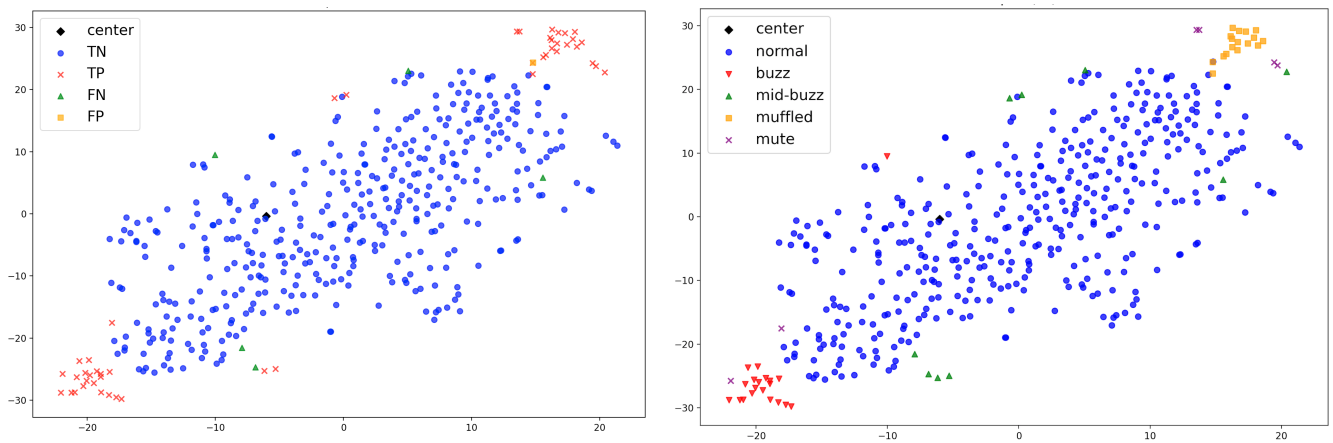
[3] 三浦駿, 安藤敏彦. Ar を用いたギター演奏学習支援システムの改良. 第 82 回全国大会講演論文集, pp. 529–530, feb 2020.

[4] 古庄優樹, 北原鉄朗. ギターの弦を正しく押さえるための初心者支援システム. Technical Report 35, 日本大学文理学部, 日本大学文理学部, mar 2021.

[5] Karola Marky, Andreas Weiß, Florian Müller, Martin Schmitz, Max Mühlhäuser, and Thomas Kosch. Let's frets! mastering guitar playing with capacitive sensing and visual guidance. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–4, 2021.

[6] Jakob Karolus, Hendrik Schuff, Thomas Kosch, Paweł W Wozniak, and Albrecht Schmidt. Emguitar: Assisting guitar playing with electromyography. In *Proceedings of the 2018 Designing Interactive Systems Conference*, pp.

© 2022 Information Processing Society of Japan



(a) テスト結果を色分けした場合　(b) 「演奏ミスをした音」を分類した場合

図 3: 特徴空間の可視化図

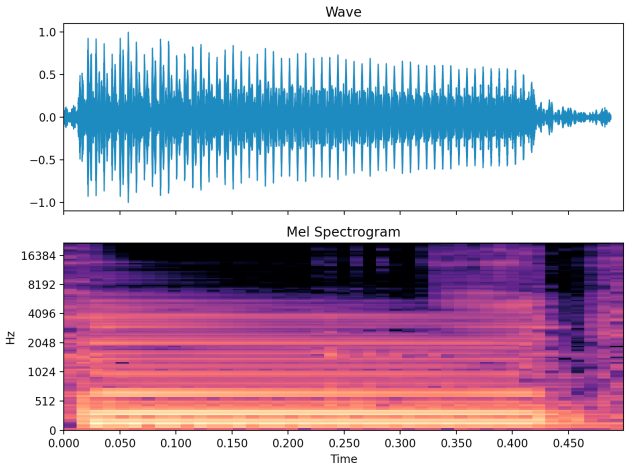


図 4: 偽陰性の「演奏ミスをした音」

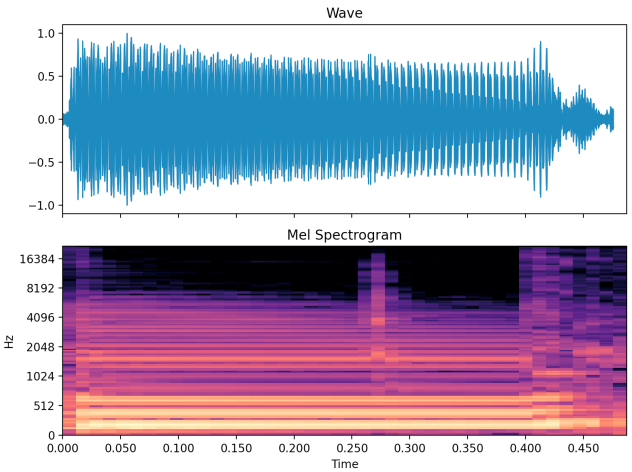


図 5: 偽陽性の「適切に演奏された音」

651–655, 2018.

[7] 坂牛和里, 甲藤二郎. カメラとセンサの利用による演奏音を判定するギター演奏支援システムの検討. Technical Report 4, 早稲田大学, 早稲田大学, feb 2017.

[8] 有賀竣哉, 後藤真孝, 矢谷浩司. Strummer：インタラクティブなギターコード練習システム. Technical Report 24, 東京大学大学院工学系研究科, 産業技術総合研究所, 東京大学大学院工学系研究科, feb 2017.

[9] 下尾波輝, 矢谷浩司. エレキギター演奏自動評価のための音響的特徴量の調査. Technical Report 3, 東京大学大学院工学系研究科, 東京大学大学院工学系研究科, nov 2017.

[10] 下尾波輝, 矢谷浩司. エレキギター演奏におけるミスの自動検出. 第 80 回全国大会講演論文集, pp. 131–132, mar 2018.

[11] Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. Deep learning for anomaly detection: A review. *ACM Computing Surveys (CSUR)*, Vol. 54, No. 2, pp. 1–38, 2021.

[12] Lukas Ruff, Robert A. Vandermeulen, Nico Görnitz, Lucas Deecke, Shoaib A. Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *Proceedings of the 35th International Conference on Machine Learning*, Vol. 80, pp. 4393–4402, 2018.

[13] Lukas Ruff, Robert A. Vandermeulen, Nico Görnitz, Alexander Binder, Emmanuel Müller, Klaus-Robert Müller, and Marius Kloft. Deep semi-supervised anomaly detection. In *International Conference on Learning Representations*, 2020.

[14] Brian McFee, Alexandros Metsai, Matt McVicar, Stefan Balke, Carl Thomé, Colin Raffel, Frank Zalkow, Ayoub Malek, Dana, Kyungyun Lee, Oriol Nieto, Dan Ellis, Jack Mason, Eric Battenberg, Scott Seyfarth, Ryuichi Yamamoto, viktorandreevichmorozov, Keunwoo Choi, Josh Moore, Rachel Bittner, Shunsuke Hidaka, Ziyao Wei, nullmightybofo, Adam Weiss, Darío Hereñú, Fabian-Robert Stöter, Pius Friesch, Matt Vollrath, Taewoon Kim, and Thassilo. librosa/librosa: 0.9.1, February 2022.

[15] David MJ Tax and Robert PW Duin. Support vector data description. *Machine learning*, Vol. 54, No. 1, pp. 45–66, 2004.

[16] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pp. 448–456. PMLR, 2015.

[17] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[18] Andres Segovia. *Diatonic major and minor scales*. Columbia Music Company, 1953.

© 2022 Information Processing Society of Japan

[19] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, Vol. 9, No. 11, 2008.

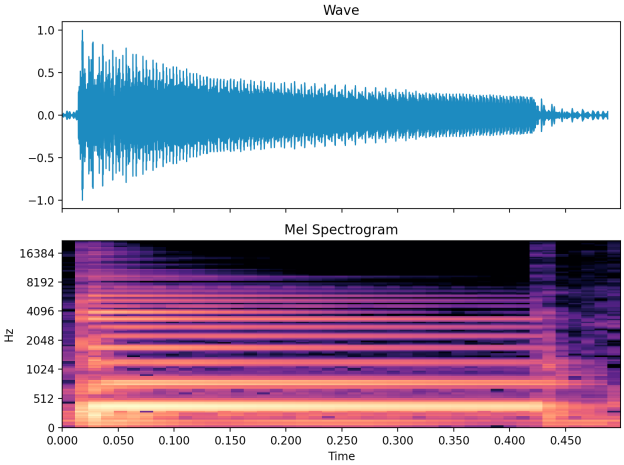


図 6: 適切に演奏された音（2 弦 5 フレット）

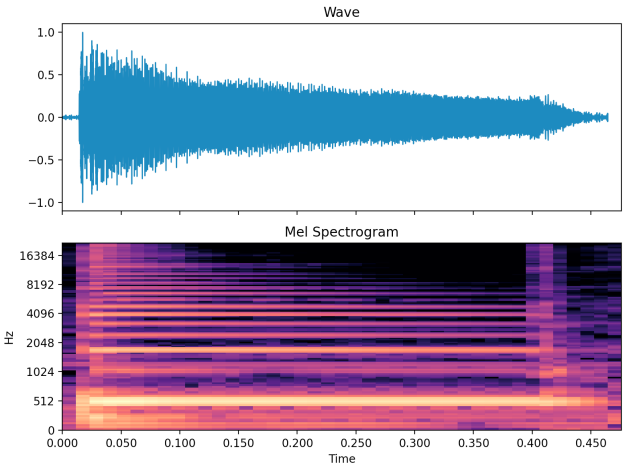


図 7: 弾弦時にビビリの生じた音（2 弦 10 フレット）．特に高周波数帯域でノイズが発生している．

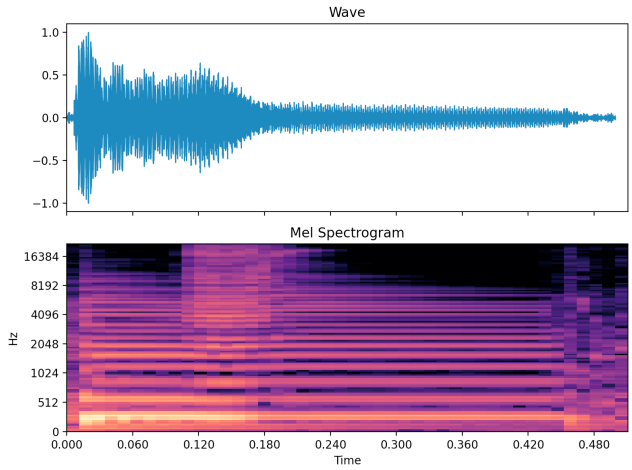


図 8: 押弦継続中にビビリの生じた音（4 弦 9 フレット）．発音時ではなく，押弦を継続させて調波音が鳴っている途中でノイズが発生している．

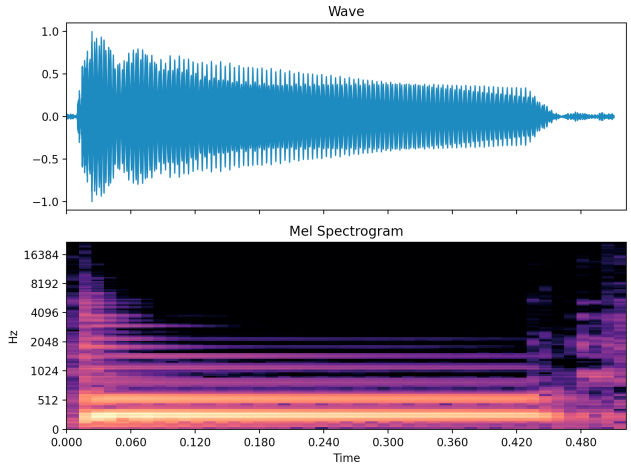


図 9: こもり音（4 弦 8 フレット）．基本周波数以外の高次の倍音成分の減衰が早く，高周波数帯域が削減されている．

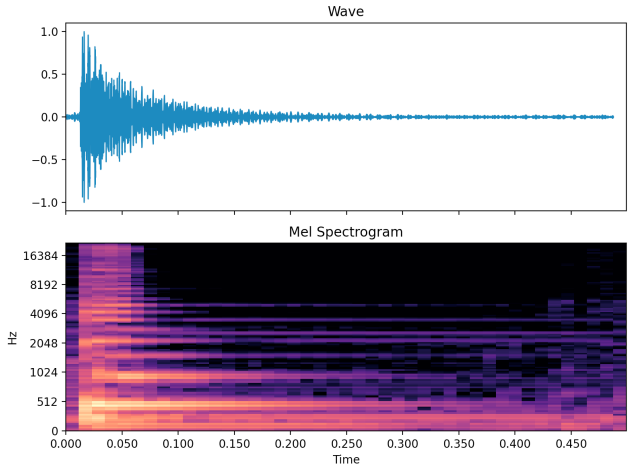


図 10: ミュート音（2 弦 9 フレット）．発音時に打楽器のような音が目立ち，調波音が急激に減衰する．

© 2022 Information Processing Society of Japan