Vol.2022-MUS-135 No.3
2022/9/14

バンド譜に対する無段階で難易度調整可能な深層ピアノ編曲

寺尾 萌夢[1,a)]　中村　栄太[1,2,b)]　吉井 和佳[1,3,c)]

概要：本稿では，深層ニューラルネットワーク (DNN) を用いて，ポピュラー音楽のバンド譜を，無段階で指定した難易度を持つピアノ譜に編曲する手法について述べる．バンド譜のピアノ編曲においては，元のバンド譜を上下に 1 オクターブシフトした拡張バンド譜から，ピアノ譜に用いる音符を選択するスコアリダクションアプローチが有効である．従来，バンド譜と初級あるいは上級のピアノ楽譜からなるペアデータを用いて，拡張バンド譜に対するマスクを推定する DNN を，難易度条件付きで教師あり学習する方法が提案されていた．しかし，学習時には初級かそうでないかの区別が見出され、推論時に中間的な難易度を指定しても、常に上級のピアノ譜が出力される問題があった．そこで，本研究では，低難度のピアノ譜は高難度のピアノ譜の部分集合であるという仮定し，DNN を用いて拡張バンド譜の各音符の基本重要度を推定し，難易度に応じた冪指数を用いて基本重要度の冪を計算した後，難易度に非依存の一定の閾値以上の音符を選択する方法を提案する．上級と初級に対応する冪指数は DNN と同時に最適化を行い，中級に対応する冪指数は線形補完によって求める．さらに，推論時における中級のピアノ譜の生成能力を強化するため，初級と上級のピアノ譜しか与えられていない学習時においても，中級のピアノ譜の生成・評価を可能にする手法を提案する．実験により，無段階の難易度調整における提案法の有効性を確認した．

1. はじめに

音楽情報処理分野では，楽曲のメロディなど重要な要素を保ったまま演奏楽器を変換する編曲タスクが研究されている．演奏楽器の変換を行う際は，曲のジャンルや対象となる楽器の特性などの専門知識が必要となるため，この過程を自動化することは依然として難しい課題である．これまで，ピアノ編曲 [1–6] や，ギター編曲 [7–9], オーケストラへの編曲 [10, 11] が行われている．本稿では，ポピュラー音楽の自動ピアノ編曲に取り組む．

自動ピアノ編曲では主に，入力楽譜から音符を選択するスコアリダクションの手法が取られている．これには入力楽譜の音符の音高，位置の情報を画像として捉え，ピアノ譜として重要な音符をカバーするような，マスク推定が用いられる．入力となるポピュラー音楽のバンド譜と，出力となるピアノ譜の間の音符の関係性の調査を行った結果 [12] より，バンド譜を上下 1 オクターブシフトすることで拡張されたバンド譜から，深層ニューラルネットワーク (DNN) で音符を選択する手法の妥当性が確認された．

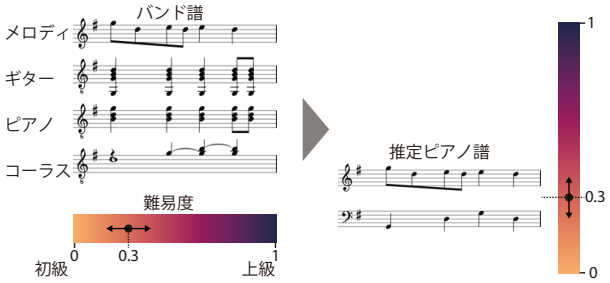

図 1　提案するピアノ編曲手法

しかし，ポピュラー音楽の難易度別ピアノ編曲を行う際，一般的に訓練データとして扱うピアノ譜の難易度は初級と上級の二つのみと，限られている．そこで二つの難易度を 0,1 の条件としてモデルに与え，その難易度条件によってモデルの振る舞いが変わるように学習をさせると，中級となるはずの，0.5 のような未知の条件を与えた場合でも上級の条件を与えた場合と同じ振る舞いをすることがわかった．しかし，難易度別ピアノ編曲に求められる要件として，ユーザーの技量に合わせたより柔軟なピアノ譜の難易度調整が挙げられる．

本研究ではまず，難易度条件下で学習した場合のピアノ譜の振る舞いを統計的に調査する．この結果に基づき，難

1　京都大学 大学院情報学研究科
2　京都大学 白眉センター
3　科学技術振興機構 戦略的創造研究推進事業（さきがけ）
a)　terao@sap.ist.i.kyoto-u.ac.jp
b)　enakamura@sap.ist.i.kyoto-u.ac.jp
c)　yoshii@i.kyoto-u.ac.jp

易度条件下で学習するという手法ではなく，バンド譜のメロディ・伴奏パートからそれらの音符がどれほど重要かを推定する DNN を学習する．この DNN から直接推定されたものを基本重要度とする．さらに目的の難易度によって，高難度の場合は重要度も高く，低難度の場合は重要度も低くなるように，冪指数を設定し，冪乗することで重要度を設定し直すという手法で実験を行う．最終的にこれを難易度に非依存の閾値で二値化することで，目的の難易度に沿ったピアノ譜を得ることができる．加えて，正解楽譜としては存在しない中級楽譜の生成を，推定時だけでなく，学習時にモデルに経験させるようランダムな難易度を与える実験と，推定されるピアノ譜が統計的に適切な性質を持つような誘導を行う実験を行う．実験結果は，提案法が柔軟な難易度の調整を可能にすることを示している．

2. 関連研究

ここでは，ピアノ編曲の手法を自動ピアノ編曲，ピアノ演奏の難易度の観点から総括する．

2.1 自動ピアノ編曲

ピアノ編曲の一つの方法として，スコアリダクションがある．これは，原曲の音を減らすことで目的の楽器の楽譜に編曲することである．Huang ら [3] は，このスコアリダクションを用いて，フレーズの特定，役割の割り当て，フレーズ選択により，自動的に目的の楽器への編曲を行う．反対に，高森らは [4] では，スコアリダクションではなく，メロディ，コード，リズム，音符の数という音楽要素を用いていくつかの伴奏テクスチャをあらかじめ定義し，原曲の特徴を反映したピアノ編曲を行う．また，彼らは [13] で，原曲の楽譜からではなく，音声からコード，メロディ，リズムを抽出し、ピアノ楽譜への変換を行っている．しかし，あらかじめ定義された伴奏は固定的であるため，Wang ら [14] は，音声からピアノ楽譜への変換を，より柔軟に行うように発展させた．

2.2 ピアノ演奏の難易度

ピアノ譜の難易度分析に着目した研究をいくつか紹介する．Chiu ら [15] は，楽譜全体に対し，いくつかの難易度へピアノ楽譜を分類する方法を提案している．Ramoneda ら [16] は，楽譜全体，または局所的にも，ピアノ運指アルゴリズムに基づいて難易度の分析を行っている．また，中村ら [5] は難易度を考慮し，スコアリダクション，オクターブシフトを用いて，統計的にピアノ編曲を行う．彼らは，従来研究で，演奏可能性の観点からピアノ編曲が行われていたことに対し，演奏可能性は演奏者の技術や，テンポに依存するという点に着目し，演奏難易度と音楽的忠実性に基づいたピアノ編曲を提案した．

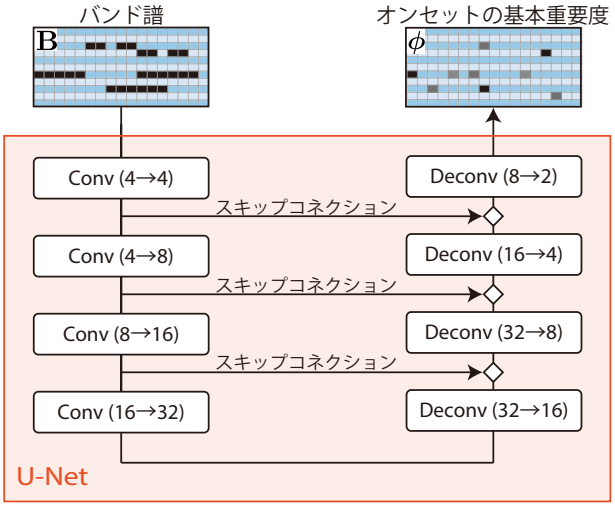

図 2 U-Net の構造．括弧内の数字はチャンネル数の変化を示している．また，スキップコネクションはチャンネルを連結することで入力に近い情報を取り込む．

3. 提案手法

本章では，バンドのポピュラー音楽の楽譜を，音符の重要度推定による滑らかな難易度調整に着目して，ピアノ譜に編曲する手法を提案する（図 1）．

3.1 問題設定

$\mathbf{B}$ をバンド譜，$\hat{\mathbf{P}}$ をピアノ譜とする．バンド譜は $\mathbf{B} \triangleq \{\mathbf{B}_\mathrm{A}, \mathbf{B}_\mathrm{M}\}$ と表され，$\mathbf{B}_\mathrm{A}$ は伴奏パートを，$\mathbf{B}_\mathrm{M}$ はメロディパートを表している．ピアノ譜は $\hat{\mathbf{P}} \triangleq \{\hat{\mathbf{P}}_\mathrm{L}, \hat{\mathbf{P}}_\mathrm{R}\}$ と表され，$\hat{\mathbf{P}}_\mathrm{L}$ は左手パートを，$\hat{\mathbf{P}}_\mathrm{R}$ は右手パートを表している．ここで，各パートは $*^\mathrm{o}$ と $*^\mathrm{p}$ で表現される，オンセットと音符を表す行列から成る．P をピッチ数（$P = 128$），N を楽譜の長さとすると，これらの行列の大きさは $P \times N$ である．本稿では各小節を時間方向に 16 個にわけた 1 単位をテイタムと表す．また，一つの曲の長さはを 12 小節，つまり，$N = 16 \times 12 = 192$ テイタムで構成する．従って，バンド譜は $\mathbf{B}_\mathrm{A} \triangleq \{\mathbf{B}^\mathrm{o}_\mathrm{A}, \mathbf{B}^\mathrm{p}_\mathrm{A}\}$ と $\mathbf{B}_\mathrm{M} \triangleq \{\mathbf{B}^\mathrm{o}_\mathrm{M}, \mathbf{B}^\mathrm{p}_\mathrm{M}\}$ と表され，ピアノ譜は $\hat{\mathbf{P}}_\mathrm{L} \triangleq \{\hat{\mathbf{P}}^\mathrm{o}_\mathrm{L}, \hat{\mathbf{P}}^\mathrm{p}_\mathrm{L}\}$ and $\hat{\mathbf{P}}_\mathrm{R} \triangleq \{\hat{\mathbf{P}}^\mathrm{o}_\mathrm{R}, \hat{\mathbf{P}}^\mathrm{p}_\mathrm{R}\}$ と表される．例えば，$\mathbf{B}^\mathrm{o}_\mathrm{M}(p, n) = 1$ はピッチ p でテイタム n でオンセットが存在することを示している．また，$\mathbf{B}^\mathrm{p}_\mathrm{M}(p, n) = 1$ はピッチ p でテイタム n でそのピッチの音符が鳴っていることを示している．両手のパートをまとめて表す際には $\mathrm{h} \in \{\mathrm{L}, \mathrm{R}\}$ を用いる．また，すべてのデータに対して初級上級の両方の正解楽譜があるわけではないことを考慮し，$\hat{\mathbf{P}}_{\mathrm{h,elm}}$, $\hat{\mathbf{P}}_{\mathrm{h,adv}}$ のように, $*_{\mathrm{h,elm}}$ は正解楽譜として初級ピアノ譜が対応しているもの，$*_{\mathrm{h,adv}}$ は正解楽譜として上級ピアノ譜が対応しているものとして表す．

本実験の目的は $\mathbf{B}$ を $\hat{\mathbf{P}}$ に変換することであるが，具体的な流れとしてまず，直接ピアノ譜 $\hat{\mathbf{P}}$ を求める代わりに

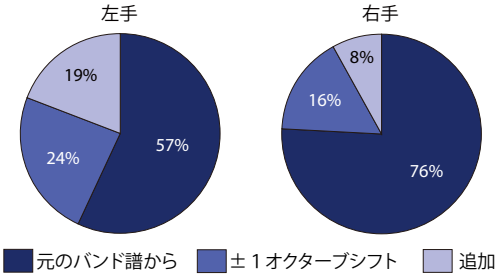

図 3 ピアノ譜とバンド譜の音符の関係性.

オンセットのマスク行列 $\boldsymbol{\phi} \triangleq \{\boldsymbol{\phi}_{\mathrm{L}}, \boldsymbol{\phi}_{\mathrm{R}}\}$ を求める．これは U-Net [17]（図 2）を用いて求めたソフトマスクの行列である．$\boldsymbol{\phi}_{\mathrm{L}} \in [0,1]^{P\times N}$ は左手パート，$\boldsymbol{\phi}_{\mathrm{R}} \in [0,1]^{P\times N}$ は右手パートをそれぞれ表している．この値は $[0,1]$ の範囲に収まるようにシグモイド関数に通して出力される．このマスク行列から，拡張バンド譜（節 3.2）の各音符から音符を選ぶための重要度を計算する．オンセットの重要度による音符の選択後，音価として拡張バンド譜から対応する音符の音価を採用する．

3.2 拡張バンド譜の計算

初めにバンド譜とピアノ譜のペアデータに対して，ピアノ譜の音符がバンド譜のどのパートから使われている音符なのかという統計的な調査を行った．図 3 が結果を表している．左の図がピアノ譜の左手パートの結果，右の図が右手パートの結果を表している．この図によると，右手の約 76%，左手の約 57%の音符は元のバンド譜の音符がそのまま使われていることがわかった．これより，元のバンド譜から音符を選択するだけでピアノ譜を構成するのは難しいと考えられる．一方で，右手の約 92%，左手の約 81%の音符はバンド譜を上下 1 オクターブシフトした拡張バンド譜の音符が使われていることがわかった．そこで，拡張バンド譜から音符を選び，不必要な音符を消すことでピアノ譜が得られるという仮定が立てられる．

$\mathbf{Z}^{\mathrm{o}} \in \{0,1\}^{P\times N}$ を拡張バンド譜のオンセット行列とする．$\mathbf{Z}^{\mathrm{o}}$ は次のように計算される．

$$\mathbf{Z}^{\mathrm{o}}(p,n) = \max_{j\in\{-12,0,12\}} (\mathbf{B}^{\mathrm{o}}_{\mathrm{A}}(p+j,n), \mathbf{B}^{\mathrm{o}}_{\mathrm{M}}(p+j,n)) \quad (1)$$

ここで $j \in \{-12,0,12\}$ はオクターブシフトを表している．

3.3 音符の重要度推定

本稿では拡張バンド譜の音符の重要度の推定に U-Net [17] を用いた．まず，$\boldsymbol{\phi}_{\mathrm{h}}$ に，難易度毎に違う冪指数の値を用いて冪乗を適用する．これを $\bar{\boldsymbol{\phi}}_{\mathrm{h}}$ とする．難易度別の冪指数の値を α_{elm}，α_{adv} とし，それぞれ，初級，上級に対する冪指数の値を表している．すると，$\bar{\boldsymbol{\phi}}_{\mathrm{h}}$ は次のように求められる．

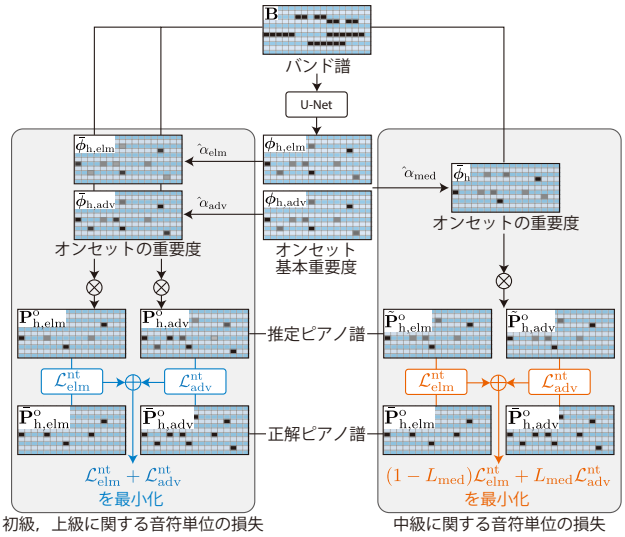

図 4 音符単位の損失の 2 種類の計算法．左側は 3.4.1 の初級上級の損失を用いた手法，右側は 3.4.2 の中級の損失を用いた手法を表す.

$$\bar{\boldsymbol{\phi}}_{\mathrm{h},*} = \boldsymbol{\phi}_{\mathrm{h},*}^{\alpha_*} \ (* \in \{\mathrm{elm}, \mathrm{adv}\}) \quad (2)$$

$\boldsymbol{\phi}_{\mathrm{h}} \in [0,1]$ であることを考慮すると，冪指数の値が大きくなれば $\bar{\boldsymbol{\phi}}_{\mathrm{h}}$ の重要度は全体的に低くなり，冪指数の値が小さくなれば $\bar{\boldsymbol{\phi}}_{\mathrm{h}}$ の重要度は全体的に高くなる．高難度の方が重要度が高くなってほしいので，$\alpha_{\mathrm{elm}} \geq \alpha_{\mathrm{adv}}$ である．

次に節 3.2 で言及した拡張バンド譜 $\mathbf{Z}^{\mathrm{o}}$ と $\bar{\boldsymbol{\phi}}$ の要素積を計算することで，拡張バンド譜に含まれる音符のみを出力することができる．これを，$\mathbf{P}^{\mathrm{o}} \triangleq \{\mathbf{P}^{\mathrm{o}}_{\mathrm{L}}, \mathbf{P}^{\mathrm{o}}_{\mathrm{R}}\}$ とする．$\mathbf{P}^{\mathrm{o}}_{\mathrm{L}}, \mathbf{P}^{\mathrm{o}}_{\mathrm{R}} \in [0,1]^{P\times N}$ はそれぞれ左手パート，右手パートを表している．

$$\mathbf{P}^{\mathrm{o}}_{\mathrm{L}} = \bar{\boldsymbol{\phi}}_{\mathrm{L}} \odot \mathbf{Z}^{\mathrm{o}}, \quad \mathbf{P}^{\mathrm{o}}_{\mathrm{R}} = \bar{\boldsymbol{\phi}}_{\mathrm{R}} \odot \mathbf{Z}^{\mathrm{o}} \quad (3)$$

$\odot$ は要素積を表している．この操作により，$\mathbf{P}^{\mathrm{o}}$ は拡張バンド譜に含まれる音符のみとなっている．最終的なピアノ譜として，$\hat{\mathbf{P}}^{\mathrm{o}} \triangleq \{\hat{\mathbf{P}}^{\mathrm{o}}_{\mathrm{L}}, \hat{\mathbf{P}}^{\mathrm{o}}_{\mathrm{R}}\}$ は $\mathbf{P}^{\mathrm{o}} \triangleq \{\mathbf{P}^{\mathrm{o}}_{\mathrm{L}}, \mathbf{P}^{\mathrm{o}}_{\mathrm{R}}\}$ を閾値処理により 2 値化することで得られる．また，音符を表す行列である $\hat{\mathbf{P}}^{\mathrm{p}} \triangleq \{\hat{\mathbf{P}}^{\mathrm{p}}_{\mathrm{L}}, \hat{\mathbf{P}}^{\mathrm{p}}_{\mathrm{R}}\}$ はバンド譜の音符を表す行列 $\mathbf{B}^{\mathrm{p}} \triangleq \{\mathbf{B}^{\mathrm{p}}_{\mathrm{L}}, \mathbf{B}^{\mathrm{p}}_{\mathrm{R}}\}$ と $\hat{\mathbf{P}}^{\mathrm{o}}$ より決定される（節 3.5）．

3.4 音符単位の損失

節 3.3 で求めた重要度を用いて，音符単位の損失を計算する．本稿では，2 種類の損失を提案する（図 4）．一つは，対応する正解ピアノ譜の難易度に合わせた初級または，上級の損失を計算し，ミニバッチ内で足し合わせる方法である（節 3.4.1）．もう一つは，α の値をランダムに変化させ，中級のピアノ譜を出力し，初級との損失，上級との損失を加重平均で計算する方法である（節 3.4.2）．

3.4.1 初級，上級に関する音符単位の損失

まず一つ目の，対応する正解ピアノ譜の難易度に合わせた初級または，上級の損失を計算し，ミニバッチ内で足し

合わせる方法について説明する．音符レベルの損失 $\mathcal{L}^{\mathrm{nt}}$ は次のクロスエントロピーによって定義される．

$$\mathcal{L}^{\mathrm{nt}} = -\sum_{h\in\{\mathrm{L,R}\}}\sum_{p=1}^{P}\sum_{n=1}^{N}\Big(w\cdot\bar{\mathbf{P}}^{\mathrm{o}}_{\mathrm{h}}(p,n)\log\mathbf{P}^{\mathrm{o}}_{\mathrm{h}}(p,n) + \big(1-\bar{\mathbf{P}}^{\mathrm{o}}_{\mathrm{h}}(p,n)\big)\log\big(1-\mathbf{P}^{\mathrm{o}}_{\mathrm{h}}(p,n)\big)\Big) \tag{4}$$

ここで，$\bar{\mathbf{P}}^{\mathrm{o}} \triangleq \{\bar{\mathbf{P}}^{\mathrm{o}}_{\mathrm{L}}, \bar{\mathbf{P}}^{\mathrm{o}}_{\mathrm{R}}\}$ は正解ピアノ譜のオンセット行列を表しており，$\bar{\mathbf{P}}^{\mathrm{o}}_{\mathrm{L}}, \bar{\mathbf{P}}^{\mathrm{o}}_{\mathrm{R}} \in \{0,1\}^{P\times N}$ はそれぞれ左手パート，右手パートである．また，$w \geq 0$ は正例に対する重みであり，オンセットとオンセットでない部分の数の格差を補正する役割がある．ここでは，$w = 4$ とする．

3.4.2 中級に関する音符単位の損失

節 3.3 では，初級，上級の 2 種類の難易度しか扱えなかったが，学習時にも中級のピアノ譜を扱うことを考える．$L \in [0,1]$ を対応する正解ピアノ譜の難易度とする．難しくなるほど L は大きくなり，初級は $L = 0$，上級は $L = 1$ である．中級を評価するために，ミニバッチ毎に DNN にランダムな難易度ラベル L_{med} を与える．この L_{med} に対応する α_{med} を次のように計算する．

$$\alpha_{\mathrm{med}} = (1-L_{\mathrm{med}})T_{\mathrm{elm}} + L_{\mathrm{med}}\,T_{\mathrm{adv}} \tag{5}$$

基本重要度を，この α_{med} で冪乗し，拡張バンド譜と掛け合わせることで，難易度ラベル L_{med} に対応する重要度 $\tilde{\mathbf{P}}^{\mathrm{o}}_{\mathrm{h}}$ を得る．ここで注意すべきことは，$\tilde{\mathbf{P}}^{\mathrm{o}}_{\mathrm{h,elm}}$，$\tilde{\mathbf{P}}^{\mathrm{o}}_{\mathrm{h,adv}}$ はそれぞれ，対応する正解ピアノ譜の種類を表しているだけであり，実際には中級となる難易度ラベル L_{med} に対応するピアノ譜だということである．

次に，$\mathcal{L}^{\mathrm{nt}}_{\mathrm{elm}}$ を初級ピアノ譜に対する音符レベルの損失とする．また，$\mathcal{L}^{\mathrm{nt}}_{\mathrm{adv}}$ は上級ピアノ譜に対する音符レベルの損失とする．

$$\mathcal{L}^{\mathrm{nt}}_{*} = -\sum_{h\in\{\mathrm{L,R}\}}\sum_{p=1}^{P}\sum_{n=1}^{N}\Big(w\cdot\bar{\mathbf{P}}^{\mathrm{o}}_{\mathrm{h},*}(p,n)\log\tilde{\mathbf{P}}^{\mathrm{o}}_{\mathrm{h},*}(p,n) + \big(1-\bar{\mathbf{P}}^{\mathrm{o}}_{\mathrm{h},*}(p,n)\big)\log\big(1-\tilde{\mathbf{P}}^{\mathrm{o}}_{\mathrm{h},*}(p,n)\big)\Big) \quad (*\in\{\mathrm{elm,adv}\}) \tag{6}$$

曲によって，対応する難易度の正解ピアノ譜 $\bar{\mathbf{P}}^{\mathrm{o}}_{\mathrm{h}}$ が存在しない場合には 0 となる．そして，$\mathcal{L}^{\mathrm{nt}}_{\mathrm{med}}$ は中級の損失とすると，次のように計算される．

$$\mathcal{L}^{\mathrm{nt}}_{\mathrm{med}} = (1-L_{\mathrm{med}})\mathcal{L}^{\mathrm{nt}}_{\mathrm{elm}} + L_{\mathrm{med}}\mathcal{L}^{\mathrm{nt}}_{\mathrm{adv}} \tag{7}$$

3.5 音価の求め方

ここでは，オンセット行列 $\hat{\mathbf{P}}^{\mathrm{o}}$ と拡張バンド譜 $\mathbf{Z}$ から音符を表す行列 $\hat{\mathbf{P}}^{\mathrm{p}} = \{\hat{\mathbf{P}}^{\mathrm{p}}_{\mathrm{L}}, \hat{\mathbf{P}}^{\mathrm{p}}_{\mathrm{R}}\}$ を求める（図 5）．

まず，拡張バンド譜のオンセット行列 $\mathbf{Z}^{\mathrm{o}}$ からピッチとオンセットのペアの系列 $\{(p_a, n_a)\}_{a=1}^{A}$ を計算する．A は拡張バンド譜内の音符の数であり，p_a，n_a はそれぞれ a 番目の音符のピッチと，オンセットのテイタムを表している．$\mathbf{A} \triangleq \{p_a\}_{a=1}^{A}$ を拡張バンド譜のピッチの系列とする．$\tilde{\mathbf{Z}} \triangleq \{\tilde{\mathbf{Z}}^{\mathrm{o}}, \tilde{\mathbf{Z}}^{\mathrm{p}}\}$ をオンセットと音符を表す行列とする．$\tilde{\mathbf{Z}}^{\mathrm{o}}, \tilde{\mathbf{Z}}^{\mathrm{p}} \in \{0,1\}^{A\times N}$ であり，$\mathbf{Z}^{\mathrm{o}}$ によって決定される．$n_a = n$ の場合に $\tilde{\mathbf{Z}}^{\mathrm{o}}(a,n) = 1$ であり，a 番目の音符がテイタム n で鳴っている場合に $\tilde{\mathbf{Z}}^{\mathrm{p}}(a,n) = 1$ である．

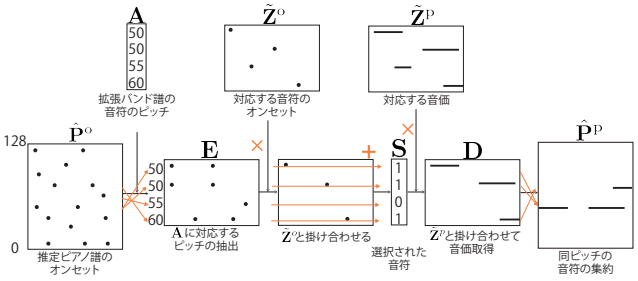

図 5 K の計算方法

$\mathbf{E} \triangleq \{\mathbf{E}_{\mathrm{L}}, \mathbf{E}_{\mathrm{R}}\}$ を $\hat{\mathbf{P}}^{\mathrm{o}}$ から $\mathbf{A}$ を使って各音符ごとにオンセット行列を抽出した行列とする．$\mathbf{E}_{\mathrm{L}}, \mathbf{E}_{\mathrm{R}} \in \{0,1\}^{A\times N}$ はそれぞれ左手パート，右手パートを表す．$\mathbf{E}_{\mathrm{h}}$ の a 番目の行は $\hat{\mathbf{P}}^{\mathrm{o}}_{\mathrm{h}}$ の p_a 番目の行である．つまり，$\mathbf{E}_{\mathrm{h}}(a) = \hat{\mathbf{P}}^{\mathrm{o}}_{\mathrm{h}}(p_a)$ となる．

$\mathbf{E}_{\mathrm{h}}$ と $\tilde{\mathbf{Z}}^{\mathrm{o}}$ の要素積を時間軸方向に足したものを $\mathbf{S} \triangleq \{\mathbf{S}_{\mathrm{L}}, \mathbf{S}_{\mathrm{R}}\} \in \{0,1\}^{2\times A}$ とする．これは，選択された音符を表しており，$\mathbf{S}_{\mathrm{L}}(a) = 1$ または $\mathbf{S}_{\mathrm{R}}(a) = 1$ はそれぞれのパートに a 番目の音符が存在していることを示す．音価を表す行列 $\mathbf{D} \triangleq \{\mathbf{D}_{\mathrm{L}}, \mathbf{D}_{\mathrm{R}}\} \in \{0,1\}^{2\times A\times N}$ は，$\mathbf{S}$ と $\tilde{\mathbf{Z}}^{\mathrm{p}}$ をかけ合わせることで得られる．最終的に $\hat{\mathbf{P}}^{\mathrm{p}} = \{\hat{\mathbf{P}}^{\mathrm{p}}_{\mathrm{L}}, \hat{\mathbf{P}}^{\mathrm{p}}_{\mathrm{R}}\} \in \{0,1\}^{2\times P\times N}$ は，$\mathbf{D}$ をノートナンバー順に並べることで得られる．

3.6 統計分布

U-Net の正則化に，同時発音数と音符密度の 2 つの音符統計量の分布を用いた．この分布は各ミニバッチ毎に微分可能な処理を行って計算する．$\mathbf{C}^{\mathrm{lv}}_{\mathrm{h}}(n)$ は推定ピアノ譜の n 番目のテイタムの同時発音数，$\mathbf{C}^{\mathrm{ds}}_{\mathrm{h}}(m)$ は推定ピアノ譜の m 番目の小節のオンセットの数，を表し，次のように計算される．

$$\mathbf{C}^{\mathrm{lv}}_{\mathrm{h}}(n) = \sum_{p=1}^{P}\hat{\mathbf{P}}^{\mathrm{p}}_{\mathrm{h}}(p,n), \mathbf{C}^{\mathrm{ds}}_{\mathrm{h}}(m) = \sum_{p=1}^{P}\hat{\mathbf{P}}^{\mathrm{o}}_{\mathrm{h}}(p,m) \tag{8}$$

これらの分布と統計量の損失の求め方を示す．I^{lv} は同時発音数の最大値，I^{ds} は音符密度の最大値とする．また，ここで同時発音数，音符密度を共通の式で表すため，$\mathrm{d} \in \mathrm{lv, ds}$ とする．各値 i $(0 \leq i \leq I^{\mathrm{d}})$ についてまず，$\mathbf{G}^{\mathrm{lv}}_{\mathrm{h}}(i)$，$\mathbf{G}^{\mathrm{ds}}_{\mathrm{h}}(i)$ を次のように計算する．

$$\mathbf{G}^{\mathrm{d}}_{\mathrm{h}}(i) = \sum_{n=1}^{N}\mathrm{ReLU}\left(-\mathbf{C}^{\mathrm{d}}_{\mathrm{h}}(n) + i\right) \tag{9}$$

ここで，$\mathrm{ReLU}(-\mathbf{C}^{\mathrm{d}}_{\mathrm{h}}(n) + i)$ は $\mathbf{C}^{\mathrm{d}}_{\mathrm{h}}(n) < i$ の時，正の値を

取り，その他の場合は 0 となる．$\mathbf{G}_{\mathrm{h}}^{\mathrm{d}}(i^{\mathrm{d}})$ は，同時発音数が j であるテイタムの頻度を表す $\mathbf{F}_{h}^{\mathrm{lv}}(j)$，もしくは，音符密度が j である小節の頻度を表す $\mathbf{F}_{h}^{\mathrm{ds}}(j)$ を用いて，次のように表現することもできる．

$$\mathbf{G}_{\mathrm{h}}^{\mathrm{d}}(i) = \sum_{j=0}^{i-1} (\mathbf{F}_{\mathrm{h}}^{\mathrm{d}}(j) \times (i-j)) \tag{10}$$

ここで，

従って，$\mathbf{F}_{h}^{\mathrm{d}} \in [0,1]^{I^{\mathrm{d}}+1}$ は再帰的に次のように求められる．

$$\mathbf{F}_{\mathrm{h}}^{\mathrm{d}}(0) = \mathbf{G}_{\mathrm{h}}^{\mathrm{d}}(0) \tag{11}$$

$$\mathbf{F}_{\mathrm{h}}^{\mathrm{d}}(i) = \mathbf{G}_{\mathrm{h}}^{\mathrm{d}}(i+1) - 2\mathbf{G}_{\mathrm{h}}^{\mathrm{d}}(i) + \mathbf{G}_{\mathrm{h}}^{\mathrm{d}}(i-1) \quad (i \geq 1) \tag{12}$$

同時発音数の分布 $\mathbf{Q}_{\mathrm{h}}^{\mathrm{d}} \in [0,1]^{I^{\mathrm{d}}+1}$ はこの頻度 $\{\mathbf{F}_{\mathrm{h}}^{\mathrm{d}}(i)\}_{i=0}^{I^{\mathrm{d}}}$ を正規化することで求められる．

3.6.1 初級，上級に関する統計量の損失

統計量損失 $\mathcal{L}^{\mathrm{lv}}$，$\mathcal{L}^{\mathrm{ds}}$ は JS ダイバージェンス $\mathcal{D}_{\mathrm{JS}}$ を用いて，正解分布 $\bar{\mathbf{Q}}_{\mathrm{h}}^{\mathrm{d}}$ と推定分布 $\mathbf{Q}_{\mathrm{h}}^{\mathrm{d}}$ の距離を測る．

$$\mathcal{L}^{\mathrm{d}} = \sum_{\mathrm{h}\in\{\mathrm{L,R}\}} \mathcal{D}_{\mathrm{JS}}(\bar{\mathbf{Q}}_{\mathrm{h}}^{\mathrm{d}} \parallel \mathbf{Q}_{\mathrm{h}}^{\mathrm{d}}) \tag{13}$$

これらを踏まえて音符単位の損失に統計量の損失を加えた $\mathcal{L}$ は β^{lv}，β^{ds} をそれぞれの重みとして，次の式で示される．

$$\mathcal{L} = \mathcal{L}^{\mathrm{nt}} + \beta^{\mathrm{lv}}\mathcal{L}^{\mathrm{lv}} + \beta^{\mathrm{ds}}\mathcal{L}^{\mathrm{ds}} \tag{14}$$

これを最小化するように U-Net を学習する．

4. 評価実験

本章では，提案手法の有効性を評価するために行った実験について報告する．まず，ポピュラー音楽を用いて従来手法である難易度条件付きで学習を行った場合の同時発音数と音符密度の分布について調査を行う．次に，冪乗を用いた提案手法の性能を検証する．

4.1 実験条件

実験には，バンド譜とピアノ譜の MIDI データ 184 ペア（初級 85 ペア，上級 99 ペア）を使用した．138 ペアは学習データとして，46 ペアはテストデータとして用いた．楽譜の拍子は 4/4 拍子として扱い，データ拡張として，1 から 11 までオクターブシフトを行った．U-Net は 4 層の畳み込み層と 4 層の逆畳み込み層で構成される．全ての層でバッチ正則化を行った．カーネルサイズは 4，ストライドは 2，パディングは 1 に設定した．過学習を防ぐため，全ての逆畳み込み層でドロップアウト（p = 0.5）を適用した．ネットワークの最適化には Adam（lr = 10^{-4}）を用いた．閾値は，学習時には 0.5 を用い，推論時には，検証

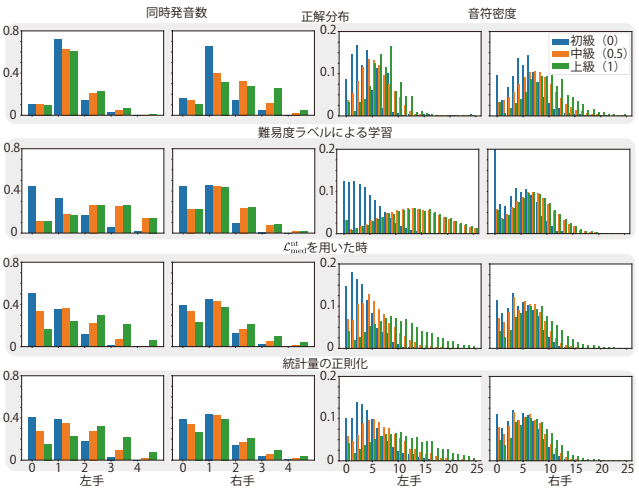

図 6 統計量の分布．左二列が同時発音数，右二列が音符密度を表している．一段目は正解ヒストグラム，二段目は難易度条件下で学習した場合のヒストグラム，三段目は $\mathcal{L}_{\mathrm{med}}^{\mathrm{nt}}$ で学習を行った場合のヒストグラム，四段目は統計量に関する正則化を含めた場合である．

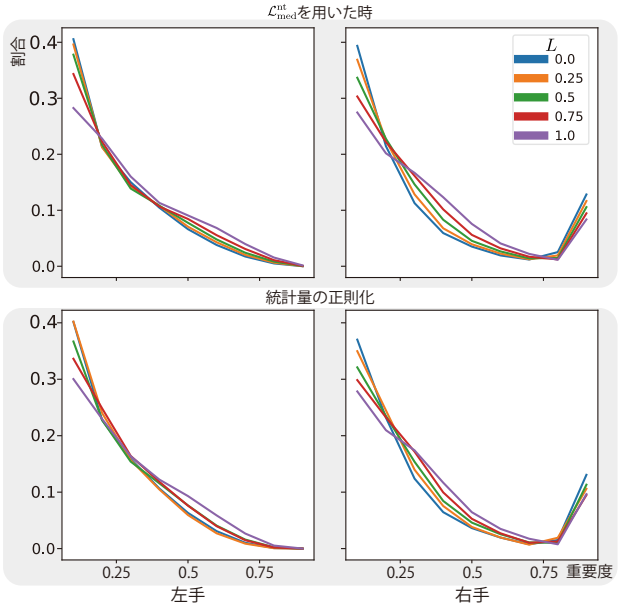

図 7 難易度ラベル L を変化させた時の重要度のグラフ．上が，$\mathcal{L}_{\mathrm{med}}^{\mathrm{nt}}$ で学習を行った場合，下が，統計量に関する正則化を含めた場合である．

データに対して左右のパートでそれぞれ $\mathcal{F}$ を最大化するものを使用した．各統計量に関する損失の重みパラメータは，$\beta^{\mathrm{lv}} = 5$，$\beta^{\mathrm{ds}} = 0$ のパターン，$\beta^{\mathrm{lv}} = 0$，$\beta^{\mathrm{ds}} = 5$ のパターン，$\beta^{\mathrm{lv}} = \beta^{\mathrm{ds}} = 2.5$ のパターンを試した．はじめは $\alpha_{\mathrm{elm}} = \alpha_{\mathrm{adv}} = 1$ に固定して学習を行い，100 エポックの学習後に 5 エポック毎に α_{elm} と α_{adv} の更新を行う．評価尺度には $\mathcal{F}$ 値を利用し，推定結果の統計的妥当性を評価するため $\mathcal{L}^{\mathrm{lv}}$，$\mathcal{L}^{\mathrm{ds}}$ を計算した．

表 1 実験結果

使用する損失				冪指数の値		$\mathcal{F}$		JS Div. 同時発音数 (×10)			JS Div. 音符密度 (×10)		
$\mathcal{L}^{\mathrm{nt}}$	$\mathcal{L}^{\mathrm{nt}}_{\mathrm{med}}$	$\mathcal{L}^{\mathrm{lv}}$	$\mathcal{L}^{\mathrm{ds}}$	α_{elm}	α_{adv}	左手	右手	初級	中級	上級	初級	中級	上級
✓				1.34	0.752	**27.3**	59.9	1.22	0.947	1.61	0.504	0.709	1.86
	✓			0.990	0.544	26.6	59.6	1.41	0.925	**1.32**	0.473	**0.538**	**1.24**
✓		✓		1.13	0.754	26.6	**60.2**	1.14	0.929	1.68	0.498	0.756	2.09
✓			✓	1.07	0.631	**27.3**	59.7	1.15	0.888	1.38	**0.436**	0.570	1.46
✓		✓	✓	1.04	0.691	26.3	59.7	**1.09**	**0.861**	1.44	0.575	0.723	1.77

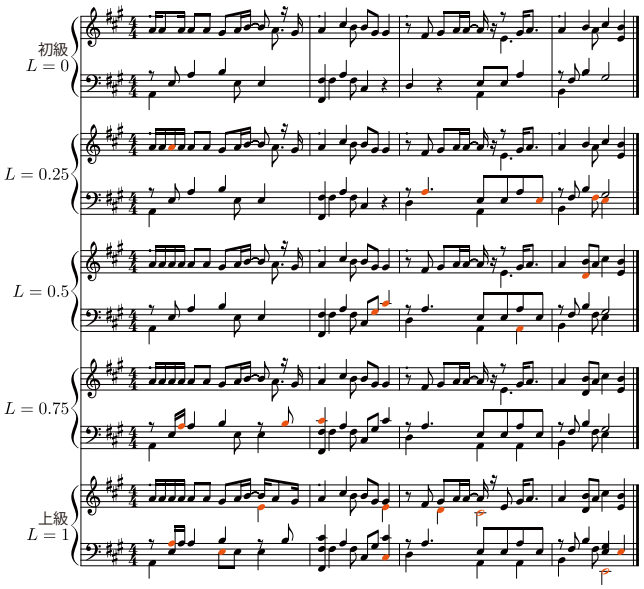

図 8 中級に関する損失を用いて学習を行った場合での楽譜の例．下に向かって，初級から上級へ，変化している．オレンジ色の音符は一つ前の楽譜と比べて増えた音符を表している．

4.2 実験結果

表 1 は U-Net を $\mathcal{L}^{\mathrm{nt}}$，$\mathcal{L}^{\mathrm{nt}}_{\mathrm{med}}$ でそれぞれ学習した場合と，$\mathcal{L}^{\mathrm{nt}}$ に統計量に関する正則化を一つずつと，両方加えた場合での最適化された冪指数の値と，$calF$ と JS ダイバージェンスによる評価値を示す．まず，中級に関する音符単位の損失 $\mathcal{L}^{\mathrm{nt}}_{\mathrm{med}}$ で学習を行った場合は，$\mathcal{L}^{\mathrm{nt}}$ で学習を行った場合に比べ，統計量の分布に対して効果があることが分かる．また，統計量による正則化は，同時発音数の JS ダイバージェンスの値はあまり改善していないが，正則化を入れたにもかかわらず，$\mathcal{F}$ が改善した点もあることを考えると，統計量の正則化による相乗効果による性能向上に期待できる．

図 6 は, 実験によって得られた統計量の分布を示している．一段目は正解分布を表している．青色で示された初級と，緑色で示された上級の分布はデータセット内の楽譜を用いて計算されたものである．また，オレンジ色で示された中級に関しては上記二つの分布からワッサーシュタイン計量の重心を用いて計算したものである．二段目は，従来手法である難易度条件下で学習を行った場合の同時発音数，音符密度の分布を表す．青色のグラフで示された初級の分布は大きく異なることがわかるが，オレンジで表された中級と緑で表された上級の分布はほとんど同じであることがわかる．しかし，提案手法である，三段目と四段目では，難易度ごとにヒストグラムが区別されていることがわかる．また，四段目の統計量の正則化を加えた場合，初級の時に値が 0 である場合が少なくなった点が数値が改善した点であると考えられる．

図 7 は, 難易度ラベルを変化させた時の，重要度の分布の変化をまとめたグラフを示している．このグラフでは 0（初級）の場合から，0.25 ずつ難易度の値を増やして 1（上級）まで，五つのグラフを比較している．上の，中級に関する損失を用いた場合は，下の，統計量の正則化を用いた場合に比べ，重要度分布が均等に分布が変化していることがわかる．また，図 8 は, 中級に関する損失を用いた場合に出力された楽譜例を示している．オレンジ色で示された音符に注目すると，少しずつ音符が増えていることがわかる．これらより，学習時にランダムに中級のラベルを与える手法が滑らかに難易度を変化させていくに効果があることがわかる．

5. おわりに

本稿では，深層学習を用いて，無段階に難易度を調整するため，音符の重要度に着目したピアノ編曲の手法を提案し，重要度の推定や，学習時の中級の生成の有効性が示された．今後の課題としては，初級と上級の統計分布から中級分布を求め，音符単位の損失だけでなく統計量の正則化にも中級を組み込むことや，初級と上級の楽譜の組み合わせからの中級楽譜の生成，主観評価実験の実施などが挙げられる．

謝辞 本研究の一部は，JSPS 科研費 No. 19H04137, No. 20K21813 および JST さきがけ No. JPMJPR20CB の支援を受けた．

参考文献

[1] Chiu, S.-C., Shan, M.-K. and Huang, J.-L.: Automatic system for the arrangement of piano reduction, *Proc. International Symposium on Multimedia*, pp. 459–464 (2009).
[2] Onuma, S. and Hamanaka, M.: Piano Arrangement System Based On Composers' Arrangement Processes, *Proc. International Computer Music Conference*, pp. 191–194 (2010).
[3] Huang, J.-L., Chiu, S.-C. and Shan, M.-K.: Towards an

automatic music arrangement framework using score reduction, *ACM Transactions on Multimedia Computing, Communications, and Applications*, Vol. 8, No. 1, pp. 8:1–8:23 (2012).

[4] Takamori, H., Sato, H., Nakatsuka, T. and Morishima, S.: Automatic arranging musical score for piano using important musical elements, *Proc. Sound and Music Computing Conference*, pp. 35–41 (2017).

[5] Nakamura, E. and Yoshii, K.: Statistical Piano Reduction Controlling Performance Difficulty, *APSIPA Transactions on Signal and Information Processing*, No. e13, pp. 1–12 (2018).

[6] Wang, Z., Chen, K., Jiang, J., Zhang, Y., Xu, M., Dai, S., Gu, X. and Xia, G.: POP909: A Pop-Song Dataset for Music Arrangement Generation, *Proc. International Society for Music Information Retrieval*, pp. 38–45 (2020).

[7] Tuohy, D. and Potter, W.: A genetic algorithm for the automatic generation of playable guitar tablature, *Proc. International Computer Music Conference*, pp. 499–502 (2005).

[8] Yoshinaga, Y., Fukayama, S., Kameoka, H. and Sagayama, S.: Automatic arrangement for guitars using hidden Markov model, *Proc. Sound and Music Computing Conference*, pp. 450–456 (2012).

[9] Hori, G., Kameoka, H. and Sagayama, S.: Input-output HMM applied to automatic arrangement for guitars, *Journal of Information Processing*, No. 2, pp. 264–271 (2013).

[10] Maekawa, H., Emura, N., Miura, M. and Yanagida, M.: On machine arrangement for smaller wind-orchestras based on scores for standard wind-orchestras, *Proc. International Conference on Music Perception and Cognition*, pp. 268–273 (2006).

[11] Crestel, L. and Esling, P.: Live orchestral piano, a system for real-time orchestral music generation, *Proc. Sound and Music Computing Conference*, pp. 434–442 (2017).

[12] Terao, M., Hiramatsu, Y., Ishizuka, R., Wu, Y. and Yoshii, K.: Difficulty-Aware Neural Band-to-Piano Score Arrangement Based on Note- and Statistic-Level Criteria, *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 196–200 (2022).

[13] H. Takamori, T. Nakatsuka, S. F. M. G. a. S. M.: Audio-based automatic generation of a piano reduction score by considering the musical structure, *International Conference on Multimedia Modeling*, pp. 169–181 (2019).

[14] Wang, Z., Xu, D., Xia, G. and Shan, Y.: Audio-To-Symbolic Arrangement Via Cross-Modal Music Representation Learning, *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 181–185 (2022).

[15] Chiu, S.-C. and Chen, M.-S.: A Study on Difficulty Level Recognition of Piano Sheet Music, *2012 IEEE International Symposium on Multimedia*, pp. 17–23 (2012).

[16] Ramoneda, P., Tamer, N. C., Eremenko, V., Serra, X. and Miron, M.: Score Difficulty Analysis for Piano Performance Education based on Fingering, *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 201–205 (2022).

[17] Ronneberger, O., Fischer, P. and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241 (2015).