

MIDI ギターと音響信号処理の統合によるギター演奏の自動採譜の検討

大塚匡紀[†]北原鉄朗[†][†] 日本大学大学院総合基礎科学研究科地球情報数理科学専攻

1. はじめに

MIDI ギターは、ギター演奏をリアルタイムに MIDI データに変換 (本稿ではこれを採譜という) できるようにした楽器であり、DTM におけるデータ入力などに有用である。しかし、MIDI キーボードなどに比べると採譜精度が十分ではない。我々は、MIDI ギターに音響信号処理を統合することで採譜精度の向上を目指し、その第 1 段階として音響信号処理のみによる採譜手法を検討した¹⁾。本稿では、この手法と MIDI ギターから得られる情報とを統合する方法について検討したので報告する。

2. 音響信号処理によるギター演奏の採譜

文献 1) で提案した音響信号処理のみによる採譜手法について述べる。この手法では、非負値行列因子分解 (NMF)²⁾ を用いる。NMF とは、楽器演奏のスペクトログラム V を非負値行列とみなして、基底行列 W と時系列の重み行列 H の積として、 $V \cong WH$ の形に近似するアルゴリズムである。ただし、NMF は、演奏が終了した信号のスペクトログラムに対して適用するものであり、リアルタイムに適用することはできない。そこで、採譜したい演奏 (これを本演奏とする) に先立ち、各弦の各フレット (計 138 音) を一音ずつ順番に演奏してもらい (第 1 予備演奏とする)、これに対して NMF を適用して、基底行列 W を得る。本演奏を採譜する際には、フレームごとに得られるパワースペクトル v_t (t : 時刻) に対して、この基底行列 W の擬似逆行列をかけることで、重みベクトル $h_t = W^{-1}v_t$ を得る。

重みベクトルは、対応するフレットの音が鳴っている可能性を表すので、しきい値処理によってオンセットとオフセットを決めることができるが、演奏のくせなどによって適切なしきい値が異なる可能性がある。そこで、本演奏の前に、本演奏の演奏者自身があらかじめ決められた内容を演奏し (第 2 予備演奏という)、これを用いて重みベクトルとオンセットか否かの関係を学習する。具体的には、図 1 に示すニューラルネットワークに対して、重みベクトルの各ノートナンバーに対応する値から得られる特徴ベクトルとオンセットか否か (オンセットなら 1、オンセットでなければ 0) を学習する。本演奏の採譜時には、重みベクトル h_t の各ノートナンバーに対応する値に対してこのニューラルネットワークを適用し、出力ノードの値がある値以上であったらそのノートナンバーのオンセットが存在するとみなす。

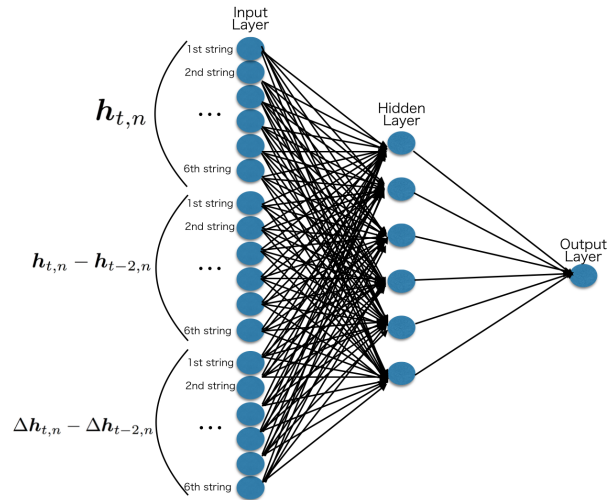


図 1 統合に用いるニューラルネットワーク

3. MIDI ギターの統合

MIDI ギターによる採譜精度はジャンルによって違いがあり、ファンクのようにゴーストノートを多用する演奏では、演奏していないはずの音が多数出力される。そのため、このような音の削減が課題となる。そこで、本稿では次の各手法を試すこととする。

(1) 後処理での論理積による統合

MIDI ギターによる採譜と音響信号処理による採譜をそれぞれ独立に行い、ある音が MIDI ギターにて発音時刻、又は発音状態にあり、かつ音響信号処理によって発音が検出されたときにのみ当該音を出力する。

(2) 前処理での弦番号による絞り込み

MIDI ギターでは各弦の振動を個別に計測して処理を行うので、どの弦が振動しているかについては信頼できると予想される。逆に言えば、MIDI ギターによる出力にある弦が含まれていなければ、その弦は弾かれておらず NMF による採譜は必要ない可能性が高い。そこで、各弦が弾かれたかを MIDI ギターから推定し、弾かれた弦のみ音響信号処理による採譜を行う。

(3) ニューラルネットワークによる統合

音響信号処理による採譜の際に用いるニューラルネットワークの入力層に MIDI ギターによる出力があるかどうかを表すノードを追加し、ここに MIDI ギターによる出力のヴェロシティを入力する。

(4) (1) と (2) の併用

(2) で述べた前処理と (1) で述べた後処理の両方を行う。

A Consideration of Automatic Transcription of Guitar Performance by Integrating MIDI Guitar and Audio Signal Processing
by Masaki Otsuka and Tetsuro Kitahara (Nihon University)

表 1 実験結果

Trk	MIDI ギター			音響信号処理のみ			統合手法 (1)			統合手法 (2)			統合手法 (3)			統合手法 (4)		
	R	P	F	R	P	F	R	P	F	R	P	F	R	P	F	R	P	F
7-1	0.60	0.27	0.37	0.98	0.06	0.12	0.88	0.58	0.70	0.98	0.10	0.18	0.63	0.26	0.37	0.63	0.82	0.71
7-2	0.67	0.25	0.36	0.98	0.06	0.12	0.82	0.45	0.58	0.98	0.10	0.18	0.64	0.27	0.38	0.61	0.76	0.68
9-1	0.61	0.25	0.36	1.00	0.08	0.15	0.98	0.43	0.59	0.99	0.13	0.23	0.74	0.31	0.44	0.69	0.70	0.69
9-2	0.59	0.24	0.34	0.82	0.07	0.13	0.68	0.34	0.46	0.74	0.11	0.19	0.53	0.31	0.39	0.51	0.67	0.58
12-1	0.45	0.56	0.50	0.93	0.14	0.25	0.84	0.47	0.61	0.92	0.22	0.35	0.54	0.37	0.44	0.47	0.81	0.60
17-1	0.45	0.24	0.31	0.90	0.08	0.15	0.52	0.36	0.43	0.85	0.14	0.25	0.73	0.26	0.39	0.42	0.45	0.43
17-2	0.51	0.32	0.39	0.96	0.10	0.18	0.75	0.45	0.56	0.91	0.15	0.25	0.60	0.31	0.41	0.54	0.69	0.61
21-1	0.59	0.07	0.13	1.00	0.03	0.05	0.95	0.38	0.55	1.00	0.05	0.09	0.68	0.20	0.31	0.91	0.67	0.77
21-2	0.79	0.12	0.21	0.97	0.03	0.07	0.90	0.36	0.51	0.97	0.06	0.11	0.62	0.24	0.35	0.90	0.68	0.78
32-1	0.56	0.15	0.23	0.90	0.06	0.11	0.71	0.52	0.60	0.74	0.09	0.15	0.28	0.15	0.20	0.32	0.49	0.39
32-2	0.63	0.19	0.30	1.00	0.06	0.11	0.95	0.55	0.70	0.95	0.10	0.18	0.58	0.29	0.38	0.60	0.83	0.69
33-1	0.58	0.23	0.33	0.91	0.09	0.16	0.69	0.32	0.44	0.89	0.14	0.24	0.44	0.27	0.34	0.41	0.67	0.50
35-1	0.39	0.13	0.19	0.95	0.06	0.12	0.84	0.35	0.49	0.93	0.10	0.19	0.77	0.32	0.46	0.68	0.67	0.67
35-2	0.49	0.35	0.41	0.84	0.14	0.24	0.61	0.57	0.59	0.76	0.23	0.35	0.45	0.42	0.43	0.41	0.76	0.53
47-1	0.38	0.28	0.32	0.88	0.12	0.21	0.50	0.61	0.55	0.84	0.20	0.32	0.67	0.40	0.50	0.38	0.75	0.51
47-2	0.60	0.48	0.53	0.89	0.13	0.23	0.70	0.56	0.62	0.85	0.21	0.33	0.55	0.33	0.41	0.51	0.74	0.60
51-1	0.46	0.49	0.47	0.87	0.15	0.25	0.75	0.55	0.63	0.85	0.24	0.37	0.44	0.37	0.41	0.45	0.74	0.56
51-2	0.56	0.49	0.52	0.95	0.12	0.22	0.85	0.51	0.63	0.94	0.19	0.31	0.71	0.38	0.49	0.56	0.76	0.65
64-1	0.38	0.14	0.21	0.89	0.08	0.15	0.60	0.45	0.51	0.87	0.14	0.24	0.79	0.37	0.51	0.54	0.69	0.61
68-1	0.26	0.25	0.25	0.95	0.11	0.19	0.48	0.27	0.34	0.88	0.15	0.26	0.46	0.20	0.28	0.34	0.38	0.36
72-1	0.50	0.18	0.26	0.96	0.06	0.11	0.98	0.32	0.49	0.98	0.09	0.17	0.52	0.24	0.33	0.59	0.48	0.53
平均	0.53	0.27	0.33	0.93	0.09	0.16	0.76	0.45	0.55	0.89	0.14	0.24	0.59	0.30	0.39	0.55	0.68	0.59

4. 評価実験

4.1 実験方法

実験には、MIDI ギターにはストラトキャスターに Roland GK-3 を取り付けたもの、MIDI インターフェースには Roland GR-55 を使用した。採譜対象の演奏は、文献3) から選んだ 21 フレーズ (BPM120) である。それらの演奏に対して、MIDI ギターのみ、音響信号処理のみ、統合手法 (1)~(4) による採譜を行った。なお、表 1 における Trk は、文献3) の付属音源の CD-Track を表す。

4.2 実験結果・考察

各々の手法を用いた採譜結果の再現率、適合率、F 値を表 1 に示す。MIDI ギターによる採譜結果の再現率が平均 0.53 であったものの適合率がすべて 0.50 未満 (平均 0.27) であったために F 値がすべて 0.60 未満 (平均 0.33) であったのに対し、統合手法 (4) では、適合率が 21 個中 10 個で 0.70 を超え、F 値の平均値が 0.59 になった。この結果は、本手法により、MIDI ギターの問題点である存在しない音の出力を削減できたことを意味している。なお、音響信号処理のみの場合は、再現率はすべて 0.80 以上と高いものの適合率が低い (すべて 0.20 未満) ために F 値が低い (すべて 0.30 未満) 結果となった。

統合手法の (1) と (4) を比較すると、(4) は (1) に比べて再現率は下がっている (平均 0.76 に対して 0.55) ものの適合率が上がる (平均 0.45 に対して 0.68) 結果となった。演奏ごとに見ると、再現率が 0.10 以上上がった演奏は 18 個あるが、一方で適合率が 0.10 以上上がった演奏は 20 あったために、結果的に 7 個で F 値が 0.10 以上向上した。

統合手法 (2) は、再現率が高く (平均 0.89) 適合率が低く (平均 0.14) なっており、音響信号処理のみとあまり変わらない結果となった。これは、音響信号処理に比べて追加されている制約が弦の情報のみであり、存在しない音の出力を削

除しきれなかったものと考えられる。

統合手法 (3) は、音響信号処理や統合手法 (2) と比べて適合率が高くなっている (平均 0.30) もの、再現率は低下してしまっており、統合手法 (1) や統合手法 (4) に比べると F 値が低い結果 (平均 0.39) となった。

統合手法 (4) について演奏ごとに見ていくと、21 個中 10 個で適合率が 0.70 以上であった。また、Track 21-1 と 21-2 は 0.90 以上の再現率を維持したまま他の手法に比べて高い適合率を実現しており、0.75 以上の F 値を実現した。これは、演奏が単音カッティングで、ブラッシングなどが原因で存在しない音が多く出力されていたが、後処理の論理積により多く削減されたと考えられる。一方、Track 17-1、68-1 と 72-1 は 0.50 未満の適合率しか得られなかった。これは、複数の音符においてピッチエラーが起きていたからである。また、Track 32-1、33-1 は MIDI ギターよりも再現率が 0.10 以上低下しており、音響信号処理において低音弦の発音を検出できていなかったことが原因だと推測される。

5. おわりに

本稿では、MIDI ギターと NMF による音響信号処理の統合手法を報告した。今後は、ニューラルネットワークの学習データなどを工夫し、更なる精度向上を目指したい。

謝辞 本研究は、科研費 26240025 の助成を受けて行われた。嵯峨山茂樹教授 (明治大学)、堀玄准教授 (亜細亜大学) から貴重なご意見をいただいた。ここに感謝する。

参考文献

- 1) M. Otsuka and T. Kitahara: "An On-line Algorithm of Guitar Performance Transcription Using Non-Negative Matrix Factorization", *Proceedings of International Conference on Signal Processing*, pp.621-624, 2014.
- 2) D.D. Lee and H.S. Seung: "Learning The Parts of Objects with Nonnegative Matrix Factorization", *Nature*, vol.401, pp.788-791, 1999.
- 3) 山口和也: "16 ビートが身につく! ファンクで覚える大人のカッティング", リットー・ミュージック, 2013.