

# 補助関数法による MIDI 信号の音楽音響信号への 詳細なフィッティング

高宗 典玄<sup>1,a)</sup> 亀岡 弘和<sup>1,2,b)</sup> 土屋 政人<sup>1,c)</sup> 嵯峨山 茂樹<sup>1,d)</sup>

## 概要 :

本報告では、楽譜が与えられているもとの、音響信号を最もよく再現する MIDI のベロシティやオンセットの時刻といったパラメータを推定する問題を扱う。MIDI パラメータの推定を行うためには、MIDI が生成する音響信号のスペクトログラムと実際の観測音響信号のスペクトログラムの距離を最小化するようにパラメータを設定する必要がある。この最適化問題は膨大な解空間を持つため、単純に大域最適解を探すのは困難である。そこで、本研究では楽譜情報を用いて正解に近いパラメータを与えた上で補助関数法を用いて効率的に局所最適解を探索する手法を提案する。音楽音響信号に対して MIDI のパラメータを提案法を用いて推定する実験を行い、提案法の特徴について議論する。

キーワード : MIDI, 音響信号への MIDI フィッティング, 補助関数法, 演奏表情

## Auxiliary function approach for optimal MIDI-fitting to audio signals

TAKAMUNE NORIHIRO<sup>1,a)</sup> KAMEOKA HIROKAZU<sup>1,2,b)</sup> TSUCHIYA MASATO<sup>1,c)</sup> SAGAYAMA SHIGEKI<sup>1,d)</sup>

**Abstract:** In this paper, we tackle to fit parameters of MIDI signals such as velocities and onset times to recorded polyphonic signals, which are performed with reference of the corresponding score. The large gap between MIDI parameters and acoustic signals exists and cannot compare simply. We regard the similarity of MIDI parameters and acoustic signals as the similarity on spectrogram, and develop the fitting algorithm which are efficient thanks to auxiliary function approach. For the sake of computation complexity, our method has information of scores associated with acoustic signals. We discuss merits and demerits of the proposed method through the experiments.

**Keywords:** MIDI, Audio to MIDI Alignment, Auxiliary function approach, Performance rendering

## 1. はじめに

本報告では、所与の音楽音響信号を MIDI (Musical Instrument Digital Interface) 信号で音響的に良く再現するよう、MIDI 信号中の各単音のベロシティやオンセット時

刻などのパラメータを自動チューニングする方法を提案する。

MIDI 信号は、個々のノートの音高、発音タイミング (オンセット時刻) や発音強度 (ベロシティ) などの情報で構成され、楽譜を書いたり楽譜上で音符を操作するような感覚で作編曲や楽曲加工を手軽に行うことが可能である。一方で、表情豊かな人間の演奏らしい演奏を MIDI で打ち込むのは、非常に入念な人手のチューニングを要し、必ずしも容易ではない。しかし、表情豊かな人間らしい演奏を自動で MIDI に打ち込めるようになれば、MIDI の用途は格段に広がるはずであろう。本発表では、楽曲の楽譜とそれが

<sup>1</sup> 東京大学大学院情報理工学系研究科  
Graduate School of Information Science and Technology,  
The University of Tokyo

<sup>2</sup> NTT コミュニケーション科学基礎研究所  
NTT Communication Science Laboratory

a) takamune@hil.t.u-tokyo.ac.jp

b) kameoka@hil.t.u-tokyo.ac.jp

c) tsuchiya@hil.t.u-tokyo.ac.jp

d) sagayama@hil.t.u-tokyo.ac.jp

実際に演奏された音楽音響信号が与えられている状況のもとで、音楽音響信号を「音響的」に再現するよう、MIDI信号中の各単音のペロシティやオンセット時刻などのパラメータを詳細に自動チューニングする方法を提案する。近年、歌声合成システムでユーザ歌唱を真似るように、合成歌唱のパラメータ（音高と音量）を自動調整するシステムが提案された [1] が、提案法はこれの多重音版と位置付けられよう。

## 2. 音楽音響信号のスペクトルと MIDI 信号のフィッティング

### 2.1 MIDI パラメータ推定問題の定式化

単音のスペクトログラムの和によって音楽音響信号が生成されているという仮定を置くと、MIDI パラメータの推定は以下のように観測音響信号と MIDI が生成する音響信号とのスペクトログラム上の距離を示す目的関数  $J(\Theta)$  を最小化する問題として捉えられる。

$$J(\Theta) = \sum_{\omega,t} D \left( Y_{\omega,t} \left| \sum_k X_{\omega,t,k} \right. \right), \quad (1)$$

$$X_{\omega,t,k} = f(\theta_k). \quad (2)$$

ここで、 $\omega$  と  $t$  は周波数と時間方向の添え字であり、 $k$  は音符の添え字である。また、 $Y_{\omega,t}$  は観測音響信号のスペクトログラム、 $X_{k,\omega,t}$  は MIDI パラメータ  $\theta_k$  から生成される音響信号の  $k$  番目の単音のスペクトログラムである。 $D$  はスペクトログラム間の距離を表し、例えば二乗誤差、I ダイバージェンス、板倉斎藤距離

$$D_{EU}(y|x) = (y - x)^2, \quad (3)$$

$$D_{KL}(y|x) = y \log \frac{y}{x} - y + x, \quad (4)$$

$$D_{IS}(y|x) = \frac{y}{x} - \log \frac{y}{x} - 1 \quad (5)$$

や、[2] で用いられている

$$D_{KA}(y|x) \equiv \begin{cases} \beta/\gamma & (|y-x| \leq \beta-c) \\ \frac{(y-x-|\beta-c|)^2}{4c\gamma} + \frac{\beta}{\gamma} & (\beta-c < |y-x| \leq \beta+c) \\ |y-x|/\gamma & (|y-x| > \beta+c) \end{cases} \quad (6)$$

のような距離関数で測ることができる。

この最適化問題は膨大な解空間を持つため、単純に大域最適解を探すのは困難である。もし、目的関数が単音ごとのスペクトログラム間距離の和に分離した形をしていれば、各単音ごとに最適な MIDI パラメータを探索することができて好都合であるが、残念ながらこの目的関数ではそうはなっていない。そこで、我々は、目的関数の上限となる補助関数を反復的に降下させることで目的関数を間接的に降下していく方法をベースにし、その補助関数として単音ご

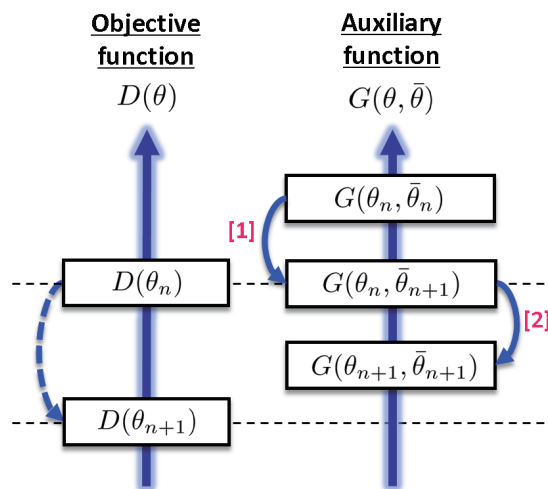


図 1 補助関数法の原理

Fig. 1 Principle of auxiliary function approach

とのスペクトログラム間距離の和に分離した形をとるものをうまく設計することで、当該最適化問題の解を見通し良く探索することができると考えた。補助関数の反復降下による目的関数の降下方法を補助関数法と呼び、音響信号処理分野で近年様々な最適化問題に適用されている [3], [4]。補助関数の定義と補助関数法の原理は以下のとおりである。

定義 1.  $\theta$  をパラメータとする目的関数  $D(\theta)$  に対し、

$$D(\theta) = \min_{\bar{\theta}} G(\theta, \bar{\theta}) \quad (7)$$

が成り立つとき、 $G(\theta, \bar{\theta})$  を  $D(\theta)$  の補助関数 (Auxiliary function),  $\bar{\theta}$  を補助変数と定義する。

定理 1 (補助関数法). 補助関数  $G(\theta, \bar{\theta})$  を、 $\bar{\theta}$  に関して最小化するステップと、 $\theta$  に関して最小化するステップ

$$\bar{\theta} \leftarrow \arg \min_{\bar{\theta}} G(\theta, \bar{\theta}), \quad (8)$$

$$\theta \leftarrow \arg \min_{\theta} G(\theta, \bar{\theta}) \quad (9)$$

を繰り返すと、目的関数  $D(\theta)$  の値は単調収束する。

証明： 反復計算のステップ数を  $n$  とし、 $\theta = \theta_n$ ,  $\bar{\theta} = \bar{\theta}_n$  から  $\theta = \theta_{n+1}$ ,  $\bar{\theta} = \bar{\theta}_{n+1}$  に更新されたときに、 $D(\theta)$  が増加しないことを示す。 $\bar{\theta}_{n+1} = \arg \min_{\bar{\theta}} G(\theta_n, \bar{\theta})$ ,  $\theta_{n+1} = \arg \min_{\theta} G(\theta, \bar{\theta}_{n+1})$  とすると、補助関数の定義より  $D(\theta_n) = G(\theta_n, \bar{\theta}_{n+1})$  であり、 $\theta_{n+1} = \arg \min_{\theta} G(\theta, \bar{\theta}_{n+1})$  より、 $G(\theta_n, \bar{\theta}_{n+1}) \geq G(\theta_{n+1}, \bar{\theta}_{n+1})$  である。更に補助関数の定義より  $G(\theta_{n+1}, \bar{\theta}_{n+1}) \geq D(\theta_{n+1})$  であるから、結局、 $D(\theta_n) \geq D(\theta_{n+1})$  である (図 1 参照)。□

ここで、距離関数  $D$  が凸関数であれば次の Jensen の不等式を用いて補助関数を求めることができる。

定理 2 (Jensen の不等式). 任意の凸関数  $g$ ,  $I$  個の実数

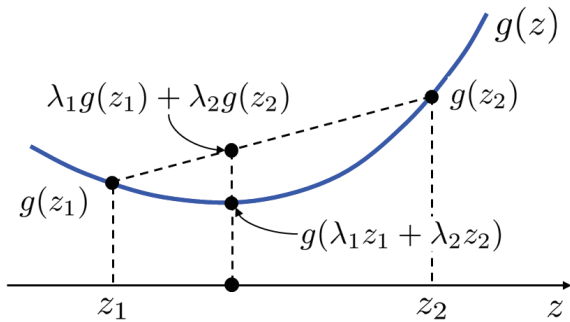


図 2 Jensen の不等式  
 Fig. 2 Jensen's inequality

$x_1, \dots, x_I, \sum_i \lambda_i = 1$  を満たす  $I$  個の正値の重み係数  $\lambda_1, \dots, \lambda_I$  のもとで,

$$g\left(\sum_i \lambda_i z_i\right) \leq \sum_i \lambda_i g(z_i) \quad (10)$$

が成り立つ (図 2 参照).

スペクトログラム間距離を二乗誤差で測った場合の補助関数の設計例は以下のとおりである.

$$\begin{aligned} J(\Theta) &= \sum_{\omega, t} \left( Y_{\omega, t} - \sum_k \frac{\lambda_{\omega, t, k}}{\lambda_{\omega, t, k}} X_{\omega, t, k} \right)^2 \\ &\leq \sum_{\omega, t} \sum_k \lambda_{\omega, t, k} \left( Y_{\omega, t} - \frac{X_{\omega, t, k}}{\lambda_{\omega, t, k}} \right)^2 \\ &= \sum_k \sum_{\omega, t} \frac{1}{\lambda_{\omega, t, k}} (\lambda_{\omega, t, k} Y_{\omega, t} - X_{\omega, t, k})^2 \end{aligned} \quad (11)$$

等号の成立は  $\lambda_{\omega, t, k} = \frac{X_{\omega, t, k}}{\sum_{k'} X_{\omega, t, k'}}$  である.

## 2.2 時間軸方向での初期値の設定

今回は目的関数 (式 1) の最適化に対して大域最適解を探索することが困難であるため, 補助関数法という高速な局所最適解探索の手法を用いた. さらに楽譜情報から得られる機械的演奏を仮定した MIDI 信号を与え, 正解に近いオンセット位置と音高を初期値として探索をすることで探索すべき解空間を大幅に狭めることができた. しかし, 人間の演奏は常にテンポ変動を含むため楽譜上の音符列情報に対して時間方向の伸縮が加わった形で演奏される. 故に楽譜情報からの機械的演奏の MIDI 信号を与えても, 実際の音響信号と MIDI 上の時間軸に対する同期をとらなければならないという課題が残っている. そこで, 動的計画法によるマッチング (DP マッチング) の結果を実時間情報の初期値として利用した. これにより DP マッチングをせずに補助関数法による最適化を行った場合よりも, 時間方向に関して推定誤りを起こしにくくなることが予想される.

しかし, スペクトログラム同士のマッチングを考えるとパワーの違いによる影響が大きくなってしまふ. そこで本報告では, Ellis の方法 [5] を利用した. この方法では, 類似度を音楽音響信号のスペクトログラムの周波数成分のう

ち, MIDI 信号の各ノートの音高に対応する周波数成分の割合で表現するため, スペクトログラムを定数倍しても類似度に変化しないという特徴がある.

## 3. 実験

### 3.1 実験条件

実験の用いたデータは, RWC のクラシック音楽データベース [6] から以下の曲の一部を用いた.

- RWC-MDB-C-2001 No. 28 の一部 (4 小節)
- RWC-MDB-C-2001 No. 32 の一部 (4 小節)
- RWC-MDB-C-2001 No. 33 の一部 (4 小節)

スペクトログラムの計算には Gabor の Wavelet 変換を用いた. 実験に用いたパラメータは以下のとおりである.

- 距離関数: 二乗誤差
- サンプリング周波数: 44100 Hz
- 時間分解能: 8 ms
- 周波数分解能: 48 cent
- 最低中心周波数: 50 Hz
- 補助関数法の反復回数: 10 回
- 考慮した MIDI パラメータの種類: 発音時刻 (オンセット), 消音時刻 (オフセット), 発音強度 (ベロシティ)
- 各反復毎のオンセット位置の検索範囲:  $\pm 400$  ms
- 各反復毎のオフセット位置の検索範囲:  $\pm 400$  ms
- 各反復毎のベロシティの検索範囲: 全通り (1~127)

### 3.2 結果

図 3 はフィッティングもとの音楽音響信号のスペクトログラムを示す. 図 4 は DP マッチングで推定した MIDI パラメータから生成される音響信号のスペクトログラムを示す. 図 5 は DP マッチングで推定した MIDI パラメータを初期値として, 補助関数法で推定した MIDI パラメータから生成される音響信号のスペクトログラムを示す.

図 3 と図 5 を比較すると音楽音響信号のスペクトログラムをよく近似できる MIDI パラメータが推定出来ていることが分かる. また, 図 3 と図 5 を図 4 と比較すると, DP マッチングでは推定することが出来なかったベロシティが, 補助関数法により推定出来ていることが分かる.

但し, 図 3 と図 4 を比較すると, ノートが低音部のみに見えているときや, 速いノートが連続する場合に, DP マッチングで推定誤りが生じていることが分かる. これは, Wavelet 変換の低周波領域における時間分解能の低さや残響の影響と考えられる. 更に, この結果を踏まえ図 5 を見ると, DP マッチングにより生じた推定誤りが補助関数法による推定に強く影響を及ぼしていることが読み取れる. このため, 今後は DP マッチングの精度の向上や, 補助関数法を行う上でアニーリングなど初期値依存性を緩和させる手法の検討をする必要がある.

#### 4. まとめ

本報告では、音楽音響信号と MIDI 信号をフィッティングする新しい手法を提案した。時間領域では比較するのが困難であるこの問題に対し、MIDI 信号から生成される波形と観測音響信号をスペクトログラム上の距離の最小化として捉え定式化した。さらに、膨大な解空間を探索しなければならないこの最適化に対して、元の楽譜から得られる機械的演奏を仮定したオンセット、音高情報と DP マッチングから得られる実時間情報を用いて解空間を効率的に探索する手法を提案した。実験ではベロシティ・オンセット・オフセットといった MIDI パラメータを推定することができ、本手法の有効性が示された。今後の研究では、MIDI パラメータの種類を増やして実験を行うことを検討している。

**謝辞** 本研究の一部は、文部科学省/学術振興会科学研究補助費 課題番号 (23240021) から補助を受けて行われた。

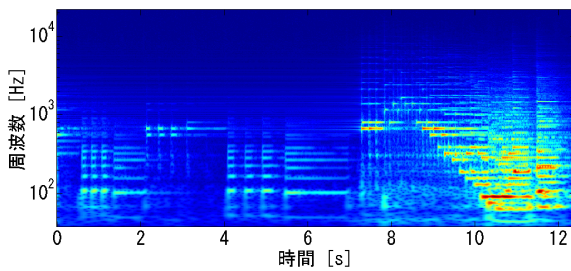


図 3 音楽音響信号のスペクトログラム  
Fig. 3 The observed spectrogram

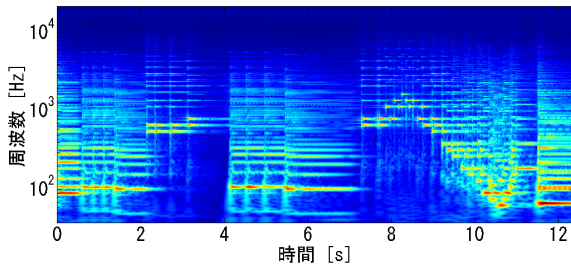


図 4 DP マッチングにより推定された MIDI パラメータから生成される音響信号のスペクトログラム

Fig. 4 The spectrogram generated from MIDI parameters estimated by DP-matching

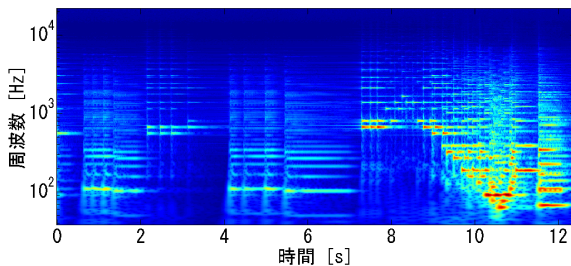


図 5 補助関数法により推定された MIDI パラメータから生成される音響信号のスペクトログラム

Fig. 5 The spectrogram generated from MIDI parameters estimated by auxiliary function approach

#### 参考文献

- [1] 中野倫靖, 後藤真孝. Vocalistener: ユーザ歌唱とその歌詞を用いた歌声合成パラメータの自動推定システム. 日本音響学会 2008 年 秋季研究発表会 講演論文集, pp. 1-4.
- [2] 亀岡弘和, 鎌本優, 原田登, 守谷健弘. 予測誤差の golomb-ricce 符号量を最小化する線形予測分析 (電気音響, 音響一般). 電子情報通信学会論文誌. A, 基礎・境界, Vol. 91, No. 11, pp. 1017-1025, 2008.
- [3] Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In *Proc. NIPS*, pp. 556-562, December 2000.
- [4] H. Kameoka, N. Ono, K. Kashino, and S. Sagayama. Complex NMF: A new sparse representation for acoustic signals. In *Proc. ICASSP*, pp. 3437-3440, April 2009.
- [5] D. P. W. Ellis. Aligning midi scores to music audio, web resource. <http://www.ee.columbia.edu/~dpwe/resources/matlab/alignmidiwav/>.
- [6] Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, and Ryuichi Oka. RWC music database: Popular, classical, and jazz music database. In *Proc. ISMIR*, pp. 287-288, October 2002.