

# 同時録音ドラム演奏音源に対するグルーブ感の評価

河田 洋人<sup>1,a)</sup> 保利 武志<sup>1,b)</sup> 中村 和幸<sup>1,c)</sup>

**概要:** 本研究では同時録音ドラム演奏音源を対象としたグルーブ感評価方法の提案を行う。従来の音源分離を用いたドラム演奏のグルーブ感評価の研究ではドラム演奏のグルーブ感をハイハット、スネア、バスドラムの音量、打点時刻のゆらぎと抽象的に定義しているため、ドラム演奏のグルーブ感として知られている、タイト、ルーズなどといった具体的なグルーブ感への評価は実現されなかった。本研究ではタイト、ルーズごとにスネア、バスドラムの打点時刻のずれを確率変数とする確率モデルを定義することで、打点時刻からタイト、ルーズへのグルーブ感の評価を行う。評価実験の結果、タイト、ルーズごとに録音した音源に対して、それぞれの結果を正しく反映するような評価値を得ることを実現した。

## 1. はじめに

本研究では同時録音ドラム演奏音源に対するグルーブ感の評価方法を提案する。

一般にグルーブ感とはリズムから生じる演奏の表情のことを指し [1], Kawase et al.[2] による主観調査の結果「ノリ」「一体感」「心が弾む」「テンポが速い」「体でリズムを感じられる」といった感覚に近いことがわかっている。特にドラム演奏においては前ノリ、後ノリ、ジャスト、タイト、ルーズといった具体的なグルーブ感がプロのドラム奏者や音楽愛好家の間で知られており、これらはドラム奏者によって表現されたドラム演奏の聴覚上の印象を表している。

ドラム演奏のグルーブ感を定量的に評価することができれば、以下のようなことが実現される。まず、音楽検索における検索情報としてドラム演奏のグルーブ感を用いることが可能になり、楽曲名やアーティスト名、作曲者名、演奏時間、歌詞、ジャンル名などのメタ情報による検索に比べ、より感覚的に楽曲を検索することが可能になる。また、演奏支援という観点からは、感覚的な暗黙知として扱われるグルーブ感をより客観的に伝え、グルーブ感のある演奏を習得することを目的とした演奏評価システムの構築が期待できる。

奥村ら [3] はグルーブ感の違いがドラムセットの内のハイハット、スネア、バスドラムの打点時刻のずれ、及び音量のゆらぎに起因すると仮定し、タイト、ルーズなグルーブ感と打点時刻、音量の関係を別録音ドラム演奏音源を対象に解析を行った。この研究が別録音ドラム演奏音源を対象としていたのに対して、Ohya et al.[4] は NMF[5] による音源分離手法を用いることによって、同時録音ドラム演奏音源に対しても、ハイハット、スネア、バスドラムの打点音量と打点時刻のずれの算出を実現した。しかし、同時録音ドラム演奏音源の音量及び打点時刻のゆらぎの抽出は実現されたものの、タイト、ルーズといった具体的なグルーブ感への評価は実現されていない。

本研究では新たに、打点時刻のずれとタイト、ルーズなグルーブ感との関係を表す確率モデルを定義することで、同時録音ドラム演奏音源に対するタイト、ルーズへの評価方法を提案する。評価実験の結果タイト、ルーズな演奏に対して、その演奏を正しく反映するような評価値を得ることを実現した。

本論文では 2 節にて手法を述べ、3 節で評価実験、4 節で結論を述べる。

## 2. 手法

### 2.1 打点時刻のずれの抽出、グルーブ感のモデル化

まず、全体に関わる仮定として、タイト、ルーズといったグルーブ感の違いはスネア、バスドラムの打点時刻のずれに起因するという仮定を置く。表 1,2 はテンポ 100 における最も基本的なドラム演奏を行った際のタイト、ルーズそれぞれの打点時刻のずれの平均と標準偏差をまとめたものである。この結果が示すように、最も基本的なドラム演奏パターンにおいて、ルーズな演奏の方が、タイトな演奏に比べて基準よりも打点時刻が遅れることがわかっている。なお表 1,2 は [3] において図として掲載されているも

<sup>1</sup> 明治大学大学院先端数理科学研究科  
〒164-8525 東京都中野区 4-21-1

a) cs181004@meiji.ac.jp

b) hori@meiji.ac.jp

c) knaka@meiji.ac.jp

表 1 タイト テンポ 100 における、打点時刻のずれの 16 分音符に対する割合

	バス 1 拍	バス 3 拍	スネア 2 拍	スネア 4 拍
平均	0.05714286	0.07142857	0.01428571	0.2857143
標準偏差	0.08571429	0.04285714	0.02857143	0.04285714

表 2 ルーズ テンポ 100 における、打点時刻のずれの 16 分音符に対する割合

	バス 1 拍	バス 3 拍	スネア 2 拍	スネア 4 拍
平均	-0.1071429	-0.1	-0.15	-0.1642857
標準偏差	0.04285714	0.07142857	0.05714286	0.04285714

のを表として記載したものである。

この仮定を踏まえ、同時録音ドラム演奏音源のタイト、ルーズへの評価という問題は 2 つの段階に分けて考えられる。

一つ目の段階は打点時刻のずれを抽出するという段階である。これは同時録音ドラム演奏音源から打点時刻のずれを抽出する問題としてとらえることができる。打点時刻のずれを抽出するためには、ずれを図る基準と打点時刻が必要となる。そこで、演奏のテンポは一定であるという仮定を置くことでずれを図る基準を算出し、スネア、バスドラムの打点時刻は NMF による音源分離手法によって算出する。

二つ目の段階はスネア、バスドラムの打点時刻のずれからタイト、ルーズへの評価を行うという段階である。これは、スネア、バスドラムの打点時刻のずれを確率変数とする確率モデルを定義することによって実現される。そして、抽出した打点時刻のずれから、タイト、ルーズそれぞれに対して定義したモデルへの尤度を計算し、これを評価値とする。

本研究では、グルーブ感の評価を上記二つの複合問題としてとらえ、これらを組み合わせた評価方法を提案する。次項以降ではその手法を具体的に説明していく。

## 2.2 NMF を用いた打点の検出

Non negative matrix factorization(NMF) は非負値行列を 2 つの非負値行列の積に分解するアルゴリズムである。非負値行列  $\mathbf{V} \in \mathbf{R}^{N \times M}$  が与えられた際に 2 つの低ランク行列  $\mathbf{W} \in \mathbf{R}^{N \times K}$ ,  $\mathbf{H} \in \mathbf{R}^{K \times M}$  の積が  $\mathbf{V}$  に近似されるように  $\mathbf{W}$  と  $\mathbf{H}$  を更新していく。NMF は音声信号処理、画像処理への応用が知られており、音声信号に応用する際には、非負の値をもつパワースペクトルを行列として表し、これを観測行列  $\mathbf{V}$  とする。分離後の行列  $\mathbf{W}$  と  $\mathbf{H}$  をそれぞれ基底行列とアクティベーション行列と呼び、基底行列は頻出するスペクトルを表す基底ベクトルをもつ行列、アクティベーション行列は各基底に対応したスペクトルのアクティベーションを表すベクトルをもつ行列として解釈され

る。本研究ではドラム演奏において主要パートを担うとされる、ハイハット、スネア、バスドラムからなるドラム演奏を扱うという仮定のもと、分離を行う。そのため、分解の際の  $K$  の値を 3 と設定し NMF を行い、得られたアクティベーション行列から打点の算出を行う。NMF の距離関数はユークリッド距離を用いる。通常ユークリッド距離の場合の NMF の更新式は

$$\mathbf{H}_{km} \leftarrow \mathbf{H}_{km} \frac{(\mathbf{W}^T \mathbf{V})_{km}}{(\mathbf{W}^T \mathbf{W} \mathbf{H})_{km}} \quad (1)$$

$$\mathbf{W}_{nk} \leftarrow \mathbf{W}_{nk} \frac{(\mathbf{V} \mathbf{H}^T)_{nk}}{(\mathbf{W} \mathbf{H} \mathbf{H}^T)_{nk}} \quad (2)$$

となるが、精度の高いアクティベーションを得るために  $\mathbf{W}$  については GarageBand のドラム音源から抽出したハイハット、スネア、バスドラムのスペクトルを列ベクトルとして持つ行列  $\mathbf{W}_0$  で固定する。そのため本研究においては更新式

$$\mathbf{H}_{km} \leftarrow \mathbf{H}_{km} \frac{(\mathbf{W}_0^T)_{km}}{(\mathbf{W}_0^T \mathbf{W}_0 \mathbf{H})_{km}} \quad (3)$$

を繰り返すことで、アクティベーション行列のみを計算する。

なお、基底行列内の基底ベクトルの  $L_1$  ノルムは基底のスケール任意性を回避するため、1 に揃えてある。

$$\forall k \sum_n w_{nk} = 1 \quad (4)$$

## 2.3 ずれを測る基準について

スネア、バスドラムの打点時刻から、打点時刻のずれを算出するためには、ずれを測るための基準が必要となる。本研究では固定したテンポでの演奏を仮定しているため、ずれを測る基準の時間間隔は一定となる。また、ずれを測る基準の初期時刻はバスドラムの最初の打点から、平均的なバスドラムのずれを引いたものとする。そのため扱うドラム演奏は 1 拍目にバスドラムが叩かれるという仮定を置くことになるが、殆どのドラム演奏において 1 拍目はバスドラムが置かれるので大きな問題とはならない。

## 2.4 二変量正規分布を用いたグルーブ感のモデル化

タイト、ルーズそれぞれのスネア、バスドラムの打点時刻のずれはこれらを確率変数とした二変量正規分布に従うと仮定する。今回はタイト、ルーズごとに理想的なずれの期待値があるとし、人間の演奏によってずれがその値の周辺に分布するという考えのもと二変量正規分布に従うという仮定を置いた。また、スネア、バスドラムのずれに相関はないと仮定し、共分散は 0 であるとした。

そこで、スネア、バスドラムの打点時刻のずれの平均、分散がそれぞれ  $(\mu_s, \sigma_s^2)$ ,  $(\mu_b, \sigma_b^2)$  であり、スネア、バスドラムの打点時刻のずれを表す確率変数を  $X_s, X_b$  とするとこのモデルは

$$\begin{pmatrix} X_s \\ X_b \end{pmatrix} \sim \text{MultiNormal} \left( \begin{pmatrix} \mu_s \\ \mu_b \end{pmatrix}, \begin{pmatrix} \sigma_s^2 & 0 \\ 0 & \sigma_b^2 \end{pmatrix} \right) \quad (5)$$

のように表現でき、この密度関数は

$$f(x, y) = \frac{1}{2\pi\sigma_s\sigma_b} \exp\left\{-\frac{(x - \mu_s)^2}{2\sigma_s^2} - \frac{(y - \mu_b)^2}{2\sigma_b^2}\right\} \quad (6)$$

となる。

実際にタイト、ルーズに対して式 (5) を定義する際にはタイト、ルーズそれぞれに対して、平均、分散のパラメータが必要となるが、これらは表 1,2 を用いて算出する。この表は先述した通り、テンポ 100 におけるスネア、バスドラムの打点時刻のずれの 16 分音符の長さに対する割合を示している。そこで、この割合がテンポによらず一定であると仮定し、演奏のテンポからタイト、ルーズそれぞれに対してスネア、バスドラムの打点時刻のずれの平均、分散を算出しこのモデルのパラメータを定める。

## 2.5 対数尤度関数によるグルーブ感の評価

対数尤度関数は  $N$  個のサンプル  $x_1, \dots, x_N$  に対してパラメータ  $\theta$  の確率分布  $f(x|\theta)$  が与えられている時

$$\log L(\theta) = \log \prod_{i=1}^N f(x_i|\theta) = \sum_{i=1}^N \log f(x_i|\theta) \quad (7)$$

と定義される。

通常パラメータ  $\theta$  は最尤推定などを行う際に変数として扱われるが、本研究ではこの尤度関数のパラメータをタイト、ルーズごとに固定し得られる定数値をタイト、ルーズへの評価値とする。

対数を取らない通常の尤度を評価値とすることも考えられるが値が大きい、もしくは 0 に近くなりすぎるといった問題点があるため、本研究では対数尤度を採用した。

## 2.6 グルーヴ感の評価手順

図 1 は本研究におけるグルーブ感評価方法をブロック図にして図示したものである。入力となる音源に対する仮定をまとめると、ハイハット、スネア、バスドラムからなる同時録音ドラム演奏音源で、テンポは一定である。出力は対数尤度に基づいたタイト、ルーズそれぞれに対する評価値となっている。NMF を用いる際の前処理として観測行列を作成する必要があるため、対象音源に対して短時間フーリエ変換 (STFT) を行いパワースペクトログラムを算出する。これを観測行列とした NMF による音源分離を行うことで、バスドラム、スネアドラムのアクティベーションを得る。得られたアクティベーションのピークから打点時刻を算出し、基準とのずれを求める。打点時刻のずれからタイト、ルーズそれぞれの確率モデルに対して尤度計算を行い、その値をドラム演奏音源のグルーブ感評価値とする。

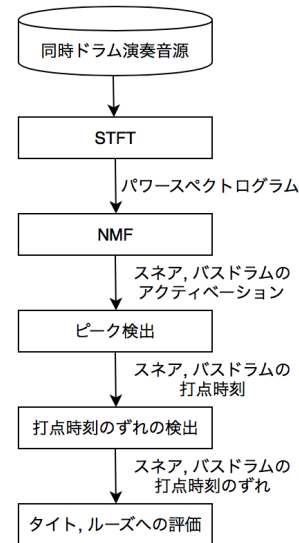


図 1 本研究におけるグルーブ感評価方法のブロック図



図 2 評価実験に用いたドラム演奏音源のスコア

## 3. 評価実験

### 3.1 実験条件

タイト、ルーズな演奏に対して、2 節で提案したドラム演奏のグルーブ感評価方法を用いることによってどのような評価値が得られるかを検証するために、タイト、ルーズそれぞれに対して演奏の録音を行い評価を行った。ドラム演奏経験 10 年以上のアマチュアドラム奏者にタイト、ルーズそれぞれのグルーブ感でメトロノームを聞きながら演奏を行ってもらい、サンプリングレート 192 kHz で同時録音を行った。グルーブ感の違いの検証のためグルーブ感以外のテンポ、奏者、使用楽器、スコアといった条件は全て同一である。図 2 は演奏に用いたスコアでこれはドラム演奏における最も基本的なスコアとされる。このスコアをもとにタイト、ルーズともにテンポ 80 で 16 小節分を繰り返した演奏を行った。

### 3.2 NMF を用いたアクティベーション計算結果

まず、NMF の前段階として観測行列を作成した。計算速度向上のため音源のサンプリングレートを 192 kHz から 48 kHz にダウンサンプリングし、16 小節分のデータを取り出すと 2304000 サンプル数の音源データを得た。この音源データに対して、STFT を行った。STFT をする際

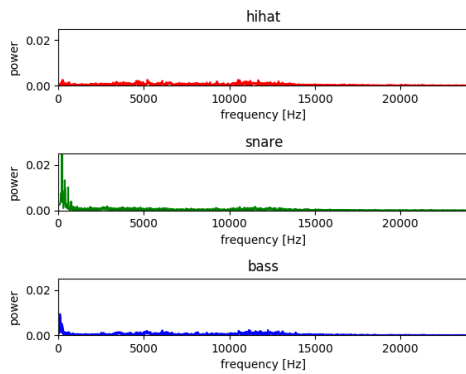


図 3 学習に用いた基底スペクトル

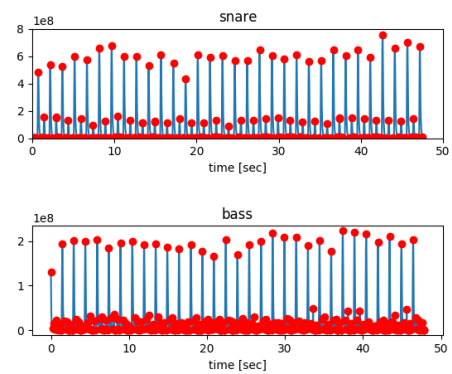


図 6 タイトな演奏のアクティベーションのピーク

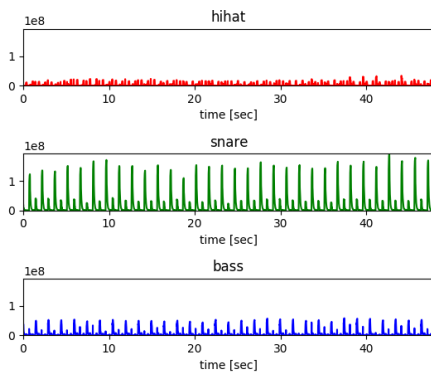


図 4 タイトな演奏のアクティベーション

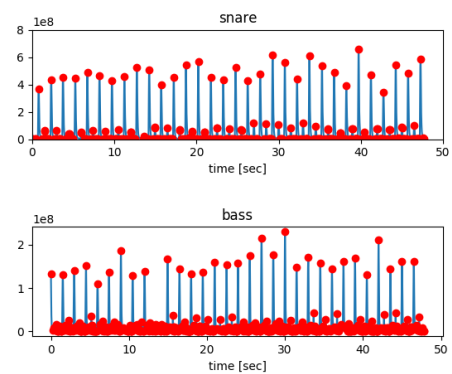


図 7 ルーズな演奏のアクティベーションのピーク

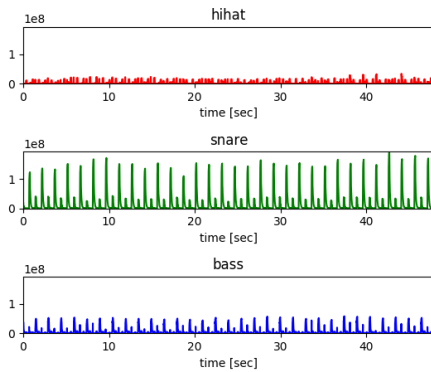


図 5 ルーズな演奏のアクティベーション

のパラメータを、フーリエ変換を行う回数 3833、窓幅 4800、シフト幅 600 と設定した。これにより、観測行列  $\mathbf{V}$  ( $2400 \times 3833$ ) を得た。この観測行列に対し、式 (3) を用いてアクティベーション行列の計算を行った。図 3 は固定した基底行列  $W_0$  に用いるために GarageBand からハイハット、スネア、バスドラムのスペクトルを抽出し図示したものである。

図 4, 5 は NMF によって得られたアクティベーション行列をタイト、ルーズそれぞれに対して、ハイハット、スネア、バスごとに図示したものである。

### 3.3 ピーク検出による打点時刻の計算

得られたアクティベーションからピークを検出するために、スネア、バスドラムのアクティベーションに対して、サンプル窓を後ろ 4 サンプルとし、1 サンプルずつフレームシフトさせた移動平均をとることで平滑化処理を行った。平滑化処理を行ったアクティベーションに対して、ピークを検出した結果が図 6, 7 である。

得られたピークに対してスネア 200000000、バス 60000000 と閾値を設定し打点を抽出した。

### 3.4 打点時刻のずれの計算

2.4 項で述べたように、ずれを測る基準を計算することで、打点時刻のずれを算出した。今回は BPM80 での演奏であるため、ずれを測る基準の 4 分音符の間隔は時間単位にすると  $60/80 = 0.75(\text{sec})$  となる。これをずれを測る基準とし、実際の打点時刻との差を打点時刻のずれとした。

### 3.5 確率モデルを用いた尤度によるタイト、ルーズへのグルーブ感評価結果

今回はテンポ 80 での演奏のため、モデル式 (5) のパラメータに用いる平均と標準偏差は表 3, 4 のようになった。そして、このモデル式への尤度は表 5 のようになった。

表 3 テンポ 80 における, タイトのモデル式のパラメータ

	スネア	バス
平均	0.004017857	0.012053571
標準偏差	0.006696429	0.013392857

表 4 テンポ 80 における, ルーズのモデル式のパラメータ

	スネア	バス
平均	-0.029464286	-0.019419643
標準偏差	0.009375	0.010714286

表 5 タイト, ルーズ評価値

	タイトな演奏	ルーズな演奏
タイトへの評価値	-2803.187033	-508.9759071
ルーズへの評価値	-3873.604362	-169.2979704

### 3.6 考察

表 5 よりタイトな演奏に対しては, ルーズへの評価値よりもタイトへの評価が大きくなり, ルーズな演奏に対してはタイトへの評価よりもルーズへの評価が大きくなる結果となった. これより本研究で扱った音源に対しては, タイト, ルーズな演奏の違いを正しく反映した評価を行えたといえる. しかしながら, 尤度計算を用いた評価のため値の大きさが正規化されておらず, この数値自体に絶対的な意味を与えることは難しい. あくまでも行った演奏が相対的にタイト, ルーズどちらに近いかを評価するための指針として解釈される.

## 4. 結論

本研究では, NMF を用いた音源分離手法と, タイト, ルーズそれぞれに対してスネア, バスドラムの打点時刻のずれを確率変数とした確率モデルの定義を組み合わせることによって, 同時録音ドラム演奏音源からタイト, ルーズへの評価を実現した.

本研究の目的は新しい手法の提案にあるため, 限定的な演奏状況を仮定した. メトロノームを聞いていない状況でもずれを測る基準を算出することができれば, メトロノームを聞いていない状態での演奏に対してもグルーブ感への評価を行うことができる. また, モデル化するグルーブ感にはタイト, ルーズのみを扱ったが, 他にもジャストなどが知られているため, このモデル化も検討される. そして, 最も基本的なスコアによる演奏でのみ評価実験を行ったため, 今後その他の演奏スコアに対する評価実験も行っていく.

**謝辞** 本研究は JSPS 科研費 17H00749 の支援を受けた. また, 本研究を進めるにあたり, 明治大学院先端数理科学研究科の嵯峨山研究室のゼミに参加させて頂き, 様々な貢献を頂いた. 嵯峨山茂樹先生および嵯峨山研究室の皆様へ, 謹んで感謝の意を表す.

## 参考文献

- [1] C. Keil: "Motion and Feeling through Music," The Journal of Aesthetics and Art Criticism, Vol.24, No.3, pp. 337-349, (1966).
- [2] S. Kawase and K. Eguchi: "The Concepts and Acoustical Characteristics of 'Groove' in Japan," PopScriptum 11 The Groove - Issue, pp.1-45, (2010).
- [3] 奥村啓太, 平田圭二, 片寄晴弘: "ポップス系ドラム演奏の打点時刻及び音量とグルーブ感の関連について" IPSJ SIG Technical Report MUS, pp.21-27, (2004).
- [4] Y. Ohya, K. Nakamura, T. Tokunaga: "Extraction of groove feelings from drum data using non-negative matrix factorization" IEEE SCIS-ISIS, pp.125-130, (2012).
- [5] D. Lee, H. Seung: "Algorithms for Non-negative Matrix Factorization" NIPS, (2000).