

# 発音時刻とアクセントパターンに基づくグルーヴの達成度評価システム

上嶋 秀典

Ueshima Syusuke

法政大学情報科学部デジタルメディア学科

E-mail syusuke.ueshima.9k@stu.hosei.ac.jp

## Abstract

‘On-top’ and ‘laid-back’, types of groove are musical expression make good impression for a listener. A score does not have information of onset time lags and accents, A drummer cannot play with groove well. Therefore, proposing the training system invests this information to a score and, a user imitates the target performance. For onset time detection, using time variation of High Frequency Contents, and calculate lags with imagination onset time based on tempo. For accents decision, making a model of high-hats beating, and we treat strong cases as accents. There were three subjects to use this system, and achievements before the training and after one were compared. In result, achievements after one increased average xx percent more than achievement of three subjects not use this system. To sum up, training with this system is effective to imitate onset time lags and accents.

## 1 序論

リズムに関する音楽表現のひとつとして、グルーヴと呼ばれるものがある。これは、各個人が他の奏者のグルーヴを聞きながら合わせることによって生まれる。グルーヴに関する従来研究として、タイトやルーズと呼ばれるグルーヴを定量的に扱った研究がある [1]。奥平らによれば、タイトはまとまりや明瞭さを聴衆に感じさせ、ルーズはゆったりとした滑らかな印象を聴衆に感じさせる。また、これらのグルーヴは打点時刻や音量を意識したものである。そこで本研究では、ドラム演奏の発音時刻のズレとアクセントに着目する。

ドラム奏者が、楽譜をもとにしてプロの奏者の演奏を模倣する際、楽譜に発音時刻のズレについての情報が明記されていない。また、アクセントについても情報がない場合もある。これは、プロのドラム奏者が演奏楽譜にない表現を行うためである。そのため、ドラム奏者どのようにして発音時刻のズレやアクセントを付けてよいかわからないため、このような特徴を持った演奏を行うことが難しい。また、発音時刻にズレが生まれることによって、音価が等間隔でなくなってしまうことで、奏者が意図したズレを作る必要がある。これは、聴覚の等時性があるため難しい。アクセントに関しても、複雑なパターンになるに連れて演奏の難易度は高くなる。しかし、後者の問題点は発音時刻のズレやアクセントのつけ方がわかっていなければならない。よって、楽譜に発音時刻のズレやアクセントの情報が明記されていない問題を先に解決する必要がある。

そこで、本研究ではドラム奏者を対象としたグルーヴの達成度判定システムを提案する。ドラム演奏の発音時刻のズレとア

クセントパターンによるグルーヴのモデルを作る。目標音源とシステム利用者の演奏のグルーヴのモデルを比較し、達成度の判定、楽譜に可視化を行うことで練習支援を行う。ドラムはバンド演奏のリズムの軸であることから、ドラム奏者を対象とする。

## 2 発音時刻のズレとアクセントパターンに基づくグルーヴのモデル化

発音時刻のズレとアクセントパターンによるグルーヴのモデル化を行う。発音時刻のズレについて利用者は、実際の発音時刻が楽譜どおりの発音時刻から、どの程度ずれているかを知ることが重要である。そこで、発音時刻推定によって求めた実際の発音時刻と、その平均から推定した奏者の意図した発音時刻(以下、仮想発音時刻)を、小節の音符の位置毎に求める。

アクセントパターンは、ロックドラムではハイハットにアクセントをつけることが多いことからハイハットに付くアクセントを対象とする。アクセントは強弱の2パターンとし、音符毎に求める。

### 2.1 発音時刻のズレの計測

#### 2.1.1 beat spectrum を用いたテンポ推定

楽譜上の発音時刻の決定のために、音響信号のテンポ推定を行う。本研究では、Foote らが提案する mfcc[3] による音の微量の自己相関である beat spectrum[4] を用いてテンポ推定を行う。計算速度向上のために、44.1kHz の音響信号を 16kHz にダウンサンプリングする。

図.1 は、4 小節のドラム演奏音響信号に対して FFT 長を 512 点、フレームシフト長を 256 点として beat spectrum を求めたものの一部である。図.1 で、青い曲線は beat spectrum の

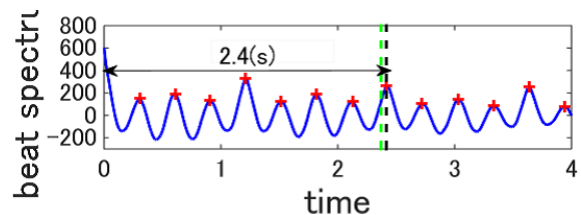


図 1. ドラム演奏の beat spectrum

変動、赤い点は 1 小節の長さの候補、緑の破線は、楽譜上のテンポから計算した楽譜 1 小節の長さ、黒の破線は実際のテンポから計算した 1 小節の長さを示す。beat spectrum からピークを検出する。より大きなピークから 4 分音符の長さよりも距離が小さいピークを除き、残ったピークを 1 小節の長さの候補列とする。楽譜の情報からテンポは既知である。そのため、beat spectrum の候補列の中で、正確な 1 小節の長さに近い候補を選択することで、演奏音源の実際のテンポの推定を行う。図.1

の音響信号の実際のテンポは 98.5(bpm) であった。

$i$  小節目の 16 点で分割したときの  $j$  点目の音価を  $b_{i,j}$ 、16 分小節の長さを  $d$  とする。このとき、発音タイミングであれば  $b_{i,j} = 1$ 、そうでなければ  $b_{i,j} = 0$  とする。発音タイミング  $\mathbf{b} = b_{1,1}, b_{1,2}, \dots, b_{n,m} (n = 8, m = 16)$  とすると、楽譜上の発音時刻列  $\mathbf{y} = y'_{1,1}, y'_{1,2}, \dots, y'_{n,m}$  は、

$$\mathbf{y} = 0, db_{1,2}, 2db_{1,3}, \dots, (nm - 1)db_{n,m} \quad (1)$$

と表される。

beat spectrum によるテンポ推定の評価実験を行った。打ち込みによって作った 4 小節と 8 小節のドラム演奏を 5 パターン用意しテンポ推定を行った (表.1)。5 パターンの演奏のテンポ推定し、実際のテンポとの相対誤差を求めた。相対誤差は最大

表 1. テンポの相対誤差

	演奏 A	演奏 B	演奏 C	演奏 D	演奏
相対誤差	0.002	-0.004	-0.004	-0.003	-0.003

で 0.004 であったが、本システムにおいて悪影響はなかった。

## 2.2 HFC の時間変動を用いた発音時刻推定

ドラムなどの打楽器は、発音の瞬間に高周波に急激な変動がある。そこで、打楽器の発音時刻推定に適した高周波成分に重み付けを行った特徴量 (HFC) の時間変動による方法 [5] を利用する。

短時間フーリエ変換によって得られたスペクトルを  $X_i$  とし、HFC を

$$h = \sum_{i=1}^N |X_i| \quad (2)$$

と定義する。これを音響信号に対してフレーム毎に行い、ピークの点を発音時刻の候補とする。サンプリング周波数 44.1kHz の音響信号に対し、FFT 長は 1024 点、フレームシフト長は 256 点とする。図.2 はドラム演奏の HFC の時間変動を求めたものである。

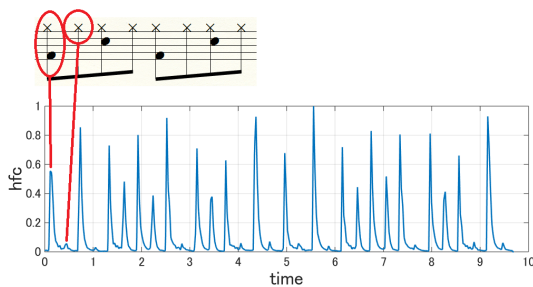


図 2. ドラム演奏の HFC の時間変動

図.2 では、楽譜上の各音符に対応してピークが発生している。HFC の値の大きさは楽器の違いや音量によって変わる。高周波が強い楽器や、音量が大きい場合に値が大きくなる。本研究が対象とする演奏は音符の最小単位が 16 分音符である。そこで、大きいピークから 16 分音符の長さよりも近いピークを無視し、残ったピークの発生時刻を各音価の実際の発音時刻の候補列  $\mathbf{a}$  とする。次に、候補列の各要素の時刻を小節の頭として楽譜上の発音時刻を決定する。ピークの点との誤差が最小になるよう楽譜上の発音時刻と対応づける。しかし、この時点で、小節の頭が間違っている可能性があるため、5 点目の候補まで繰

り返し、最も誤差が小さい組み合わせを選択する。図.2 上部は HFC の時間変動に対応した楽譜である。赤い線は、音符とピークが対応していることを表す。例えば、1 小節目の最初の音符の発音時刻はピークの点の時刻となる。音要素  $a_l$  を小節の頭としたときの楽譜上の発音時刻  $y_{i,j}$  と候補  $a_l$  のユークリッド距離を  $d(y_{i,j}, a_l)$  とする。

要素  $a_l$  を小節の頭としたときの楽譜上の発音時刻列  $\mathbf{y}_l$  と候補列  $\mathbf{a}$  の誤差  $e_l$  は、

$$e_l = \sum_{i,j} \min d(y'_{i,j}, a_l) \quad (3)$$

となる。このときの  $\mathbf{a}$  の組み合わせを実際の発音時刻  $\mathbf{r}$  とする。仮想発音時刻  $\mathbf{m}$  は、実際の発音時刻を  $\mathbf{r}$  を小節の  $j$  番目毎に平均をとることで求める。

$$m_j = \sum_i r_{i,j} \quad (4)$$

また、実際の発音時刻と仮想発音時刻の誤差を最小にするため、最小二乗法を用いて仮想発音時刻を修正する。このとき、テンポが変わらないよう、傾きは固定したまま最小二乗法を行う。このときの仮想発音時刻候補と実際の発音時刻候補のズレの大きさ  $e$  は、

$$e = \sum_{i,j} m_j - r_{i,j} \quad (5)$$

$e$  が最小になるときの  $\mathbf{m}$  と  $\mathbf{r}$  をそれぞれ、仮想発音時刻と実際の発音時刻とする。

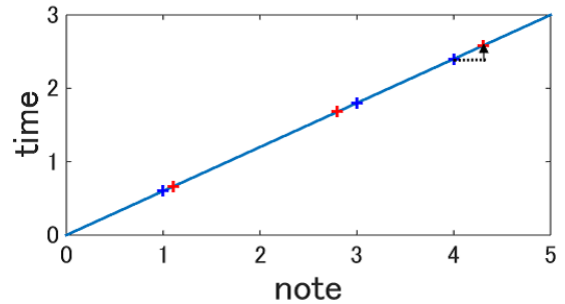


図 3. 仮想発音時刻と実際の発音時刻のズレ

図.3 は、仮想発音時刻と実際の発音時刻のズレを求めるイメージ図である。青線は、楽譜上の発音時刻で、この直線上に音符があった場合は、ジャストである。赤の点は仮想発音時刻であり、黒の矢印で表された楽譜上の発音時刻と仮想発音時刻のズレが発音時刻のズレとなる。

HFC の時間変動を用いた発音時刻推定の評価実験を行った。打ち込みによって作った 4 小節と 8 小節のドラム演奏を 5 パターン用意し発音時刻推定を行った。表.2 はその結果である。5 パターンの演奏の各音符の発音時刻を推定し、実際の発音時刻とのズレを求め、各音符の発音時刻のズレの平均を求めた。各音符の発音時刻のズレの平均は最大で 26.94(ms) であった。

表 2. 発音時刻検出のズレの平均

	演奏 A	演奏 B	演奏 C	演奏 D	演奏
平均 (ms)	- 0.83	22.42	-26.94	-22.24	21.41

このズレはテンポ推定時のズレから生まれたものだと考えられるが、本システムにおいて悪影響はなかった。

### 2.3 アクセントパターンの推定

アクセントパターンの推定を行うために、各発音時刻でのハイハットパワーの大きさを推定する。そのため、スネアドラムやバスドラム等の他の打楽器の影響を小さくする。そこで前処理として、音響信号に 4000Hz のハイパスフィルターをかける。

短時間フーリエ変換によって得られた各フレームのスペクトルを  $X_i$  とするとそのフレームのパワーは

$$p = \sum_{i=1}^N |X_i| \quad (6)$$

で求められる。これをサンプリング周波数が 44.1kHz の音響信号に対して、FFT 長 1024 点、フレームシフト長 256 点のフレーム毎に求めることで、パワーの時間変動を求める。大きいピークから 16 分音符の長さよりも近いピークを無視し、残ったピークの大きさを各音価のパワーとする。また、発音時刻同様、各音符と各ピークとの対応関係がわかっていない。そこで、実際の発音時刻との近傍点を取っていくことで対応付けを行う。

図.4 は、ドラム演奏の振幅の時間変動を表している。図.4 で、

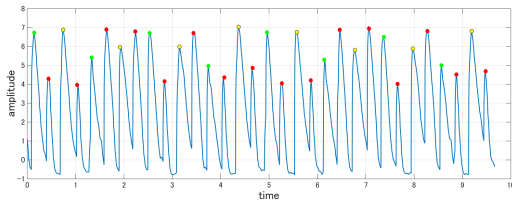


図 4. ドラム演奏の振幅の時間変動

青線はドラムの振幅の時間変動を、丸はそれぞれドラムの振幅の時間変動のピーク、つまり各音符でのパワーである。赤の点はハイハットのみ、緑の点はハイハットとバスドラム、黄色の点はハイハットとスネアドラムが鳴っている。

次に、アクセントの強弱を混合ガウス分布を用いたクラスタリングを行うことで判定する。しかし、スネアドラムやバスドラムの音がハイハットと重なっている場合、ハイハットの場合に比べてパワーが大きくなってしまふ。そこで、ハイハットのみ、ハイハットとスネアドラム、ハイハットとバスドラム、これら 3 つの打楽器が同時になっている 4 つの場合に分ける。それぞれの場合について、音符のパワーをこれら 2 個の分布の混合分布であると仮定し、EM アルゴリズムを用いてモデルの推定を行う。このモデルを基にアクセントの有無を判定する。

パワーの大きい分布をアクセントを付けて叩いた音符、パワーの小さい分布をアクセントをつけないで叩いた音符とする。この分布を元にクラスタリングを行う。4 つのパターンで同様の処理を行い、結果を統合する。

アクセントパターンの推定の評価実験を行った。8 小節のドラム演奏を 5 パターン用意しアクセントパターンの推定を行った (表.3)。各演奏のアクセント位置の検出精度を F 値で求めた。演奏 B は F 値が 1.00 であり、すべてのアクセントを検出

表 3. アクセント位置の推定精度の F 値

	演奏 A	演奏 B	演奏 C
F 値	0.71	1.00	0.83

できたが、演奏 A と演奏 C について全てのアクセントを検出出来なかった。検出出来なかったアクセントでは全てハイハットとスネアドラムが同時になっていた。聴覚上ハイハットのパワーの違いがわかりにくいことから、本システムでは除外する。これらのことから、本システムでは悪影響のない結果であった。

### 3 グルーヴの達成度評価システム

ドラム奏者のグルーヴのある演奏会得のために、楽譜を用いた発音時刻のズレとアクセントの強弱パターンの可視化と達成度の評価を行う。竹川らのリズム学習に関する演奏学習支援の研究では、音長情報を楽譜に可視化することによってシステム利用者の音長理解に効果を得ている [6]。本システムについても、音長情報を扱っているため、楽譜に可視化するシステムを用いる。

発音時刻のズレについては音響信号全体で行い、平均の値を判定する。強弱パターンについては、各音価毎に行い、目標音源とシステム利用者の演奏それぞれをアクセント記号を用いて表示を行う。達成度の評価は発音時刻のズレとアクセントパターンで別に行う。発音時刻のズレは目標音源の発音時刻のズレとシステム利用者の演奏の発音時刻のズレとの MAE で評価を行う。MAE は 0 に近ければ達成度が高くなる。アクセントパターンについては、F 値を用いて評価を行う。ここで用いる F 値は適合率を  $p$ 、再現率を  $r$  としたとき

$$F = \frac{2pr}{p+r} \quad (7)$$

とする。アクセントは全て正確な位置で叩く必要があるため、適合率と再現率はどちらも重要である。よって F 値は上記のものを利用する。F 値が 1 に近ければ達成度が高くなる。

図.5 は、練習システムの GUI の一部である。実際は、発音時刻のズレを 1 小節分 (平均と分散)、アクセントパターンを 8 小節、それぞれの達成度を表示しているが、図.5 では発音時刻の揺らぎとアクセントパターンを表示する部分を抜粋している。

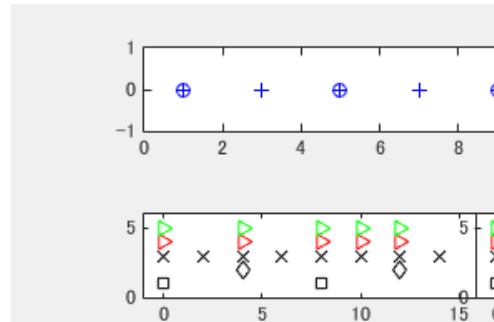


図 5. 発音時刻の揺らぎとアクセントの強弱パターンの可視化 (一部)

図.5 上部は発音時刻のズレを表している。青の点は仮想発音時刻を表している。また、青の丸で囲まれた点は拍を表している。目標演奏の発音時刻のズレの平均は赤の点、利用者の演奏の発音時刻のズレの平均は緑の点で表される。また、発音時刻のズレの分散については、各平均を表す点の同色の縦線で表される。

図.5 下部はアクセントパターンを表している。黒の四角はバスドラ、黒のひし形はスネアドラム、黒のバツはハイハットを表している。目標音源のアクセントは赤の三角で、利用者の演奏は緑の三角で表される。

音響信号は、4小節か8小節のロックドラムの演奏による音響信号を対象とする。ドラムの構成は、ロックで良く使用されるハイハット、スネア、バスドラムとする。また、システム利用者は、楽譜通りにドラムを演奏出来ることを前提とする。システム利用者は、グルーヴの達成度評価を行いたい音響信号と、目標とするグルーヴを持った音響信号、これらに対応した楽譜を入力する。ここで、楽譜とは、テンポ、使用楽器、各音価がわかるものである。システムは楽譜とそれぞれの音響信号を基に、発音時刻のズレの計測とアクセントパターンの計測を行う。次に、目標音源と利用者の演奏それぞれの発音時刻のズレとアクセントパターンを楽譜に付加することで、可視化を行う。システム利用者は、その楽譜を見ることで目標演奏との発音時刻のズレやアクセントの位置を修正する。システム利用者は、練習とシステムによる解析を繰り返し行い、発音時刻のズレの達成度が小さくなるよう、アクセントパターンの達成度が大きくなるよう演奏を目標演奏に近づけていく。

#### 4 達成度評価システムの評価実験

本練習システムで提案した発音時刻のズレとアクセントパターンの可視化と達成度評価による練習システムが目標音源のグルーヴの模倣に効果的であるか評価実験を行った。

被験者3人に、発音時刻のズレがあるパターンとアクセントがあるパターンの2つのドラムパターンで、次の練習を行ってもらった。始めに、模倣を行ってもらい演奏の楽譜を渡し、1時間練習をしてもらった。発音時刻のズレとアクセントパターンの模倣が目的であることを伝えた。プロの奏者の演奏を聞いてもらった。メトロノームを聞きながらプロ奏者の演奏の模倣を行ってもらい録音を行った。これを練習前の演奏とする。次に、楽譜、プロ奏者の演奏音源、メトロノーム、本システムを利用して1時間練習を行ってもらった。練習後に同様に演奏を模倣してもらい録音を行った。これを練習後の音源とする。録音では楽譜を2週叩いてもらい、2週目の演奏を解析した。3人の被験者はドラムを含めた打楽器の経験者で、奏者Aの経験年数はそれぞれ、ドラム3年、マーチングドラム10年、奏者Bの経験年数はそれぞれ、ドラム5年、パーカッション3年、奏者Cの経験年数はドラム3年である。

練習前の演奏の達成度と練習後の演奏の達成度を本システムで判定を行った結果が表4である。

表4. 評価実験による練習前後の2拍目の裏の発音時刻のズレの平均と標準偏差 (ms)

	奏者 A		奏者 B		奏者 C	
	前	後	前	後	前	後
平均	-22.10	-19.95	-20.92	-18.18	-42.43	-56.32
標準偏差	13.96	18.84	42.43	20.86	19.73	24.84

表4は、奏者三人にシステムを用いて練習をしてもらった前後の発音時刻のズレで2拍目の裏の平均と標準偏差を求めたものである。2拍目の裏は、目標演奏の中で大きく発音時刻が遅れている部分であったため着目した。練習前後の平均を比較すると、奏者Aと奏者Bはそれぞれ平均が2.15(ms)、2.74(ms)小さくなった。それに対し、奏者Cは平均が13.89(ms)0大きくなった。練習前後で平均に差があるか練習前の達成度と練習後の平均でt検定を行った。結果、練習前後の母平均は違ふとみなせた。よって、奏者A、Bについては練習後の方が平均が小さくなったため、練習後の達成度は練習前に比べて良くなっ

たといえる。また、奏者Cについては練習後の平均が大きくなったため、練習後の達成度は練習前に比べて悪くなったといえる。

アクセントのあるドラムパターンについて奏者A、B、Cの三人にシステムを用いて練習をしてもらった。練習前後の達成度(F値)を比較すると、奏者Aと奏者Cは0.96から1.00となり、0.04増加した。それに対し、奏者Bは1.00から1.00と変化がなかった。達成度の増加は平均で0.026であった。練習前後で達成度に差があるか練習前の達成度と練習後の達成度でt検定を行った結果、練習前後の母平均は違ふとみなせた。よって、練習後の方が平均が大きくなったため、練習後の達成度は練習前に比べて良くなったといえる。

発音時刻のズレについては、発音時刻のズレの平均が小さくなった。発音時刻のズレの可視化による練習支援は有用であったと考えられる。しかし、奏者Aと奏者Cの2拍目の裏の発音時刻のズレを比較すると、奏者Cは目標演奏とのズレを小さくすることが出来なかった。奏者Cは奏者A、Bに比べてドラムの経験年数が短かったことから、熟練者は可視化による支援によって発音時刻の模倣の達成度が良くなるが、初心者などの熟練度が低い奏者は、可視化による支援では効果がない可能性がある。

アクセントパターンについては、F値の平均が大きくなったことから、アクセントパターンの可視化による練習支援は有用であったと考えられる。しかし、練習前後のF値の上昇率が小さかった。これは、アクセントは、音源を聞けばアクセントがどの位置にあるかわかることが原因であると考えられる。

#### 5 結論

本論文では、発音時刻のズレとアクセントパターンに基づき、利用者の演奏と目標演奏のグルーヴの差異をもとに達成度の表示、楽譜を用いて可視化を行う練習システムを提案した。このシステムを3人のドラム奏者に利用してもらった。結果、奏者二人について、2拍目の裏の発音時刻のズレの平均がそれぞれ、2.15(ms)、2.74(ms)だけ0に近づいた。また、アクセントパターンの達成度は0.026増加した。これらのことから、発音時刻のズレとアクセントパターンの模倣の補助が出来た。しかし、発音時刻のズレの模倣については効果がなかった被験者もいた。今後の課題として、熟練度の低い奏者が発音時刻の発音時刻のズレの達成度が良くなるよう、簡易なドラムパターンでの発音時刻のズレの再現を行うことによる練習補助を行う。

#### 参考文献

- [1] 奥平他, "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(第3報)-データの基礎的分析とドラム演奏生成システムの実装-", 情報処理学会音楽情報科学研究会研究報告, 2006-MUS-64, pp.53-58, 2006
- [2] 渡辺他, "ドラム演奏のグルーブ感の解析", 情報処理学会研究報告 2006-MUS-67, 2006
- [3] Rabiner L., et. al., "Fundamentals of Speech Recognition, Englewood Cliffs", NJ, 1993
- [4] J.Foote., et. al., "The Beat Spectrum: a New Approach to Rhythm", IEEE Int. Conf. on Multimedia and Expo, 2001
- [5] J.B.Bello., et. al., "A tutorial on onset detection in music signals.", IEEE Trans. Audio Speech Lang. Process., vol.13, no.5, pp.1035-1047, 2005.
- [6] 竹川他, "リズム学習を考慮したピアノ演奏学習支援システムの構築", 情報学会論文誌, 54巻, 4号, pp.1383-1392, 2013