

リアルタイム発音検出つきメトロノームを用いた リズム練習支援システムの基礎的検討

Fundamental Study for Rhythm Practice Supporting System with Electronic Metronome and Real-time Onset Detection

山田 昌尚^{*1}, 松尾 章弘^{*1}, 土江田 織枝^{*1}
Masanao YAMADA^{*1}, Akihiro MATSUO^{*1}, Orie DOEDA^{*1}

^{*1} 釧路工業高等専門学校

^{*1} National Institute of Technology, Kushiro College

Email: ym@elctro.kushiro-ct.ac.jp

あらまし: 本研究は管楽器などの単音を対象としてマイクからの入力をリアルタイムに発音検出し可視化することで、リズム練習を支援するシステムを構築するものである。このシステムは音楽初心者のためのリズム認識の練習や音楽経験者のリズム感向上のための練習などに使うことを目的としている。システム構築には ChuckK と Processing を使用する。本稿ではその基礎的検討としてオンセット検出とシステム全体の構成について述べ、楽器音での発音検出に関して行った実験について報告する。

キーワード: 演奏支援, 演奏の可視化, ChuckK, Processing, ASIO

1. はじめに

音楽演奏にかかる能力の多くは身体知のひとつである。演奏した音高を知る方法としてはチューナーが普及しているが、リズムに関してはそのような外部的な補助手段で客観的に数値化、可視化する方法は一般化していない。関連研究として、ピアノ演奏^(1,2) やドラム演奏⁽³⁾ に関するものがあるが、これらは MIDI 楽器などを発音検出の手段としており、管楽器等には適用できない。本研究は、単音の楽器音を対象として、演奏された発音タイミングをリアルタイムに可視化することで練習の補助となるシステムの構築を目指すものである。

2. オンセット検出

楽器の発音をオンセットといい、音響信号から発音のタイミングを得ることをオンセット検出という。オンセット検出は一般的に、前処理、検出関数、ピーク抽出の3段階からなる⁽⁴⁾。第1段階の前処理としては正規化を施す。第2段階の検出関数として、周波数ごとの信号スペクトル強度を利用する方法や、位相の変化を利用する方法など、各種が提案されている。本研究では、周波数ごとの信号スペクトル強度の変化が大きい場合にオンセットとみなすスペクトルフラックスを使用する。スペクトルフラックスは次式で表される。

$$SF(n) = \sum_{k=1}^{\frac{N}{2}-1} \max(0, |X(n, k)| - |X(n-1, k)|)$$

ここで N は FFT フレームのデータ数, n は時間, k は周波数であり, $X(n, k)$ はスペクトログラム (時間周波数信号) を表す。0 との max をとることでスペクトル強度が増加する場合のみを対象としている。第3段階のピーク抽出では、しきい値を超える局所最

大値 (local maxima) を検出し、その時刻をオンセットとする。しきい値として次式による動的しきい値を用いる。

$$TH(n) = \delta + \lambda \cdot \text{median}(SF(n - v_1 : n + v_2)) + \alpha \cdot \text{mean}(SF(n - v_1 : n + v_2))$$

ここで δ はしきい値の定数項, λ および α はそれぞれ中央値, 平均値に対する重みであり, v_1, v_2 は動的しきい値の対象幅を表す。このしきい値を用いて、検出関数 $DF(n)$ およびオンセット位置 $OD(n)$ を次のように求める。

$$DF(n) = SF(n) - TH(n)$$

$$OD(n) = \begin{cases} 1, & DF(n) > 0 \text{ and} \\ & \arg\max_{w_1 < m < w_2} d(m) = n \\ 0, & \text{otherwise} \end{cases}$$

ここで w_1, w_2 は local maxima の探索範囲である。

3. システム構成

図1にシステム構成を示す。構築環境は Windows である。電子メトロノーム音の生成とマイク入力からの信号処理を ChuckK で行い、表示には Processing を用いる。ChuckK はプリンストン大学で開発された実時間音楽処理用のプログラミング言語である。

マイクから入力した信号の A/D 変換にはフリーの ASIO エミュレーションドライバである ASIO4ALL

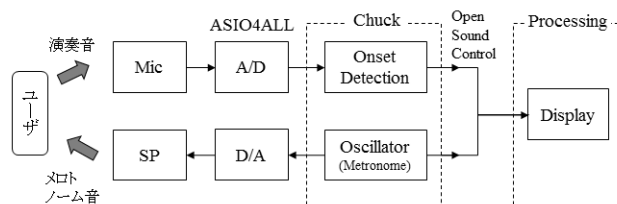


図1 システム構成

を使用する。これは、A/D 変換に Windows 標準のオーディオエンジンを使用すると 100~200ms のレイテンシーが生じるためである。サンプリング周波数は 44.1KHz でモノラル入力とし、この信号を 1024 サンプルのフレーム幅でハニング窓をかけて短時間フーリエ変換 (STFT) により時間周波数情報に変換した。STFT のホップ幅は 10ms であり、これがシステムの時間解像度となる。この信号を前節で述べた方法でオンセット検出した。

メトロノーム音は ChucK で生成し、スピーカから出力する。強拍と弱拍は周波数で区別する。システムのユーザはこのメトロノーム音を聴きながら楽器を演奏する。ChucK から Processing へメトロノーム音と検出したオンセットのタイミングを送るのには Open Sound Control (OSC) を用いる。OSC は電子楽器およびコンピュータ間で音楽演奏データ等を送受信するための通信プロトコルである。OSC では URL 形式でデータを送るため、メトロノームのタイミングとオンセットを容易に区別して情報伝達できる。

4. 実験

オンセット検出の動作を確認するため、2 種類の実験を実施した。なおすべての実験は室内残響の影響を排除するため無響室で測定している。

1 つめの実験は、スピーカから出力したメトロノーム音をそのままマイク入力し、リアルタイムで発音検出するものである。その結果、約 20ms の一定した遅延があるものの、100%の発音検出ができた。

2 つめの実験として、フルート、トランペット、ホルンの 3 種類の楽器で録音した音のオンセット検出を実施した。録音したパターンは変ロ長調で 1 オクターブのタンギングとレガートの演奏 (48 音、約 15 秒) である。前節で述べたパラメータは $v_1, w_1, w_2=50\text{ms}$, $v_2=0$ とした。正解とする発音タイミング (ground truth data) は入力信号の振幅エンベロープとスペクトログラムにもとづいて 3 名でアノテーションした時刻の平均を用いた。一般的には、この ground truth data に対して検出したオンセットが前後 25~30ms の範囲に入っているかどうかで認識率を評価することが多いが、本研究では前に 25ms、後に 55ms の範囲でオンセット検出されていれば True Positive と判断した。管楽器では、息を吹き込み始めた瞬間が ground truth として判断されるのに対して、実際に音として知覚されるタイミング、つまり音楽演奏として意味のある発音時刻はそれよりも遅くなることを考慮したためである。ただし、検出されたオンセット時刻と聴覚上の発音と判断され

るタイミングの関係は現時点では明らかでない。

それぞれの楽器でのオンセット検出結果を表 1 に示す。トランペットの再現率がやや低い、おおむね検出できているといえる。これらのオンセット検出率を得るためのしきい値に関するパラメータ (δ, λ, α) は、楽器ごとに異なることがわかった。

図 2 はフルートの最初の 3 秒についてのオンセット検出の様子である。上から、(a) 入力信号、(b) スペクトログラム、(c) スペクトルフラックス SF (実線) およびしきい値 TH (点線)、(d) 検出関数 DF (点線) とオンセット位置 OD (実線)、(e) ground truth data との比較、をそれぞれ示している。フルートは他の管楽器に比べて音の立ち上がりが遅いため、ground truth から後に 55ms という判断基準でも 7 音目および 9 音目が誤検出になっている。

5. おわりに

楽器でのリズム練習支援システムの開発を目的としてオンセット検出を行った。実験により、検出に適したパラメータは楽器によって異なることがわかった。このパラメータを簡便に設定する方法について検討することが今後の課題である。また、リズム練習システムとして使用するために、検出関数によって得られるオンセット位置と聴覚上発音として感じられるタイミングの関係についても詳細に調査する必要がある。

参考文献

- (1) 辻靖彦, 西方敦博: “リズムと打拍フォームに基づく打楽器学習支援システムの開発と評価”, 電子情報通信学会論文誌 Vol.J88-D1, No.2, pp.508-516 (2005)
- (2) 岩見直樹, 三浦雅展: “MIDI 楽器を用いたドラム演奏練習支援システムの提案”, 情報処理学会研究報告[音楽情報科学] pp.85-90 (2007)
- (3) 竹川佳成, 寺田努, 塚本昌彦: “リズム学習を考慮したピアノ演奏学習支援システムの設計と実装”, 情報処理学会論文誌, 第 54 巻, 第 4 号, pp. 1383-1392 (2013)
- (4) Bello, Juan Pablo, et al.: “A tutorial on onset detection in music signals”, Speech and Audio Processing, IEEE Transactions, Vol.13, No.5, pp.1035-1047 (2005)

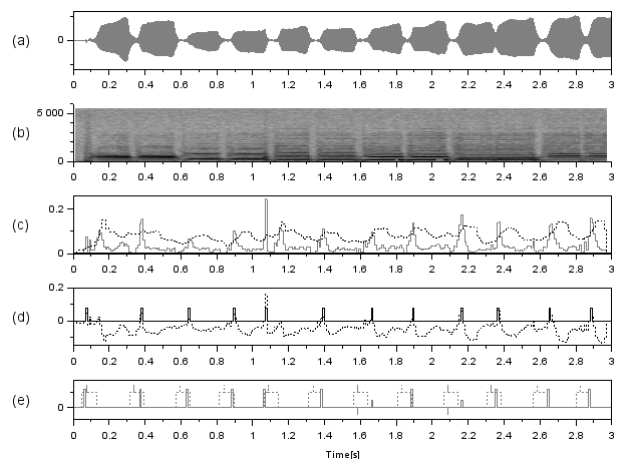


図 2 フルートのオンセット検出

表 1 楽器でのオンセット検出結果

楽器	適合率	再現率	F 値
Flute	0.900	0.938	0.918
Trumpet	0.951	0.796	0.867
French Horn	0.941	1.000	0.970