

短報

12平均律を用いた一般化調和解析による打楽器音の音高抽出の試み

田中 元志**

A Pitch Extraction of Percussion Sound by Generalized Harmonic Analysis with Twelve-Tone Equal Temperament

Motoshi Tanaka**

Abstract

Humans often recognize someone's footsteps, though the footstep sound contains many noise and partial tone components. Also, the high and low of footsteps can sometimes be recognized, which indicates the footsteps have a pitch. It may be similar to drum and percussion sound. In this paper, a pitch extraction of those percussion sounds was discussed by GHA (Generalized Harmonic Analysis) with twelve-tone equal temperament. And each pitch of the sound was subjectively evaluated by 4 subjects. In results, the pitch felt by subjects was included in frequency components extracted by the GHA.

1. はじめに

ヒトは足音で歩行者を特定できる場合があり, 足音に関する検討が行われている⁽¹⁾. 歩行者を識別できれば, 見守りやセキュリティシステムへの応用が期待できる. ヒトが足音を識別する要素の一つとして, 音高や音色の時間的な変化が考えられる. しかし, 足音には雑音成分が多く含まれている. 筆者らは, 計算量が多いが, 離散フーリエ変換に比べて周波数分解能が高く, 雑音低減特性を持つ一般化調和解析(GHA)⁽²⁾の利用を検討した⁽³⁾. その結果, いくつかの倍音構造を持つ周波数成分を抽出できた. 音高に着目すると, 抽出周波数を12平均律の音高に限定することが考えられる. 計算量の削減も期待できる.

本研究では, 足音の音高抽出の前段階の検討として, 雑音で, ヒトが音高の違いを比較的に感じることもできる打楽器音(ドラムなど)を例に, 12平均律を用いたGHAによる音高抽出を試みた. 本稿では, その解析方法と抽出結果について述べる.

2. 12平均律を用いた一般化調和解析

解析方法を以下に示す. ①~⑤は, 区間長 N で観測された信号 $x(t)$ に対するGHAアルゴリズムである⁽²⁾.

① フーリエ係数 $A(f)$ と $B(f)$ を計算する. ここで, f は

周波数, $T = 1/f$, $N \geq nT$, n は整数である.

$$A(f) = \frac{2}{nT} \int_0^{nT} x(t) \sin(2\pi ft) dt \quad (1)$$

$$B(f) = \frac{2}{nT} \int_0^{nT} x(t) \cos(2\pi ft) dt \quad (2)$$

② 原信号 $x(t)$ と, $A(f)$ と $B(f)$ で合成される信号の残差エネルギー $E(f)$ を求める.

$$E(f) = \int_0^N e(t, f)^2 dt \quad (3)$$

$$e(t, f) = x(t) - A(f) \sin(2\pi ft) - B(f) \cos(2\pi ft)$$

③ f を変えて①②を繰り返し, $E(f)$ が最小になる周波数 f_1 を抽出する.

④ 原信号から③で求めた周波数成分を取り除く.

$$x'(t) = x(t) - x_1(t) \quad (4)$$

$$x_1(t) = A(f_1) \sin(2\pi f_1 t) - B(f_1) \cos(2\pi f_1 t)$$

⑤ 残差信号 $x'(t)$ が0 ($x'(t) \cong 0$) になるまで①~④を繰り返す. 抽出された周波数を f_k とすると, パワースペクトルは $P(f_k) = A(f_k)^2 + B(f_k)^2$ で求められる.

本研究では, 12平均律の音高に対応する周波数

$$f_m = f_R \times \sqrt[12]{2^{m-m_R}} \quad (f_R = 440 \text{ Hz}) \quad (5)$$

のみを抽出対象とする(m : 音階番号, $m_R = 69$). ⑤の繰り返しを F 回とした. 大きい振幅を持つ周波数成分が順に F 個だけ抽出される. 条件によるが, 計算時間は1 Hz毎のGHAに比べて1/30以下となった.

3. 打楽器音の音高抽出の試み

打楽器音として, RWC 研究用音楽データベース⁽⁴⁾のドラム音など7音(標本化周波数 44.1 kHz, 16 bit

2018年7月31日受理

** 秋田大学大学院理工学研究科数理・電気電子情報学専攻, Department of Mathematical Science and Electrical-Electronic-Computer Engineering, Graduate School of Engineering Science, Akita University

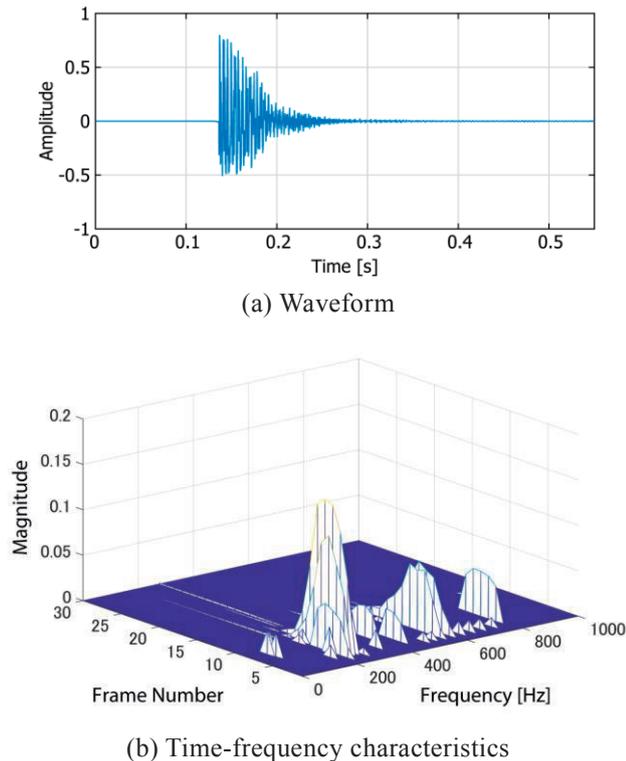


Fig.1 Example of the result (Snare Drum).

量子化)を用いた。GHAの解析周波数範囲をE0(41.2 Hz, $m = 28$)~B8(7,902 Hz, $m = 119$)とし、抽出する周波数(音高成分)の数を10個とした。また、解析フレーム長100 ms, フレーム周期20 msとした。

解析結果の例(スネアドラム)をFig. 1に示す。連続5フレーム以上で抽出された成分の中で、ピークの振幅が最も大きい成分を音高候補①, 継続時間が最も長い成分を候補②, 倍音構造を有する基音成分(複数存在した場合は、ピークの振幅が最も大きい基音)を候補③として抽出した。また、絶対音感を持たない楽器演奏の経験がある被験者4名にA~G#の純音を聞かせて、打楽器音の音高を答えさせる主観評価実験を行った。結果をTable 1にまとめる。表中の主観評価は多数決による音高である。候補①~③の音高と一致していない場合が多いが、被験者が感じた音高がGHAによって抽出された音高成分の中に含まれていた割合は約97%であった。音高に関連する成分を抽出できていると考えられる。また、評価結果の傾向から、比較的継続時間が長い音では、短時間で振幅が大きい成分より、ある程度の振幅を持ち長く継続している成分が音高として認識されやすい可能性が考えられる。例えば、Fig.2に示すハイタムの場合、約10フレーム目に大きなピークを持つ成分GやG#ではなく、F#が主観評価で選ばれた。

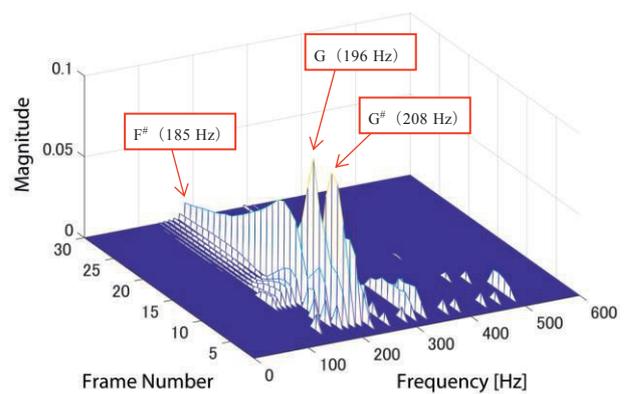


Fig.2 Time-frequency characteristics of High Tom.

Table 1 Extracted and subjectively evaluated pitch.

	候補①	候補②	候補③	主観評価
スネアドラム	B	C	G	D
バスドラム	C#	C#	E	C#
ハイタム	G	G	E	F#
ミッドタム	D#	D#	C#	C#
ロータム	C#	A#	A#	A#
木魚	G	G	F	G#
ハンドクラップ	G#	E	E	D

4. まとめ

解析周波数を12平均律の音高に対応させたGHAの利用を提案し、打楽器音の音高抽出を試みた。今後、本方法の利用、および足音の音高抽出について検討する予定である。最後に、実験に協力頂いた本学卒業生 吉原駿氏, 今井祥太氏, ならびに被験者の方々に感謝する。また、本研究の一部はJSPS科研費(15K00746, 18K02236)の助成で行われた。

参考文献

- (1) 例えば, 磯村太郎, 野村祐基, 板井陽俊, 安川博 (2008): 屋内環境下における歩行足音の検出手法に関する一検討, 信学技報, SIS2007-81.
- (2) 牛山聡, 東山三樹夫, 飯塚昌弘, 平田能睦 (1994): 一般調解析による波形分析, 信学技報, EA93-103.
- (3) 島崎興平, 田中元志 (2014): 一般調解析による足音の特徴抽出に関する検討, 平成26年東北地区若手研究者研究発表会, YS-12-A3.
- (4) Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R. (2003): RWC Music Database: Music Genre Database and Musical Instrument Sound Database, *Proc. ISMIR2003*, pp. 229-330.