Vol.2016-MUS-113 No.16
2016/10/14

音高系列の標準偏差を用いた 基本周波数のピーク分類と弾弦時刻・演奏区間の抽出

竹渕 瑛一[†1,a)]　梶並 知記[†2]　徳弘 一路[†3]　速水 治夫[†3]

概要：本研究では，音高系列の標準偏差から基本周波数や倍音のピークを分類し，音高系列と標準偏差の振幅比の時間変化から弾弦時刻と演奏区間の抽出を行った結果を報告する．音高系列とは，パワースペクトルを音高の周波数に対応させた系列である．先行研究では，基本周波数の検出は正確であるが処理時間が長い．弾弦時刻と演奏区間の抽出も基本周波数の検出手法では不可能である．本論文の手法は，普遍的な手法をそれぞれ共通して用いているため，計算リソースの効率化が期待できる．実験の結果，エレキギターの単音演奏では，基本周波数のピークを確実に分類し，弾弦時刻と演奏時間を正確に抽出できることを示した．

Using MIDI Sequence's Standard Deviation Classifying a Fundamental Frequency And Extracting a Picking Time and a Playing Time

Eiichi Takebuchi[†1,a)]　Tomoki Kajinami[†2]　Ichiro Tokuhiro[†3]　Haruo Hayami[†3]

Abstract: Our study classified a fundamental frequency and overtones from MIDI sequence's standard deviation. And, our study extracted a picking time and a playing time from an amplitude ratio by the MIDI sequence and MIDI sequence's standard deviation. The MIDI sequence is an array of indexing a power spectrum to the pitch frequency. Previous studies detects a fundamental frequency, correctly. However, previous studies are not faster than the realtime. Previous studies impossible detects a picking time and a playing time. Our study is a single method for a fundamental frequency, overtones, a picking time and a playing time. Our study resolved each output from MIDI sequence's standard deviation. Therefore, our study is able to promise to make efficient computing resources. Our evaluation experiment was picking electing guitar. As a result, our study classifies fundamental frequency's peak, correctly. And, our study extracts a picking time and a playing time, correctly.

1. はじめに

標本から大きく外れている統計値として外れ値 [1] が知られている．外れ値は，株価の急騰や暴騰などの景気予測や，工場の生産ラインでの不良品検知など，極端な値を検出するために用いられている．

楽器音の周波数スペクトルには，基本周波数とその整数倍の倍音が必ず含まれており，それらはピークとして現れる．本研究では，基本周波数や倍音のピークを外れ値として考え，効率的にピークを分類する方法を検討した．

標準偏差は確率や統計で標本のばらつき具合を求めるために利用されている．標準偏差は計算機が苦手である除算

†1 現在，神奈川工科大学大学院博士後期課程
Presently with Doctor's Course of Kanagawa Institute of Technology
†2 現在，岡山理科大学
Presently with Okayama University of Science
†3 現在，神奈川工科大学
Presently with Kanagawa Institute of Technology
a) nanashi4129@gmail.com

や累乗を何度も用いず，2 回のループだけで簡単かつ速やかに求められる．実装も簡単であり，多くの数学ライブラリで実装されている．

また，周波数領域から標準偏差を求めるのは非効率であるため，パワースペクトルを音高の周波数に対応させた音高系列を用いる．MIDI でも 128 音しか利用されておらず，楽音を対象とした高速フーリエ変換（Fast Fourier Transform：FFT）では音高系列を用いたほうが効率的である．

本論文では，エレキギターの楽器音を対象に，音高系列の標準偏差を閾値として，音高系列に含まれる基本周波数や倍音のピークが分類できるか報告する．また，振幅比の時間変化から，弾弦時刻や演奏区間を抽出できるか報告する．振幅比とは，音高系列とその標準偏差の比である．

本研究の貢献は，統計において標準的に用いられている標準偏差で，音高系列から複数の情報の抽出が行えたことである．特に，弾弦時刻や演奏区間の抽出は，他の手法との組み合わせが必要である [2][3]．ピークの分類や演奏区間の抽出などを同じ方法で解決できれば，解決に必要な計算資源（処理時間やメモリ量など）が少なくなることが期待できる．

本論文で報告する手法は，瞬時に基本周波数や弾弦時刻や演奏区間を判定したい場合に有用である．音高系列はわずか 128 音しかなく，その標準偏差を求めるのに必要な処理時間はごく僅かである．

2. 研究背景

ソフトウェアにおけるリアルタイム性とは，定められた時間内に制御が完了することである．

例えば，ゲームでは 1 秒間に 60 フレームの更新があり，毎フレームごとに描画までの処理を終わらせる必要がある．1 フレームあたり 16.7 ミリ秒の猶予があるものの，ゲームでは様々な機能が複合的に実行される．一つのアルゴリズムで処理時間の割合が大きくなると，他の機能や描画に割くための処理時間が少なくなる．

音楽に関しても，人間の可聴域が 15〜18kHz 程度であることから，44.1kHz のサンプリング周波数が一般化されている．人間は 22kHz の周波数は聴取できず，44.1kHz の遅延を感知することは不可能である．人間と計算機が同期的に演奏を行うには，その処理時間の猶予はわずか 45 マイクロ秒未満であると考えられる．

一方で，Desktop Music（DTM）の作曲環境の遅延は人間系と比較するとシビアではない．サンプリング周波数が 44.1kHz，バッファ・サイズが 256 点で遅延は約 5.8 ミリ秒となるが，演奏を聴き返す用途では深刻な遅延にはならない．

これらのことから，用途によってリアルタイムの定義が変わるものの，いずれにせよミリ秒単位の極めて短い時間で処理を完了させる必要がある．先行研究では，まだゲーム程度の処理時間に迫るアルゴリズムはなく，単純でも高速で確実に演奏情報を得られる研究が求められている．

本研究の目的は，楽器演奏者が一般的に用いている環境において，リアルタイムでエレキギターの楽音の情報を抽出することである．楽音には，基本周波数や倍音構造，弾弦時刻や演奏区間が含まれている．これらを短い時間で抽出できれば，演奏情報に関する様々な応用が考えられる．

例えば，演奏練習支援や即興演奏支援である．文献 [4] では，運指の最適化を行った上で演奏方法を逐次提案する練習システムである．文献 [4] は，MIDI ギターが前提である必要がある．Rock Smith[5] は，エレキギター向けの演奏の正確性を競うゲームである．Rock Smith は専用ケーブルを繋ぐだけで，演奏者が所有しているギターで遊べる特徴がある．しかし，適当に演奏しても加点されるため，演奏内容の評価には至っていない．

これらの演奏練習支援や即興演奏支援には，音響信号から高速かつ正確に演奏情報を得るようなアルゴリズムが必要である．

周波数領域上で演奏された楽音を推定する研究は既に数多く取り組まれている [6][7][8][9][10][11]．楽音の検出・推定の精度が目覚ましいほど向上しているが，先行研究の問題点に処理時間の長さと弦楽器に対する有効性がない点が挙げられる．

単音であれば MPM[8] が高速に基本周波数を検出できるが和音に対しては有効ではない．文献 [12][13] では，高速かつ和音に対する有効性が報告されているが，あくまでベース音の検出のみに特化している．

弦楽器は弦長と弾弦位置によって倍音構造が変化する．また，弦楽器は弦の剛性によって倍音周波数がわずかに変化する性質がある [14]．

Harmonic Product Spectrum(HPS[6]) は，倍音系列の内積で表現されるため他の手法と比べて処理時間が短い特徴がある．一方で，インハーモニシティ [14] が発生する条件下では適用できないため，弦楽器に対しては有効ではない．

Specmurt 分析 [7] は共通調波構造と観測された周波数スペクトルとの畳み込みで倍音を除去する手法である．ピアノであれば弦とハンマーは固定されているため倍音構造は変化しないが，ギターは弾弦位置を自由に変えて演奏できる．

3. 音高系列と標準偏差

MIDI の規格では 128 音の楽音の基本周波数が定められている．たかだか 128 音であるため，周波数領域上で楽音を扱うよりも計算効率が良い．

音高周波数は基準音から定められている．平均律の音高周波数を式 1 に示す．

$$p_n = 440 \cdot 2^{\frac{n-69}{12}} \quad (0 \leq n \leq 127) \tag{1}$$

p_n は音高周波数である．n は MIDI ノート番号である．

楽器音の音響信号を $x = \{x_0, x_1, ..., x_{w-1}\}$ としたとき，FFT を式 2 に示す．

$$x_i = \frac{1}{w} \sum_{k=0}^{w-1} X_k \cdot e^{i2\pi \frac{i}{w}} \tag{2}$$

X は振幅スペクトルの集合，w は分析フレーム数である．式 2 は正の領域のみ考慮している．音高系列の集合を式 3 に示す．

$$s := \{s_n = X_{\lfloor 2wp_n/f \rfloor}\} \tag{3}$$

式 4 で s から標準偏差が求められる．

$$\sigma^2 = \frac{1}{n} \sum_{i=0}^{127} (s_i - \mu)^2 \tag{4}$$

$$\mu = \frac{1}{n} \sum_{i=0}^{127} s_i \tag{5}$$

σ は標準偏差，μ は算術平均である．

本論文では，音高系列と標準偏差の比を振幅比としている．振幅比を式 6 に示す．

$$AR := \{AR_n = s_n/\sigma\} \tag{6}$$

振幅比の時間変化を表すため，ある時刻 t の音高系列を $s_n(t)$，標準偏差を $\sigma(t)$ と表す．ある時刻の振幅比 $AR_n(t)$ を式 7，振幅比の時刻の微分 $AR'_n(t)$ を式 8 に示す．

$$AR_n(t) = s_n(t)/\sigma(t) \tag{7}$$

$$\frac{dAR_n(t)}{dt} = \lim_{t \to 0} \frac{AR_n(t+\Delta t) - AR_n(t)}{\Delta t} \tag{8}$$

$$= AR'_n(t) \tag{9}$$

4. ピーク分類の評価実験

本章では，エレキギター 5 本を使用し，標準偏差による音高系列のピーク分類が有効であるか評価実験を行った．使用したエレキギターの諸元を**表 1** に示す．

表中の音数とは，弦の数とフレット数を掛けた数である．

実験では共通して，サンプリング周波数は 44,100Hz，FFT の分析フレーム数は 16,384 点，窓関数はハニング窓を用いている．FFT の窓幅と分析フレーム数は同じである．FFT は cuFFT[15] を用いた．以上の条件で，NVIDIA GeForce GTX 970 を用いて FFT を実行した結果，その処理時間は 1 マイクロ秒未満であった．

表 1 評価実験で利用したエレキギター
Table 1 Guitars used in our evaluation experiment.

ギター名	PU 数	動作方式	フレット数	音数
F 社 S	3	シングル	22	138
E 社 S	2	ハムバック	22	138
F 社 T	2	シングル	21	132
G 社 1	2	ハムバック	22	138
G 社 3	2	ハムバック	20	126

4.1 確率密度による音高系列の有効性

音高系列は音楽に必要な音高の周波数のみ利用している．従って，本来存在するはずの周波数スペクトルの成分は失われることになる．本章では，単一楽音の音高系列から確率密度を求め，基本周波数や倍音のピークが標準偏差によって分類可能か述べる．

確率変数を音高系列のパワースペクトルとしたとき，確率密度は式 10 で求められる．

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{10}$$

図 1 は，F 社 S の 4 弦 2 フレット（220Hz）のパワースペクトルと音高系列から確率密度を求めた図である．

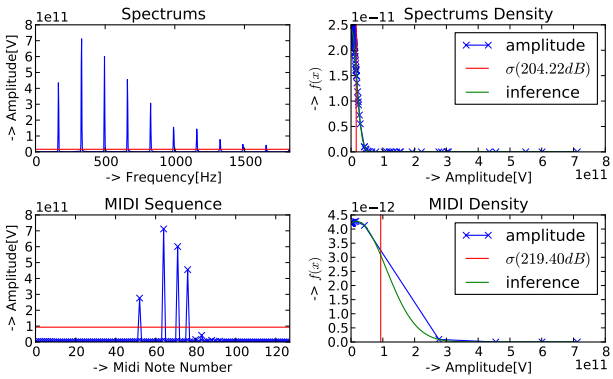

図 1 パワースペクトルと確率密度の標準偏差の関係

図 1 上段はパワースペクトルである．図 1 下段は音高系列である．音高系列の横軸は，式 1 の MIDI ノート番号に対応している．図 1 赤線が各々の標準偏差である．なお，図 1 の振幅は相対電圧である．図 1 の確率密度では，共通して正規分布となっており，ほとんどが非ピークに分類されていることがわかる．図 1 上段では，標準偏差より高いパワースペクトルが多く，確実にピークを分類できていない．ピークの裾野ではなく，ピークの頂点が分類されることが望ましい．一方で，音高系列ではわずか 4 つのピークだけが分類されている．すべて基本周波数と倍音周波数のピークであり，それらの頂点が確実に分類されている．

図 1 より，パワースペクトルより音高系列のほうが正確にピークが分類されている．しかし，基本周波数を除いた一部の倍音のピークが分類されていない．その理由とし

て，倍音は必ずしも音高周波数に対応しないことが挙げられる．倍音は基音のオクターブ上であれば音高系列に出現する．非オクターブの倍音周波数は音高周波数と対応しない．

それでもなお，図 1 では，出現しないオクターブ上の倍音が存在する．これはインハーモニシティ [16] と呼ばれる弦の剛性によって倍音周波数がわずかに変化する現象が原因である．インハーモニシティが発生する限り，オクターブ上の倍音周波数は音高系列を成さないことがある．倍音周波数のパワースペクトルが基本周波数より強いパワースペクトルを示したとしても，音高周波数から外れていれば音高系列に現れない．

各ギターの各弦各フレット（計 642 サンプル）を弾弦し，各々の音高系列の標準偏差を閾値として基本周波数のピークを分類した．その結果，100%の確率で基本周波数が含まれていた．標準偏差によるピーク分類は，基本周波数のピークに対して確実に分類できることがわかった．

4.2 標準偏差の時間変化

本節では，標準偏差の時間変化と基本周波数のピーク分類の可否について，無音区間と演奏区間のある音響信号から変動係数を求め，音高系列と標準偏差の比（以下，振幅比）を求めた．

変動係数は，平均に対して標本値が変動する具合を示す指標であり，$CV(t) = \sigma(t)/\mu(t)$ で表す．$\sigma(t) \to 0$ と置くと，変動係数は限りなく 0 に近づくため，平均から標本値が動いていないことを示す．

図 2 は，F 社 T で 4 弦 3 フレットを 1 度だけ弾弦し，10 ミリ秒ごとに変動係数と振幅比を求めた図である．**図 3** は，図 2 と同様の条件で振幅比の時間変化を表した図である．

演奏区間は約 2.4〜14.2 秒である．約 14.2 秒でミュートした．図 2 上段は音響信号，図 2 中段はパワースペクトルと標準偏差，図 2 下段は変動係数と振幅比である．

変動係数は発音した瞬間に急激に値が上昇しているが，約 8 秒以降から徐々に減衰している．一方で，振幅比はアタックする瞬間において谷間が発生している．これは無音区間において，ある周波数のノイズと標準偏差の比が 3〜5 倍程度になることを示している．一方で，変動係数よりも振幅比のほうがミュートするまで値が伸びている．

図 3 の MIDI ノート番号 30 番あたりの無音区間において振幅比が 7 倍程度になっている．このピークは電源ノイズである．

図 4 は振幅比を微分した時間変化である．

2.4 秒付近に弾弦時のピークが発生している．発音中の 2.8〜13.5 秒付近は勾配に変化がなく，0.0 付近の値を示している．それ以外の発音していない区間については，雑音のようなノイズが発生している．

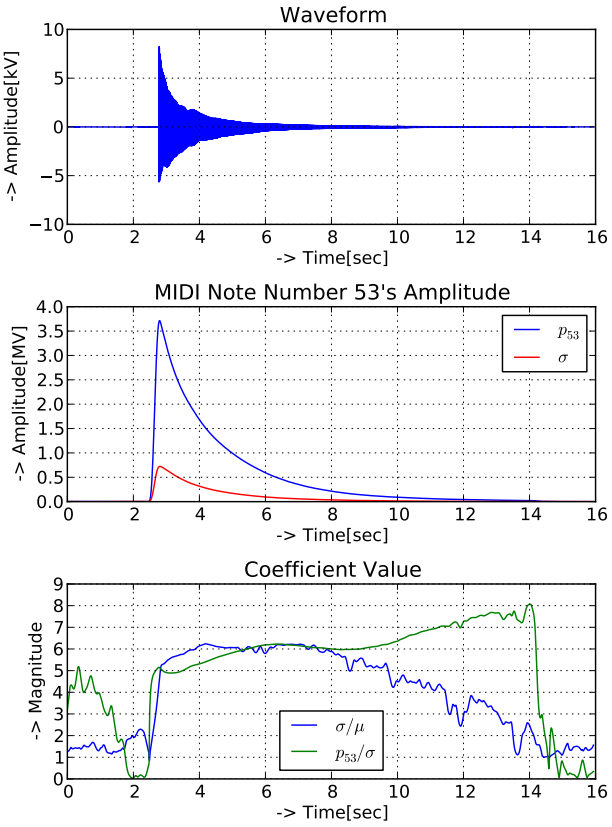

図 2 変動係数と振幅比の時間変化

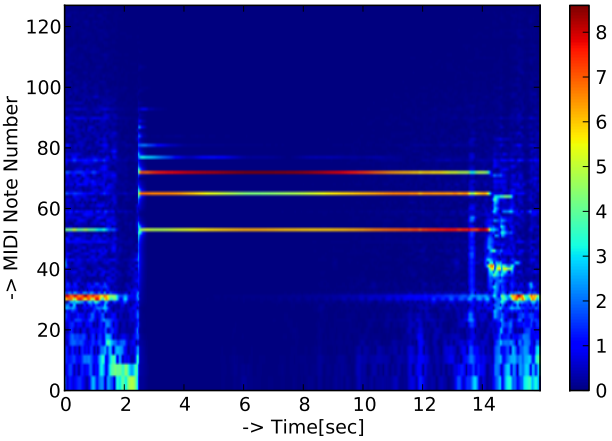

図 3 振幅比のソナグラム

一般的に，弦楽器の楽器音は強制振動をさせない限り，時間変化とともにパワースペクトルは減衰する．一方で，図 3 の振幅比は時間変化で減衰せず，演奏区間中の応答性能が高い．振幅比の勾配も，弾弦された瞬間において，基本周波数と倍音のピークが発生している．また，弾弦した瞬間だけ，MIDI ノート番号 35 番以下で負の値を示している．

Wavelet 変換は，時間変化の応答性能は高いが，減衰に対する応答性能が低い．弾弦時刻と演奏区間は Wavelet 変換でも求められるが，詳細な周波数領域の情報が得られず，フーリエ変換の結果を使い回すことができない．計算リソースが限られた環境では，Wavelet 変換とフーリエ変

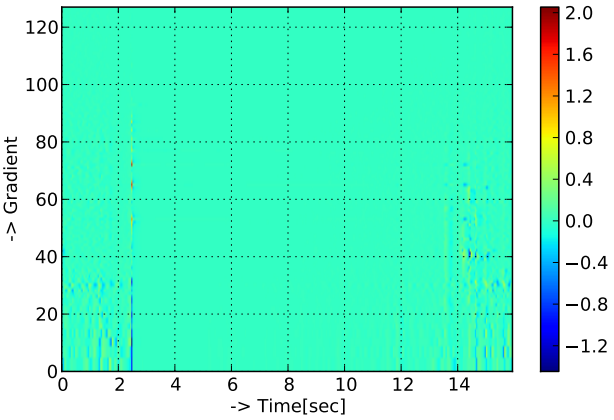

図 4 振幅比の微分のソナグラム

換の併用は冗長である．本論文で述べた手法は，フーリエ変換で得られた計算リソースを使い回せる点において有用である．

5. おわりに

本論文では，エレキギターの単一楽音の音高系列から標準偏差を閾値とし，基本周波数や倍音のピークや，振幅比とその微分から弾弦時刻や演奏区間を求めた．音高系列の標準偏差は，基本周波数のピークを分類したり，弾弦時刻や演奏区間の抽出に有用であることがわかった．

本論文の対象は単一楽音であった．和音は個々の楽音が重なりあった音響信号である．各々の楽音を事象として捉えたとき，それぞれ独立していると考えられる．楽音はヘルムホルツ共鳴 [17] やカップリング [18]，インハーモニシティ [19] など様々な影響を受けるが，その影響は音高系列に現れにくい．今後は音高系列から和音の基本周波数を分類する手法について研究を行う．

今後の展望は，演奏をリアルタイムに認識するシステムが構築できることである．従来の手法では 1 秒以内をリアルタイムと定義する研究が多く [8][20][21]，それ以上に短い処理時間を求められるシステムでは利用できない問題があった．今後，極限まで処理時間が短い方法で和音の認識ができるようになれば，楽器や音楽に関連するシステムを構築する幅が広がることが期待できる．

参考文献

[1] Grubbs, F. E.: Procedures for detecting outlying observations in samples, *Technometrics*, Vol. 11, No. 1, pp. 1–21 (1969).
[2] 浜中雅俊，後藤真孝，麻生英樹，大津展之：学習に基づくクォンタイズ: 発音時刻の楽譜上の位置の推定，情報処理学会 音楽情報科学研究会 研究報告，pp. 21–28 (2001).
[3] 後藤真孝：音楽音響信号を対象としたメロディーとベースの音高推定，電子情報通信学会論文誌 D， Vol. 84, No. 1, pp. 12–22 (2001).
[4] Ichise, M., Emura, N., Miura, M. and Yanagida, M.: Constructing an integrated system for the practice of playing the guitar., *The Journal of the Acoustical Society of America*, Vol. 124, No. 4, pp. 2490–2490 (2008).
[5] Soft, U.: ロックスミス，UBI Soft（オンライン），入手先 ⟨http://www.ubisoft.co.jp/rocksmith/⟩（参照 2016-08-09）.
[6] Noll, A. M.: Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum, and a maximum likelihood estimate, Vol. 779 (1969).
[7] Saito, S., Kameoka, H., Takahashi, K., Nishimoto, T. and Sagayama, S.: Specmurt analysis of polyphonic music signals, *Audio, Speech, and Language Processing, IEEE Transactions on*, Vol. 16, No. 3, pp. 639–650 (2008).
[8] MacLeod, P. and Wyvill, G.: A Smarter Way to Find Pitch, *International Computer Music Conference, ICMC*, Vol. 5, pp. 138–141 (2005).
[9] 後藤真孝，吉井和佳，藤原弘将，Mauch, M.，中野倫靖：Songle: 音楽音響信号理解技術とユーザによる誤り訂正に基づく能動的音楽鑑賞サービス，情報処理学会論文誌，Vol. 54, No. 4, pp. 1363–1372 (2013).
[10] Goto, M., Nakano, T., Kajita, S., Matsusaka, Y., Nakaoka, S. I. and Yokoi, K.: VocaListener and VocaWatcher: Imitating a human singer by using signal processing, *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, IEEE, pp. 5393–5396 (2012).
[11] 亀岡弘和：非負値行列因子分解の音響信号処理への応用，日本音響学会誌， Vol. 68, No. 11, pp. 559–565 (2012).
[12] 竹渕瑛一，井上寛生，徳弘一路，速水治夫：音高情報のオクターブ圧縮によるピッチ推定アルゴリズムの提案，マルチメディア，分散，協調とモバイル (DICOMO2015) シンポジウム，Vol. 33, No. 12, pp. 1103–1109 (2015).
[13] 竹渕瑛一，井上寛生，徳弘一路，速水治夫：音高情報のオクターブ圧縮によるピッチ推定アルゴリズムの提案，マルチメディア，分散，協調とモバイル (DICOMO2015) シンポジウム，Vol. 33, No. 12, pp. 1103–1109 (2015).
[14] Conklin Jr, H. A.: Piano strings and “phantom” partials, *The Journal of the Acoustical Society of America*, Vol. 102, No. 1, pp. 659–659 (1997).
[15] Nukada, A., Ogata, Y., Endo, T. and Matsuoka, S.: Bandwidth intensive 3-D FFT kernel for GPUs using CUDA, *High Performance Computing, Networking, Storage and Analysis, 2008. International Conference on*, IEEE, pp. 1–11 (2008).
[16] Fletcher, H.: Normal vibration frequencies of a stiff piano string, *The Journal of the Acoustical Society of America*, Vol. 36, No. 1, pp. 203–209 (1964).
[17] Firth, I. M.: Physics of the guitar at the Helmholtz and first top-plate resonances, *The Journal of the Acoustical Society of America*, Vol. 61, No. 2, pp. 588–593 (1977).
[18] Arthur, P., Le Carrou, J.-L., Navarret, B., Dubois, D. and Fabre, B.: A vibro-acoustical and perceptive study of the neck-to-body junction of a solid-body electric guitar, *Acoustics 2012* (2012).
[19] Fletcher, H., Blackham, E. D. and Stratton, R.: Quality of piano tones, *The Journal of the Acoustical Society of America*, Vol. 34, No. 6, pp. 749–761 (1962).
[20] Camacho, A.: SWIPE: A sawtooth waveform inspired pitch estimator for speech and music, PhD Thesis, University of Florida (2007).
[21] Tobise, H., Takegawa, Y., Terada, T. and Tsukamoto, M.: Construction of a system for recognizing touch of strings for guitar, *New Interfaces for Musical Expression, 2013. International Conference on*, NIME, pp. 261–266 (2013).