

ベース音を用いた高速な和音認識手法の提案

竹渕 瑛一¹ 梶並 知記² 徳弘 一路³ 速水 治夫³

概要: 本論文では、ベース音を用いることで、Chroma-Vector 法で得られる音階情報を音程情報に変換することで、和音認識のパターンマッチを簡略化するための手法を提案する。本研究では、和音と倍音の調波構造に着目することで、音程情報をビットフラグと見立て、演奏情報と和音とのパターンマッチングを行った。評価実験では、著者の一人が試作システムとエレキギターを用いてコードを繰り返し弾弦することで実施した。評価実験の結果、完全一致では F 値が 19.1%、部分一致では F 値が 30.8%となったが、実験中の目視によってどのコードが弾弦されたのか理解できる結果となった。従って、提案手法と機械学習の組み合わせで認識率を向上させることが可能と考えられる。今後の課題として、演奏者がうまく弾けない状況やエレキギターの調波構造を考慮することである。先行研究のように音響信号から和音を認識するアプローチから、押弦位置を認識するアプローチで課題を解決できると考えられる。

Proposed Fastest Chord Recognition Method Using Bass Pitch

EIICHI TAKEBUCHI¹ TOMOKI KAJINAMI² ICHIRO TOKUHIRO³ HARUO HAYAMI³

1. はじめに

バンドによる複数人の演奏では、一つの楽曲を同期的にリアルタイムで演奏するために、極めて高度な協調作業が要求される。

例えば、バンド音楽であれば、あらかじめ決められた音楽を正確に演奏するだけのように見えるが、実際には何度も繰り返し練習しなければ1曲を弾き通すことは極めて難しい。暗譜せずに楽譜通り演奏するには、長い経験年数と練習時間が必要とされている。ジャズの即興演奏も同様に、周囲の演奏を聴きながら次の演奏を考える必要があり、違和感なく演奏するだけでも極めて難易度が高い。

近年では、ドラムマニアやDTXManiaなど、ゲームと音楽を通じてドラムを練習するシステムが開発されている。エレキギターではRocksmithが開発されており、既存の楽曲から演奏を評価するシステムが既に確立している。

一方で、自由な演奏を認識するための技術はまだ確立されていない現状がある。Rocksmithは既存の楽曲通りに弾いているか判定しているため、現在どのような音を弾いた

のかまでは理解することができない。また、ギターソロも楽曲通りに弾かなければならず、自由な演奏までは程遠い現状がある。

また、和音認識では様々な先行研究が存在しているが、既存の楽曲を静的に解析することを目的としている研究が中心である。リアルタイムで演奏されている内容を評価するシステムでも、MIDI楽器の条件下で研究されており、多くのエレキギター演奏者が理想とするシステムに近づいていない現状がある。

そこで、自由に演奏された内容を認識し、理解・評価するためのシステムを開発するために必要となるのが、エレキギターの音響信号から演奏内容を認識するための技術である。この技術が実現することにより、MIDIギターなどの専用のハードウェアを必要とせず、手元のエレキギターとコンピュータを繋ぐことで、自分の演奏を解釈・評価し、練習できるようなシステムを実現することができるようになる。

以上のことを踏まえ本研究の目的は、Chroma-Vector法を用いることによる和音推定の精度を向上させ、リアルタイム・システム上で動作させることである。

本論文では、リアルタイムで演奏されている和音を推定し、遅延なく提示するための手法について述べる。提案手

¹ 神奈川工科大学大学院博士後期課程

² 岡山理科大学総合情報学部情報科学科

³ 神奈川工科大学情報学部情報メディア学科

法は、Chroma-Vector 法による音階情報から、ベース音をもとに音程情報に正規化する。和音の構成音を 12 ビットの値とみなし、音程情報との排他的論理和を取ることで、演奏された和音とのパターンマッチを行う。提案手法は、論理演算によってパターンマッチを行うため、計算負荷が少なく、リアルタイム処理に向いている点の特徴である。

本論文では、第 2 章で先行研究と提案手法の着眼点について述べる。第 3 章で提案手法のアルゴリズム、第 4 章で提案手法の評価と考察について述べる。

2. 先行研究と提案手法の着眼点

本章では、提案手法の先行研究となる、Chroma-Vector 法を用いた手法と提案手法の差分について述べる。また、提案手法の着眼点についても述べる。

2.1 先行研究

和音を認識するための手法として、Chroma-Vector 法が代表的な手法となっている。Chroma-Vector 法とは、音響信号の周波数スペクトルを音階領域に畳み込むことにより、音階ごとの周波数成分を分析する手法である。

文献 [12] では、Chroma-Vector 法を応用し、EM アルゴリズムから HMM のパラメータを学習することで和音を推定する方法により、約 75% の確率で和音を推定できている。文献 [6] では、和音の遷移確率を学習することで、次に遷移する和音のコード進行をモデル化している。

これらの手法では、あらかじめ HMM のパラメータを学習してモデル化する必要がある。静的な環境で有効であるが、ジャズの即興演奏のように予測の難しい環境については、複雑な和音が演奏されることもあるため、認識率が低下する可能性が十分に予想される。また、これらの研究では、ピアノによって実験が行われている。本研究が対象としているのは、エレキギターのように 1 つの和音で複数のコードが存在する楽器であり、ジャズのような即興演奏に対応することである。

文献 [17] では、個々の和音には周波数スペクトルに 7~8 個の谷間が存在することを利用し [16][18]、Chroma-Vector 法を周波数スペクトルにおける山とみなして両者を併用することで、候補の和音の中から約 2.978 位以内に目的の和音が存在していることを報告している。

文献 [17] は、周波数スペクトルに存在している谷間の周波数とのテンプレートマッチングを行っているため、機械学習などを用いるより高速に動作することが予想されている。しかし、実験では MIDI 音源を利用しているため、現実の楽器で必ずしも正しく動作するとは限らない。特に、楽器にはカップリング [1][13] と呼ばれる共振現象や、インハーモニシティ [4][15] と呼ばれる倍音の周波数のずれによる音のゆらぎが発生しやすい。エレキギターは様々な種類

が存在するため、全てに対応できるかは疑問である。

文献 [3] では、深層学習を用いて周波数スペクトルから演奏された和音の推定を行っている。Recurrent Neural Network によって実装した場合は、平均 93.6% の確率で認識に成功している。他にも非負値行列因子分解 [8][14] を用いた手法なども挙げられる。

これらの手法は、Chroma-Vector 法を一部利用している手法も存在するが [8]、計算負荷が高い手法を組み合わせているため、リアルタイムに演奏を解釈することが難しく、本論文で述べるようなリアルタイムに演奏された和音を提示する用途には向いていない。

2.2 提案手法の着眼点

和音において重要な要素にベース音とルート音がある。ベース音とは和音の中で最も低い楽音のことであり、ルート音とは和音の中で基準となる楽音のことである。ルート音とベース音は基本的には同じ楽音を示すが、転回形によってはルート音とベース音が異なる時がある。

ルート音とベース音が一般的に同じ楽音を示すと仮定した時、Chroma-Vector 法で得られた音階情報 (Pitch Class [5]) を、ベース音をもとに音程情報 (Interval Class) に正規化することによって、和音認識がより簡単になることに着目した。

また、提案手法は音程情報をビットフラグとして捉えている。音階情報は基音だけでなく倍音も含まれているため、ピークと非ピークが明確であることから、標準偏差によってそれらを分類することが可能である。このピークと非ピークをビットとして捉えれば、論理演算だけで分析することができ、和音認識の処理の高速化も望めるようになる。

3. Chroma - Vector 法の併用による和音認識手法

提案手法は、Chroma-Vector 法とベース音検出法を併用することにより、音階情報を音程情報に正規化することによって、高速に和音を認識する手法である。

提案手法は大きく 5 ステップに分けることができる。

- (1) 振幅スペクトルを MIDI ノート番号に対応させる
 - (2) ベース音を検出する
 - (3) Chroma-Vector 法により音階情報を取得する
 - (4) ベース音から音階情報を音程情報に正規化する
 - (5) 音程情報より和音とのパターンマッチを行う
- なお、本論文ではベース音は既知であるとする。本論文で用いる変数は、表 1 のとおりである。

3.1 振幅スペクトルと MIDI ノート番号の対応

提案手法では、FFT で得られた振幅スペクトルを MIDI ノート番号に対応付けてあることが重要である。Chroma-

表 1 本論文で用いる変数の一覧

Table 1

変数名	説明
x	演奏情報の音響信号 [V]
X	振幅スペクトルの集合 [V]
M	MIDI ノート番号の集合 [V]
S	Chroma-Vector 法で得られる音階情報 [V]
I	S を 2 値化した集合 [bool]
a	音程ごとの調波構造の集合 [bool]
b	ベース音の MIDI ノート番号 [整数]
c	任意のコード [コード名]
w	FFT のサンプル数 [整数]
p	任意の MIDI ノート番号 [整数]
s	サンプリング周波数 [整数]
r	余分に共起するビット数の閾値 [整数]
H	倍音の調波構造の集合 [bool]
ξ	調波構造の分布を表す集合 [bool]
ζ	σ の標準偏差 [V]
σ	I を b で正規化した音程情報 [V]

Vector 法は周波数領域のすべてが対象となっているため、FFT の分析フレーム数によっては計算効率が落ちる場合がある。そこで、FFT の分析フレーム数への依存をなくすため、FFT で得られた振幅スペクトルを MIDI ノート番号へ対応させる。

まず、演奏情報のデータを $x = \{x_1, x_2, \dots, x_w\}$ と表す。フーリエ変換を式 1 で表す。

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

式 1 は FFT の分析フレーム数を w としたとき、振幅スペクトルの集合を $X = \{X_0, X_1, \dots, X_{w-1}\}$ で表す。なお、FFT の周波数スペクトルについては負の領域を考慮しないものとする。

このとき、振幅スペクトルを MIDI ノート番号へ対応させた集合 M を式 2 に表す。 p は MIDI ノート番号である。 s はサンプリング周波数である。なお、MIDI ノート番号は $0 \leq p \leq 127$ の範囲を取る。

$$M := \left\{ M_p = \text{floor} \left(\frac{w \cdot 440 \cdot 2^{\frac{p-69}{12}}}{s} \right) \right\} \quad (2)$$

3.2 音階情報の取得と音程情報への正規化

既にベース音は既知であると仮定し、ベース音となる MIDI ノート番号を b として表す。

音階情報とは、音階ごとに振幅スペクトルの平均を取ったものである。音階情報 S は式 3 で表すことができる。式 3 は、 $0 \leq i \leq 11$ の範囲を取る。式 3 の総和を取る範囲が $0 \leq j \leq 11$ であるのは、音階の数が 12 個存在するためである。

$$S := \left\{ S_i = \frac{1}{11} \sum_{j=0}^{11} M_{11j+i} \right\} \quad (3)$$

ただし、このまま音階ごとに平均を取ると MIDI ノート番号が 132 音存在することになるため、一部の音階において正しい平均を取ることができないため、 $j \Leftrightarrow 10, i > 7$ のときは 10 で平均を取る。

音程情報を σ としたとき、音程情報は式 5 で表す。このとき σ の i は、 $0 \leq i \leq 11$ を取る。式 4 はローテート演算のことであり、任意のベース音の添字 b を i だけシフトしたときの添字を返す関数である。

$$r(b, i) = (b + i) \bmod 12 \quad (4)$$

$$\sigma := \{ \sigma_i = S_{r(b, i)} \} \quad (5)$$

音程情報は音階情報をベース音の音階を 0 に向かってシフトしただけであるため、音程情報に含まれているピークと非ピークのカテゴリができていない。音階情報には倍音も畳み込まれるため、ピークと非ピークには明確な境界が存在する。そこで、ピークと非ピークの 2 クラスを分類するために音程情報 σ の標準偏差 ζ を求める。標準偏差 ζ をもとに σ を 2 値化した集合 I を式 6 に表す。

$$I := \left\{ I_i = \begin{cases} 1 & \zeta < \sigma_i \\ 0 & \text{otherwise} \end{cases} \right\} \quad (6)$$

これにより、音程情報をピークと非ピークとで分類できるため、音程情報からそのまま音認識を行うことができる。また、2 値化することで 12 ビットのフラグとして管理できるようになるため、パターンマッチを行う上で高速に処理できるようになる。

3.3 和音の認識

提案手法における和音認識は、2 値化集合 I より、 $I_n \Leftrightarrow 1$ となる音程が構成音となる。従って、 I_3, I_7, I_{10} が 1 であれば、Cm7 コード (C は任意のルート音) であると予想される。ただし、この例は調波構造として倍音がほとんど含まれていない。Cm7 コードは、基音の 5 倍音に長 3 度が含まれるため、 I_4 も 1 となることがある。

これは見かたを変えると、構成音と調波構造によって共起する倍音は決まっていると考えられる。そのために、和音に含まれている構成音の倍音と調波構造から音程情報を求める。

調波構造は基音、3 倍音、5 倍音が和音に影響を及ぼしやすいことから、調波構造を $H = \{h_0, h_1, h_2\}$ として表す。 h_0 は基音であり、 h_1 は 3 倍音、 h_2 は 5 倍音である。2 値化集合により音程情報を数理論理的に扱えるようになった

ため h も 1 か 0 で表す.

次に, 構成音ごとに調波構造の分布を表すため, 式 7 と式 8 のような集合を考える. 式 8 は $0 \leq i \leq 11$ の範囲を取る. 集合 a は, 調波構造の分布の集合であるため, 集合の集合である. a_0 はベース音, a_4 であれば長 3 度, a_8 であれば増 5 度における音階ごとの調波構造の分布を表す. また, a は集合の集合であるため, $a_{4,7}$ であれば, 長 3 度の調波構造における完全 5 度のフラグを示している.

$$a_0 = \{h_0, 0, 0, 0, h_2, 0, 0, h_1, 0, 0, 0, 0\} \quad (7)$$

$$a_i = \{\exists j \in \mathbb{N}, 0 \leq j \leq 11, a_{0,r(i,j)}\} \quad (8)$$

集合 a は, 構成音ごとの調波構造の分布であり, かつ調波構造は 0 と 1 で表されている. ここで, 式 10 と式 11 のような, 集合の要素ごとの論理和と排他的論理和の規則を与える. 式 9 はビットを反転させるための規則である. また, 集合の中で共起しているフラグの数を表すのに, 式 12 のようなノルムの規則と, 式 13 のように集合をビットフラグとみなした時の値を求める規則を与える.

$$\overline{A_i} = \begin{cases} 1 & (A_i = 0) \\ 0 & (A_i = 1) \end{cases} \quad (9)$$

$$A + B = \{\exists i \in \mathbb{N}, 0 \leq i \leq 11, A_i + A_i \cdot B_i\} \quad (10)$$

$$A \oplus B = \{\exists i \in \mathbb{N}, 0 \leq i \leq 11, A_i \cdot \overline{B_i} + \overline{A_i} \cdot B_i\} \quad (11)$$

$$\|A\| = \frac{1}{N} \sum_{i=0}^{N-1} A_i \quad (12)$$

$$\hat{A} = \sum_{i=0}^{N-1} A_i \cdot 2^i \quad (13)$$

すなわち, メジャーコードであれば, $maj = a_0 + a_4 + a_7$ のように表すことができる. ここでコードの調波構造の分布の集合を ξ_c とする. c は任意のコードを表している. メジャーコードであれば ξ_{maj} , マイナー 7th であれば ξ_{m7} などのように表す. 与えられた 2 値化集合 I より, 任意のコード c において, ξ_c と比較した結果を s_c としたとき, $s_c = I \oplus \xi_c$ で表される.

s_c のノルムが $\|s_c\| \leq r$ (r は余分に共起するビットの数) であるとき, I が ξ_c のコードである可能性が高いと考えられる. このとき, $r = 0$ であれば完全一致となる.

表 2 は, 各コードの構成音の音程と音程情報 ξ_c をビットパターンとみなした時の値 $\hat{\xi}_c$ を表している. また, 表 3 は, $\hat{\xi}_c$ が重複するコードの一覧である. なお, 表 2 と表 3 は, $H = \{1, 1, 0\}$ を想定している.

表 3 が示すように, ビットパターンが一致するコードは極めてわずかである. 表 3 中の代表的なコード (Cm6, C6, C7-9 など) に対し, あまり用いられない裝飾音が含まれているコードが重複しているため, 仮に音程情報 I のビットパターンが重複したコードと一致していたとしても, 表 3

表 2 各コードの構成音の音程とビットパターンの値

Table 2 Intervals of Each chord's construction note and bit patterns.

コード名	R	3	5	7	9	11	13	ビット値
Cmaj	0	4	7					2197
C5	0		7					133
C-5	0	4	6					2259
C6	0	4	7	9				2709
C69	0	4	7	9	2			2709
Csus4	0	5	7					165
C7	0	4	7	10				3253
CM7	0	4	7	11				2261
C7-5	0	4	6	10				3315
C7-9	0	4	7	10	1			3511
C7sus4	0	5	7	10				1189
C7+9	0	4	7	10	3			3261
C7+11	0	4	7	10		6		3319
C7+13	0	4		10			9	3761
Cadd9	0	4	7		2			2709
C9	0	4	7	10	2			3765
C9-5	0	4	6	10	2			3831
C-9	0	4	7	10	1			3511
C-9+5	0	4	8	10	1			3515
CM9	0	4	7	11	2			2773
Caug9	0	4	7	10	2			3765
C11	0	4	7	10	2	5		3765
C13	0	4	7	10	2	5	9	3765
Cm	0	3	7					1165
Cm69	0	3	7	9	2			1693
Cm7	0	3	7	10				1197
CmM7	0	3	7	11				3277
Cm6	0	3	7	9				1693
Cm7-5	0	3	6	10				1259
Cm9	0	3	7	10	2			1709
Cm11	0	3	7	10	2	5		1709
Cm13	0	3	7	10	2	5	9	1725
Cdim	0	3	6					1227
Cdim7	0	3	6	9				1755
Caug	0	4	8					2457
Caug7	0	4	8	10				3513
CaugM7	0	4	8	11				2521
Cadd2	0	4	7		2			2709
Cadd4	0	4	7			5		2229

表 3 ビット値が重複するコードの一覧

Table 3 List of chord that bit value is duplicated.

ビット値	重複するコード
1693	Cm69, Cm6
1709	Cm9, Cm11
2709	C6, C69, Cadd9, Cadd2
3511	C7-9, C-9
3765	C9, Caug9, C11, C13

表 4 評価実験で利用したマシン構成

Table 4 Machine configuration to be used in the experiment.

	概要
OS	Windows 7 Enterprise 64bit
CPU	Intel Core-i7 2600k
メモリ	16 ギガバイト
GPU	NVIDIA GTX970
Audio IF	Presonus Audio Box USB

中の代表的なコードが演奏された可能性が高いと考えられる。なお、 $H = \{1, 0, 0\}$ の場合はさらに重複するコードは少なくなり、重複するビットパターンは 3 種類であり、重複するコードはそれぞれ 2 種類ずつである。どれも 9th 系の装飾音が含まれるコードであるため、重複による誤認識の可能性は限りなく少ないと考えられる。

4. 提案手法の評価

本研究では、提案手法がリアルタイム・システム上で和音を認識できているか確認するため、試作システムを用いて提案手法の評価を実施した。評価実験では、著者の一人（以下、演奏者）に実験対象のコードを 1,000 フレーム間にかけて 120BPM の 4 分音符で繰り返し弾弦することで、正しいコードが認識されているか評価する。本章では、評価の方法と評価結果について述べる。

4.1 実験環境

評価実験で利用したマシン構成を表 4 に示す。

Audio Box USB の入力ボリュームは 10 時、サンプリング周波数は 44,100Hz、バッファサイズは 256 サンプルに設定されている。

提案手法では FFT を実行するために、ライブラリとして cuFFT[9] を用いている。FFT のサンプル数は 16,384 点である。

サンプリング周波数が 44,100Hz、FFT のサンプル数が 16,384 点に設定されている理由として、周波数分解能は $\frac{f_s}{N}$ となる。この設定の周波数分解能は約 1.3Hz である。エレキギターのレギュラーチューニングにおいて、6 弦開放の MIDI ノート番号は 28 である。MIDI ノート番号 28 と 29 の差は $43.7 - 41.2 = 2.5$ であるため、実験環境下の周波数分解能は十分であると考えられる。

エレキギターは Fender USA 社製の Custom Shop Stratocaster (F 社 S) を用いた。F 社 S は、シングルコイル・ピックアップを 3 基搭載し、フレット数は 22 フレットである。

4.2 試作システム

本研究では、提案手法を評価するために試作システムを実装した。試作システムの GUI を図 1 に示す。試作シス

テムは 1 秒間に 60 回更新される。



図 1 試作システムの GUI

Fig. 1 The GUI of the prototype system.

図 1 の認識されたベース音は、 $b \bmod 12$ で示す値と対応している。認識された構成音は、集合 I の値と対応している。測定結果の各変数については 4.3 節で述べる。

試作システムでは、認識したいベース音と構成音を指定し、スペースキーで録音を開始する。録音中は和音の認識と評価を行っている。認識されたベース音と構成音については、マゼンタ色で提示される。録音中は録音を示すマークが提示され、1,000 フレーム間の録音が完了するとマークが消えるようになっている。録音完了後に試作システム下部の表に結果が提示される。

試作システムがリアルタイムに実行可能であるか確認するため、216,000 フレーム間だけ試作システムを放置したところ、1 フレームの処理落ちも発生しなかった。また、ベース音検出や FFT を含めた提案手法の処理速度を調べたところ、平均 1 マイクロ秒未満であることがわかった。

4.3 評価方法

評価実験では、試作システムを起動し、任意のベース音の音階とコードのビットフラグを指定して計測する。計測開始から 1,000 フレームの間だけ計測が行われ、指定したコードと現在弾弦しているコードの認識結果が正しいか評価する。計測中は指定されたコードを 120BPM の 4 分音符で、ピックを用いてダウンピッキングを繰り返す。評価には文献 [7] で示されているコードを用いた。弾弦したコードは、C ~ B までのメジャー、マイナー、メジャー M7、メジャー 7、マイナー M7、マイナー 7 コードの計 42 種類のコードである。

評価実験の演奏者は、エレキギター歴が約 10 年程度である。演奏者はミスピッキングやミュートができないなど、初心者でないことを条件に実験の協力を依頼した。

実験では、調波構造が $H = \{1, 0, 0\}$ と $H = \{1, 1, 0\}$ の

場合について同時に評価する。また、余分に共起するビット数が $r = 0$ と $r = 1$ の場合についても評価した。それぞれ、再現率と適合率を求め、最終的に再現率と適合率の調和平均である F 値を求めた。

適合率を式 15、再現率を式 16、F 値を式 17 に示す。 R は正解数、 N は実験区間 (1,000 フレーム)、 C は提案手法が認識したコードの数を表している。

$$precision = \frac{R}{N} \quad (14)$$

$$recall = \frac{R}{C} \quad (15)$$

$$F - measure = \frac{2 \cdot precision \cdot recall}{precision + recall} \quad (16)$$

4.4 評価結果

評価実験の結果を表 5、表 6、表 7 に示す。それぞれの表では、メジャー、マイナー、メジャー M7、メジャー 7、マイナー M7、マイナー 7 の平均と分散を求めている。最下段で余分に共起するビット数 r と調波構造 H の平均と分散を求めている。なお、括弧内の値は分散である。

表 5 評価実験における和音認識の適合率 (%)

Table 5 Precision of chord in the evaluation experiment.

	$r = 0$	$r = 0$	$r = 1$	$r = 1$
	$H = 100$	$H = 110$	$H = 100$	$H = 110$
maj	26.7 (2.9)	0.0 (0.0)	50.2 (38.4)	50.2 (38.4)
min	19.5 (19.3)	0.0 (0.0)	65.0 (14.5)	65.0 (14.5)
majM7	24.0 (25.8)	0.0 (0.0)	0.0 (0.0)	45.8 (30.2)
maj7	8.0 (5.7)	0.0 (0.0)	0.0 (0.0)	12.4 (20.5)
mM7	9.1 (27.4)	0.0 (0.0)	0.0 (0.0)	23.9 (32.9)
m7	14.9 (12.6)	0.0 (0.0)	48.2 (31.6)	48.2 (31.6)
平均	17.0 (7.0)	0.0 (0.0)	27.2 (27.7)	40.9 (17.5)

表 6 評価実験における和音認識の再現率 (%)

Table 6 Recall of chord in the evaluation experiment.

	$r = 0$	$r = 0$	$r = 1$	$r = 1$
	$H = 100$	$H = 110$	$H = 100$	$H = 110$
maj	27.6 (2.9)	0.0 (0.0)	12.6 (10.8)	29.4 (23.3)
min	44.1 (17.9)	0.0 (0.0)	19.5 (6.7)	62.6 (25.4)
majM7	41.3 (24.2)	0.0 (0.0)	0.0 (0.0)	19.3 (12.8)
maj7	16.0 (21.8)	0.0 (0.0)	0.0 (0.0)	12.8 (12.2)
mM7	16.9 (41.3)	0.0 (0.0)	0.0 (0.0)	36.8 (42.3)
m7	29.2 (39.2)	0.0 (0.0)	16.3 (11.1)	28.2 (19.6)
平均	29.2 (10.8)	0.0 (0.0)	8.1 (8.3)	31.5 (15.8)

図 2 は、評価実験の適合率の分布である。図 3 は、評価実験の再現率の分布である。 r と H の各々の値において実験した結果を、昇順でまとめている。

図 2 では、 $r = 0, H = \{1, 1, 0\}$ の場合は、全体のおよそ 50% から適合率 50% を超えている。 $r = 0, H = \{1, 0, 0\}$ は

表 7 評価実験における和音認識の F 値 (%)

Table 7 F-Measure of chord in the evaluation experiment.

	$r = 0$	$r = 0$	$r = 1$	$r = 1$
	$H = 100$	$H = 110$	$H = 100$	$H = 110$
maj	27.1 (2.9)	0.0 (0.0)	20.0 (17.0)	31.3 (23.4)
min	22.0 (20.4)	0.0 (0.0)	29.8 (9.2)	60.8 (14.9)
majM7	25.7 (25.7)	0.0 (0.0)	0.0 (0.0)	23.9 (11.8)
maj7	9.9 (8.3)	0.0 (0.0)	0.0 (0.0)	10.8 (14.2)
mM7	11.1 (33.5)	0.0 (0.0)	0.0 (0.0)	24.4 (32.7)
m7	19.0 (18.6)	0.0 (0.0)	24.4 (16.4)	33.8 (20.7)
平均	19.1 (6.6)	0.0 (0.0)	12.4 (12.7)	30.8 (15.3)

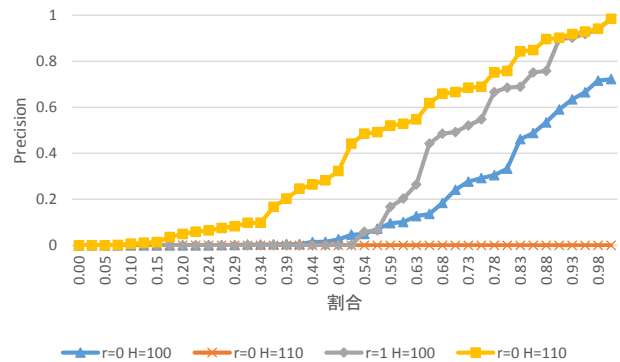


図 2 評価実験の適合率の分布

Fig. 2 Distribution of precision of evaluation experiment.

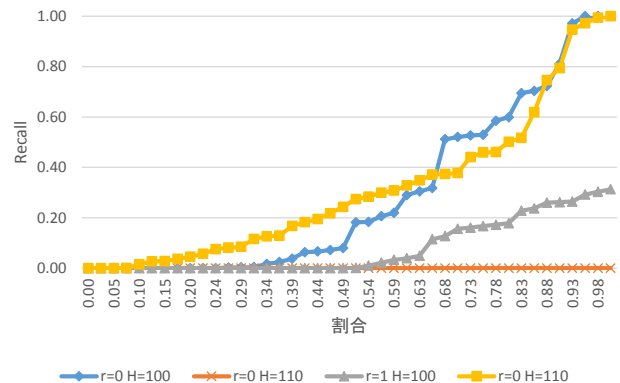


図 3 評価実験の再現率の分布

Fig. 3 Distribution of recall of evaluation experiment.

全体の 90% から適合率 50% となった。 r の値を大きくすれば、適合率は簡単に上昇すると考えられるが、再現率は大きく低下することが予想される。

図 3 では、部分一致に対する候補となった数の多さから再現率を求めているため、 $r = 1$ については再現率と比べると良い結果となった。ただし、それでも再現率が 50% を超えるのは、 $r = 1, H = \{1, 0, 0\}$ では全体の 70% からとなった。

4.5 考察

評価実験の結果では、 $r = 0, H = \{1, 1, 0\}$ の場合では F 値が 0% となった。3 倍音も共起し、完全一致するコードはなかった。また、最も良い結果となったのが

$r = 1, H = \{1, 1, 0\}$ の場合である。3倍音も共起し、1つまで部分一致する場合はF値が30.8%となった。全体的に分散が大きく、セーハコードであるか、1弦や2弦に3rdや7thに含まれているかなどで、結果にばらつきが出た。

セーハコードは、指の腹で複数弦を押弦するコードであるが、3弦や4弦は押さえにくいいため、演奏者も気づかないうちに弾いていなかったケースが多く発生した。Fメジャーコードであれば、オクターブと3rdであり、このとき3弦がうまく弾けなかった場合、長3度を弾く弦が他に存在しないため、認識に失敗することが十分に予想できる。プロであればこのような失敗はないと考えられるが、演奏者がアマチュアの場合は、弾けないコードがある。従って、和音認識を一般化するにあたり、弾いていない音を予測することも必要である。

また、Dメジャーコードのように1弦や2弦に調性感を決めるような構成音が含まれている場合、認識に失敗する確率が上がっていた。エレキギターの個体差にもよるが、1弦や2弦は3倍音の振幅が小さく、単音で弾いても完全5度が認識されなかった。1弦や2弦の周波数スペクトルを計測したところ、5弦や6弦と比較すると、3倍音の振幅スペクトルが小さくなっていた。1弦や2弦の3倍音の振幅スペクトルが小さくなったことから、 $r = 0, H = \{1, 1, 0\}$ の条件下ではF値が0%になったと考えられる。従って、弦ごとの調波構造はアルゴリズム中で考慮する必要があると考えられる。

実験中は、演奏者が試作システムを目視したうえでは、和音が認識されたと感じていた。簡易的なブラインドテストを行ったが、試作システムで提示された和音と演奏された和音は、演奏者の間で一致していた。実際には、1弦や2弦の調波構造を考慮すると、 $r = 0, H = \{1, 0, 0\}$ でもなく、 $r = 0, H = \{1, 1, 0\}$ でもない、中途半端な状態の和音が提示されていたことになる。しかし、人間が目視で和音を認識できることから、機械学習によって解決することも可能であると考えられる。

押弦位置によって認識率が変化していることから、エレキギターに限り、コードの押弦に着目することによって、和音の認識精度を向上させることが可能ではないかと考えられる。エレキギターにおける押弦するコードの組み合わせは膨大であるが、それだけ先に述べたような3倍音の振幅が小さい弦に調性感のある構成音が当たる可能性が高い。各弦の調波構造に着目しつつ、なおかつ弾いていない音を予測するには、提案手法を押弦した時のコードに置き換えながら和音認識をする必要があるのではないかと考えられる。

押弦位置の認識を試みた研究 [2][10][11] はいくつか取り組みがなされているが、その多くはHMMや機械学習などを利用しているため、既に演奏内容が認識されていることに限定している研究が少なくない。押弦位置の認識も含

めて和音を提示できれば、それらを同時に解決できると考えられる。

5. おわりに

本論文では、ベース音を利用することで、Chroma-Vector法の音階情報を音程情報に変換し、和音の調波構造とのパターンマッチングを行うことで、リアルタイムに動作するシステム上で和音を認識する手法について述べた。

評価実験では、提案手法には改良が必要であることを示す結果となった。しかし、試作システム上において目視する限りでは、正しく認識されているように見えた。これは提案手法の定義が厳密すぎるため、エレキギターのような不確定性の大きい実験条件下では、認識率が低下するものと考えられる。

今後の課題としては、和音認識ではなく、押弦認識からアプローチすることで、より確実な和音認識手法を考案する。考察でも述べたとおり、和音認識を一般化するにあたり、演奏者がうまく弾弦や押弦のできないケースや、弦によって調波構造が異なる問題を解決する必要がある。そのためには、従来通りの和音認識からのアプローチでは、解決は困難であると考えられる。

今後の展望としては、押弦位置の認識に加えて和音の認識も行うことである。先行研究 [2][10][11] では、HMMや機械学習などを用いて運指や押弦位置の認識に取り組みされているが、和音の認識と押弦位置の認識を同時に取り組んでいる研究がない。まずは押弦位置の認識に取り組み、押弦位置を提示できてから、和音の認識に取り組みたい。

参考文献

- [1] 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).
- [2] Barbancho, A. M., Klapuri, A., Tardón, L. J. and Barbancho, I.: Automatic transcription of guitar chords and fingering from audio, *Audio, Speech, and Language Processing, IEEE Transactions on*, Vol. 20, No. 3, pp. 915–921 (2012).
- [3] Boulanger-Lewandowski, N., Bengio, Y. and Vincent, P.: Audio Chord Recognition with Recurrent Neural Networks., *ISMIR*, pp. 335–340 (2013).
- [4] 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).
- [5] Lee, K.: Automatic chord recognition from audio using enhanced pitch class profile, *Proc. of the International Computer Music Conference*, p. 26 (2006).
- [6] Lee, K. and Slaney, M.: A Unified System for Chord Transcription and Key Extraction Using Hidden Markov Models., *ISMIR*, Citeseer, pp. 245–250 (2007).
- [7] mamopage: ギターコード表, mamopage (オンライン), 入手先 (<http://www.geocities.jp/mamopage/chordhyou-guitar-full.html>) (参照 2016-05-06).
- [8] Maruo, S., Yoshii, K., Itoyama, K., Mauch, M. and

- Goto, M.: A feedback framework for improved chord recognition based on NMF-based approximate note transcription, *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, IEEE, pp. 196–200 (2015).
- [9] Nvidia, C.: CUFFT library (2010).
- [10] Radicioni, D. and Lombardo, V.: Guitar fingering for music performance, *strings*, Vol. 40, No. 45, p. 50 (2005).
- [11] Radisavljevic, A. and Driessen, P.: Path difference learning for guitar fingering problem, *Proceedings of the International Computer Music Conference*, Vol. 28, sn (2004).
- [12] Sheh, A. and Ellis, D. P.: Chord segmentation and recognition using EM-trained hidden Markov models, *ISMIR*, pp. 185–191 (2003).
- [13] Weinreich, G.: Coupled piano strings, *The Journal of the Acoustical Society of America*, Vol. 62, No. 6, pp. 1474–1484 (1977).
- [14] 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).
- [15] 高澤嘉光, 徳弘一路：弦振動におけるインハーモニシティについて-Conklin の第 2 系列発生説への疑問, 音楽音響研究会資料, MA00-13, pp. 27–32 (2000).
- [16] 宮田聡, 上野佑馬, 蔵内雄貴, 松原正樹, 斎藤博昭：スペクトルディップテンプレートをを用いた和音推定, 情報科学技術フォーラム講演論文集, Vol. 8, No. 2, pp. 337–340 (2009).
- [17] 黒川奈桜子, 斎藤博昭：スペクトルディップとクロマベクトルの併用による和音推定, 情報科学技術フォーラム講演論文集, Vol. 11, No. 2, pp. 165–170 (2012).
- [18] 蔵内雄貴, 松原正樹, 大野将樹, 斎藤博昭：周波数スペクトルの谷状点に基づく和音推定, 情報科学技術フォーラム講演論文集, Vol. 7, No. 2, pp. 5–8 (2008).