

博士論文概要

論文題目

Researches on Music Feature Analysis
using Chroma Vector and its Applications

クロマベクトルを用いた音楽特徴解析と
その応用に関する研究

申請者

植村	あい子
Aiko	UEMURA

情報理工学専攻 画像情報研究

2015年12月

近年、圧縮技術やネットワークを利用した配信システムといった音楽コンテンツの提供手段の増加に伴い、人々が大量の音楽に触れる機会が増加した。それに付随して、「いつでも」「どこでも」「あらゆる」音楽を扱うことができるように、様々な音楽要素の自動認識や音楽情報検索に対する需要は高まっている。和音は和声をコンパクトに表現したものであり、和声は旋律・リズムと並んで音楽の3要素であることから、和音は音楽を構成する重要な要素である。また、和音進行をもとに楽曲構造を判断したり、調性を推定したりできることから、和音は他の音楽要素と関係が深い手掛かりになる。

和音認識は、時系列パターン認識問題として考えられることが多く、従来研究では、音響信号を入力として和音名系列の出力を得るために学習・認識の手順を踏むものが多数存在する。具体的には、音響信号と和音名系列のペアから識別器を学習したり、あらかじめ和音テンプレートと比較したりして、認識時に新たな音響信号の和音名系列を出力する。識別器には、一般的に **Hidden Markov Model (HMM)** や **Support Vector Machine(SVM)**などが用いられている。音響特徴量としては、クロマベクトルと呼ばれる、12音名の各音に対応する周波数のパワーを複数のオクターブに渡って加算した12次元のベクトルが用いられる。クロマベクトルは、オクターブで演奏されても、構成音が同一であれば同一の和音と認識されるという性質があり、和音を捉えやすいといえる。

本研究では、クロマベクトル解析で和音そのものや楽曲の構成を捉えることができることから、音楽音響信号に対し基礎と応用の2面から和音認識応用の検討を行う。基礎面では、和音認識性能について音楽理論を用いた和音認識性能の改善と音楽圧縮の性能評価を行う。また、応用面では実環境で収録された音楽番組やライブ音源に関して、時間周波数情報を用いた楽曲区間検出やカバーソング検索の検討を行う。

第1章では、本研究の位置付けを示す。

第2章では、和音認識の従来研究とクロマベクトル解析の概要を述べる。

第3章では、音楽理論を用いた和音認識性能の改善を目標として、従来手法 [Ellis 2008]では上位3候補の中に正解が含まれる可能性が高いことと、和音の類似性を表す音楽理論 **Doubly Nested Circle of Fifths (DNCOF)**を用いてラフな推定結果を用いることで、従来手法の改善を試みた。本研究では、音楽理論 **DNCOF**から生成する **DNCOF**ベクトルを提案する。これは、クロマベクトルを **DNCOF**平面上に写像したものであり、「向き」と「大きさ」の情報を持つ。この **DNCOF**ベクトルを用いて、従来手法の後処理として誤認識部分の推定と誤り補正を行う。まず、誤認識の推定には、各フレームにおいて、出力和音と **DNCOF**ベクトルのコサイン距離から閾値を用いて閾値以上であれば誤りと判定する。次に、誤りが連続しているフレームでは、補助的な **DNCOF**ベクトルを算出し、従来手法で得

られた第 3 候補までの和音の中から，DNCOF 順で DNCOF ベクトルに近い候補で補正を行う．一方，突発的な誤りのフレームに対しては，前フレームの出力和音を使用することで補正を行う．本論では，The Beatles (180 曲) の楽曲に対し，誤り部分の検出性能と和音認識性能，DNCOF 平面上における正解との角度差の評価を行った．現状，ラベルに離散化すると大きな改善効果は認められないが，DNCOF ベクトルの角度差は削減可能であることを示す．

第 4 章では，音響信号の周波数変形の一例として圧縮に着目し，圧縮信号について音質評価と学習ベースでの和音認識性能評価を行い，低圧縮やコーデックによる影響の調査結果を示す．圧縮信号では，スペクトルは圧縮の影響を受け圧縮率が上がるにつれて高周波成分を中心に影響を受けるが，クロマベクトルはオクターブ情報をまとめているため，一見圧縮による影響はないように見られる．ここで，The Beatles (180 曲)，Queen (20 曲)，C. King (7 曲) と RWC データベースからポピュラー楽曲 (100 曲) の全 307 曲に対して，MP3, AAC-LC, Ogg Vorbis の 3 コーデックについて AAC-LC のみ 12, 24kbps を含み，32～320kbps の CBR (Constant Bit Rate) モードによる圧縮信号について評価を行った．はじめに，音響信号の劣化度合いは ITU-R BS.1387-1 で定義されている PEAQ (Perceived Evaluation of Audio Quality) と呼ばれるオーディオの客観品質評価手法を用いる．実験結果により，どのコーデックでも 96kbps 以下で徐々に音質は劣化するが，HMM, SVM-HMM を用いた和音認識性能においては，4%以内の差であることを確認し，音質の劣化は和音認識に対してロバストであることを示した．また，307 曲を 3 グループに分割して，ビットレートの異なるデータを学習し評価を行い，認識性能は 1%以内の差に収まることを確認し，和音認識においてビットレートが異なるデータを使用しても，ロバストであることを確認した．以上の結果から，圧縮やクロマベクトル抽出において有意な周波数成分が残される場合は和音認識性能が低下しないことを示す．

第 5 章では，クロマベクトルの時間情報に着目し，音楽番組から楽曲区間を抽出する手法を提案する．本研究は頭出しのような視聴支援や音楽情報検索を目的として音楽番組から楽曲区間を抽出するにあたり，クロマベクトルの時系列情報に着目した．クロマベクトルでは，楽曲部分は時間方向に連続成分が続き，それ以外の区間は周波数方向に成分が続いたり不連続であったりすることから，時間方向に続く成分を抽出・指標化し，閾値処理によって楽曲区間の判別を行う．具体的には，クロマベクトルの各フレームのピーク位置において 8 近傍の値から時間方向と周波数方向の曲率を算出し，時間方向に緩やかな場合に 1，それ以外は 0 となるマスクを生成し，ノイズ除去を行うために画像処理でよく用いられる 3 種類の平滑化 (フィルタメディアンフィルタ，ガウシアンフィルタ，バイラテラルフィルタ) を適用する．実験では，標準データを用いた音響信号と 2 番組の

54.4%が楽曲である 208 分の TV 音楽番組の PCM 音源に対して、3 種類のクロマベクトル (CP: Chroma Pitch-base, CLP: Chroma Log Pitch, CRP: Chroma DCT-Reduced log Pitch) を算出し、適合率、再現率、F 値により評価を行う。本論では、提案手法により比較手法 [Zhu,2003] より正確性・網羅性が高く楽曲区間を検出でき、CLP とバイラテラルフィルタの組み合わせが最も検出性能が向上することを示す。

第 6 章では、ライブにおける演奏音源からの楽曲同定として、従来のカバーソング検索手法 [Ellis 2006, 2007] に、時系列的な関係性を付与するとともに音響シーンの判別手法を加え、演奏音源に強固な手法を提案する。これにより、従来のカバーソング手法に対してライブ音源を適用した場合に発生する 2 つの課題を解決する。ここでは、音響信号に MC などの楽曲シーン以外も含まれること、もう一方は 1 信号に対し 1 楽曲の対応である課題に対して提案を行う。楽曲同定フェーズでは、入力クエリからビート同期を行ったクロマベクトルを抽出し、データベース中の原曲群とのマッチングと後処理(重み付け処理)を経て同定楽曲が決定される。提案法では、40 秒のクエリごとに結果を出力し、時系列的な関係性を付与しているため、楽曲が切り替わる部分の同定結果がスムーズになる。一方、シーン判別フェーズにおいては、シーン判別フェーズでは非楽曲シーンと楽曲シーンで違いが現れる楽曲同定の類似度スコア、Pulseclarity, RMS の特徴の 3 つを用い、閾値による判別を行う。結果として、実際のライブ音源 8 つに対して精度 70% の同定性能を得、同定結果、Pulse clarity, RMS の 3 要素を活かすことにより、非楽曲シーン判別も行うことができた。最終的に、楽曲シーン・非楽曲シーンの判別に加え、楽曲シーン部分について楽曲名を出力するシステムを提案する。

第 7 章では本論の総括と今後の課題について述べる。

以上述べたように、本研究ではクロマベクトル解析に基づいて和音認識性能に関する基礎と応用の 2 面から検討を行った。基礎面では、和音は他の音楽要素認識への適用や周波数変形成分に対する和音認識性能の改善が期待できる。応用面では、実環境下での複数の音響要素から構成される音楽番組やライブ音源について、自動で曲名を教えてくれる視聴アプリケーション・Web サービスの提供や、歌詞同期表示や購入案内表示などインタラクティブなアプリケーション・サービスへの応用例も考えられ、本成果の適用により音楽情報処理分野への新たな展開が期待できる。

早稲田大学 博士（工学） 学位申請 研究業績書

氏名 植村 あい子 印

(2015年12月3日現在)

種 類 別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む）
○論文	Music Part Detection in Music Television Program on Chroma Vector Analysis <u>Aiko Uemura</u> , Kyota Higa, Masumi Ishikawa, Toshiyuki Nomura and Jiro Katto, IEEEJ Transactions on Image Electronics and Visual Computing, Dec.2015.
○国際 会議	Effects of Audio Compression on Chord Recognition 【査読有，採択率 46%】 <u>Aiko Uemura</u> , Kazumasa Ishikura and Jiro Katto, The 20th Anniversary International Conference on MultiMedia Modeling (Lecture Notes in Computer Science Volume 8326, 2014, pp 345-352, Springer), Jan. 2014.
国際 会議	Live Version Identification with Audio Scene Detection 【査読有】 Kazumasa Ishikura, <u>Aiko Uemura</u> and Jiro Katto, The 21th Anniversary International Conference on MultiMedia Modeling (Lecture Notes in Computer Science Volume 8935, 2015, pp. 408-417, Springer), Jan. 2015.
○国際 会議	Music Part Segmentation in Music TV Programs based on Chroma Vector Analysis 【査 読有】 <u>Aiko Uemura</u> , Kyota Higa, Masumi Ishikawa, Toshiyuki Nomura and Jiro Katto, IEEE International Symposium on Multimedia (ISM) 2012, pp. 32 - 35, Dec.2012.
○国際 会議	Chord Recognition using Doubly Nested Circle of Fifths 【査読有，採択率 49%】 <u>Aiko Uemura</u> and Jiro Katto, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2012), pp. 449 - 452, Mar. 2012.
	Music/Speech Discrimination based on Chroma Vector Analysis <u>Aiko Uemura</u> and Jiro Katto, MIREX 2015 Music/speech classification and detection task, Oct. 2015.
講演	楽器種類を拡張した打楽器の音源同定 大石皓太郎, <u>植村あい子</u> , 甲藤二郎 電子情報通信学会総合大会, A-10-17, Mar. 2015.
講演	確率モデルに基づく候補提案型鍵盤楽器システム 【学生奨励賞受賞】 村岡眞伍, <u>植村あい子</u> , 甲藤二郎 情報処理学会全国大会, 4S-08, Mar. 2015.
講演	未知楽器を含む打楽器を対象とした音源同定 大石皓太郎, 甲藤二郎, <u>植村あい子</u> , 石倉和将 電子情報通信学会 総合大会, A-10-13, 2014年3月.
講演	和音認識における学習数の影響と誤り傾向の調査 萩原秀一, 甲藤二郎, <u>植村あい子</u> , 石倉和将 電子情報通信学会 総合大会, A-10-14, 2014年3月.

早稲田大学 博士（工学） 学位申請 研究業績書

種 類 別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む）
講演	シーン判別手法を組み込んだライブバージョン楽曲同定システム 石倉和将, <u>植村あい子</u> , 甲藤二郎 電子情報通信学会 総合大会, A-15-31, 2014年3月.
講演	異なるビットレートの楽曲を用いたクロマベクトルの解析と和音認識性能評価 <u>植村あい子</u> , 石倉和将, 甲藤二郎, 情報処理学会 音楽情報科学研究会 デモセッション, 2013年8月.
講演	自動で曲名を教えてくれる楽曲同定機能付きメディアプレイヤー 石倉和将, <u>植村あい子</u> , 甲藤二郎, 情報処理学会 音楽情報科学研究会 デモセッション, 2013年8月.
講演	圧縮やダウンサンプリングがクロマベクトルと和音認識に与える影響について <u>植村あい子</u> , 石倉和将, 甲藤二郎, 情報処理学会音楽情報研究会, 2013年5月.
講演	異なるビットレートの楽曲を用いたクロマベクトルの解析 <u>植村あい子</u> , 石倉和将, 甲藤二郎, 情報処理学会全国大会, 2D-7, 2013年3月.
講演	ライブにおける楽曲の遷移を考慮したリアルタイム楽曲同定システム 石倉和将, <u>植村あい子</u> , 甲藤二郎, 情報処理学会全国大会, 3R-2, 2013年3月.
講演	実環境におけるライブ演奏楽曲の楽曲同定 石倉和将, <u>植村あい子</u> , 甲藤二郎, 電子情報通信学会 総合大会, A-10-15, 2012年3月.
講演	音楽理論を用いた誤り区間の検出・補正による和音認識 <u>植村あい子</u> , 甲藤二郎, 情報処理学会全国大会, 4S-4, 2012年3月.
講演	DNCOF ベクトルとクロマベクトルを併用した和音認識 <u>植村あい子</u> , 甲藤二郎, 情報処理学会音楽情報研究会, 2011-MUS-90, 5, 2011年5月.
講演	Doubly Nested Circle of Fifths に基づく和音情報と HMM を用いた和音認識 <u>植村あい子</u> , 甲藤二郎, 電子情報通信学会総合大会, A-10-8, 2011年3月.
講演	Doubly Nested Circle of Fifths を用いた和音認識の特性改善 <u>植村あい子</u> , 甲藤二郎, 情報科学技術フォーラム(FIT2010), 5G-3, 2010年9月.
講演	Doubly Nested Circle of Fifths を用いた和音認識 <u>植村あい子</u> , 甲藤二郎, 情報処理学会音楽情報研究会, 2010-MUS-84, 5, 2010年2月.