

和音進行を用いた音楽ハイライト検出の一手法と，組み込み機器への応用

A Method for Detecting Highlights of Music by Chord Progressions and Its Application to Embedded Systems

莪山 真一

Shinichi Gayama

要旨 和音進行を用いた音楽のハイライト検出手法を提案し，組み込み機器に実装した。音楽の特徴部分であるハイライト(サビ)の部分は，楽曲中に他より多く出現する傾向がある。そこで本手法では，楽曲から抽出した和音進行の部分的な相互相関を演算することによって楽曲中の繰り返し部分を抽出し，その結果を用いてハイライトの検出を行った。また，和音進行は少ない情報量で音楽の類似性を表現することができるため，本手法は組み込み機器に求められる効率的な演算を可能とした。その結果，従来手法では性能の確保が困難であったが，本手法は実用上十分な動作速度と性能を示し，JPOP340 曲中，308 曲に対して正確にハイライトを捉えることができた。

Summary The author has proposed a method for detecting highlights of music by chord progressions and applied it to embedded systems. A highlight of music, known by the Japanese word "sabi", which is one of the more expressive parts in a piece of music, tends to appear more frequently than others. So it can be detected by means of identifying similar parts that appear repeatedly in the music, calculating the cross-correlation of partial chord progressions extracted from the waveform of music in advance.

The author has also succeeded in achieving the efficient calculation required of embedded systems, because chord progressions are capable of representing similarity of music with little information. As a result of this implementation, although it had been difficult for previous methods to meet the above requirement, the proposed method can produce sufficient performance for practical use, and can detect highlights of 308 songs correctly among the 340 Japanese pops music used in the experiment.

キーワード : ハイライト，サビ，和音進行，繰り返し構造，相互相関演算，
組み込み機器

1. まえがき
大容量 HDD の普及と音楽圧縮技術の進化に伴い，PC のみならずカーナビゲーションシステムを始めとした組み込み機器においても，多くの音楽を蓄積することが可能となった。また，音楽配信サービスも普及の兆しを見せており，

音楽を楽しむ手段は確実に拡大しつつある。

しかしながら一方で、大量の音楽を個人で管理することや、必要に応じて楽曲を選択することが次第に困難になってきたことから、使い勝手を向上させるためのさまざまな検索手段や再生方法が提案されている。中でも音楽的印象的な部分、すなわちハイライトを検出して音楽を再生する手法は、多くの音楽を短時間で試聴することを可能とし、検索時間の短縮のみならず、新たな音楽の楽しみ方を提供するものとして期待されている。音楽ハイライト検出に係わる研究としては、後藤によるサビ区間検出手法^{(1)・(2)}が新しく、サビを含む音楽の繰り返し構造を分析し、任意の部分を再生する手法およびシステムを発表している。

一方、カーナビゲーションシステムなどの組み込み機器への応用を考えた場合、同時に動作する他の機能との兼ね合いが重要となるため、各機能に要求される演算速度とメモリ量の制約が極めて厳しくなる。従来手法では、それらの要求に対して十分な性能を確保することが困難であった。

本研究の目的は、音楽の和音進行を抽出し、組み込み機器において実用可能なハイライト検出手法を開発することである。和音進行は、音楽の基本的な構造を成すことで知られている。また、和音進行は少ない情報量で音楽の類似性を表現することができるため、音楽の繰り返し構造を検出する目的においては、効率的な特徴量である。

本稿では、筆者の和音進行抽出手法^{(3)・(4)}を用いた音楽のハイライト検出手法を提案し、アルゴリズムの概要を説明する。そして、本手法を組み込み機器に実装して得られた動作パフォーマンスと、ポップス音楽を対象としたハイライト検出性能について報告する。

2. 音楽ハイライト検出

2.1 概要

ポップスやロック音楽の多くは、いくつかの

パートの組み合わせで構成されている。例えば、歌曲で言う1番、2番という大まかな構成や、Aメロ、Bメロ、サビといった詳細な構成が一般的に用いられている。一般に、サビと呼ばれる部分は、その音楽を特徴付ける目的で用いられるため、他より多く出現するが多い。作曲家の立場からも、サビ部分をいかに効果的に繰り返すかが、音楽的印象を決定付けるコツであると述べている⁽⁵⁾。

図1に一例を示す。A、A'(以下A群)やC、C', C''(以下C群)では、同様の曲調が繰り返されている。後藤⁽²⁾が指摘している通り、最も繰り返されるC群がいわゆるサビの集合である可能性が高い。本稿においても同様の発想で、最多繰り返し区間C群を検出し、その一つをハイライトとして捉える。人間が音楽を聴くと、C群を同様なフレーズとして認識することができるが、実際には旋律の動きや伴奏、リズムが異なっている場合が多く、単純に楽音信号やスペクトルの比較を行っても繰り返しの検出はできない。しかしながら、人間がそれを可能とするのは、各部分の曲調が類似していることを認識できるからである。したがって、計算機でその共通の曲調を表現できれば、人間の理解に近い形でC群を検出することができる。

そこで筆者は、共通の曲調を表す音楽的要素として和音進行に注目した。本章では、和音進行を用いて音楽ハイライトを検出するアルゴリズムについて説明する。

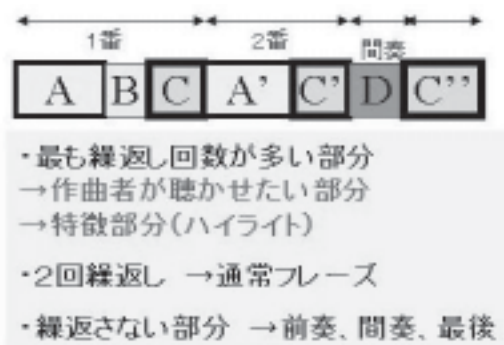


図1 繰り返しを持つ音楽構造の例

まず、最初に各楽曲の楽音信号から和音進行を抽出する。次に一曲全体の和音進行と、所定の位置から切り出した部分和音進行の相互相関演算によって、曲調の類似する最多繰り返し区間の検出を行う。そして最後に、繰り返し区間の中からハイライト位置を選択する処理を行う。以下、各々について説明する。

2.2 楽音信号からの和音進行抽出

対象とする楽曲の楽音信号に対して、筆者が提案した和音進行抽出手法⁽³⁾⁽⁴⁾を用いて和音進行を抽出する。具体的には、リニアPCM形式の楽音信号に対する周波数解析の結果から時系列の和音候補を抽出し、その後の補正処理によって和音進行を生成する。和音進行は、一部の経過和音を除くと、音楽の速さに対して比較的緩やかに変化する傾向がある。また、筆者が提案した類似楽曲検索⁽³⁾⁽⁴⁾や、本手法を実現する音楽の構造解析といった目的に対しては、すべての音程を含む複雑な和音ではなく、主要な響きを成す単純な3和音(3和音とは3つの音程で和音名が確定する和音を言う)の候補を検出する⁽³⁾⁽⁴⁾ことによって、わずかな曲調の変化に対する感度を低く抑えるとともに、和音進行の数値演算を容易にすることができる。以上の根拠に基づいて和音進行を抽出すると、そのデータ量は非常に少なく、次に述べる和音進行の相関演算など、各種演算処理の負荷を軽減する効果が期待できる。

2.3 最多繰り返し区間の検出

前項で説明した手法で楽曲から和音進行を抽出した後、図1のC群に相当する最多繰り返し区間を検出する。曲調の類似する部分(以下、繰り返し部分)の出現回数を検出するために、一曲全体の和音進行の先頭から8~13和音長の部分的な和音進行(以下、部分和音進行)を切り出し、全体の和音進行との相互相関演算を行う(図2)。以降、切り出し位置を時間方向にずらしながら同様の処理を継続する。

繰り返し部分の長さは各楽曲によって異なるため、部分和音進行の長さを一意に決定することは一般に困難である。問題となるのは、部分和音進行の長さに相当する時間長が、実際の繰り返し部分の時間長を超えている場合であるが、本手法で用いる8~13の部分と音進行列は、実時間に換算して約10~15秒に相当することが実験的に分かっている。また、多くの音楽のサビは短くても15秒程度は継続するため、本手法では楽曲全体の和音長に応じて、8~13の部分と音長を採用することとした。

以上の演算を行った結果の一例を図3に示す。対角線上の縞以外に出現する縞が、曲調の類似する繰り返し区間を示しており、対角線を含む出現回数が繰り返し回数を意味する。

なお、和音進行の相関演算においては、転調前後の繰り返し区間に対応するために、絶対和

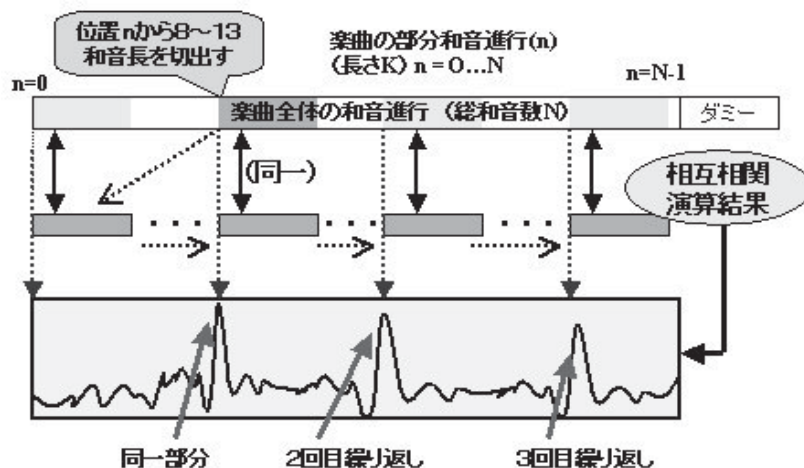


図2 部分と音進行を用いた繰り返し部分の検出

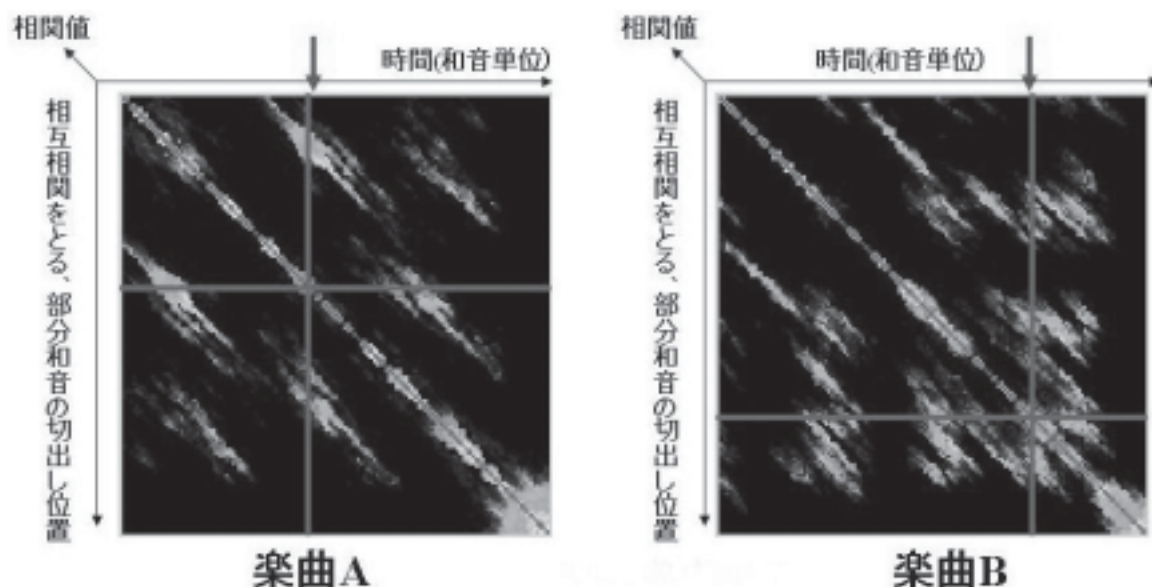


図3 繰り返し区間の演算結果とハイライト位置

音の比較ではなく、前後の和音間に生ずる度数の差分値列と属性列(メジャー、マイナー)すなわち相対的な和音の動きを用いる(以下、和音差分値列)。

両者の和音進行の相関距離としては、基本的に各和音変化点における和音差分値列のユークリッド距離を計算し、それらを加算した値を用いる。前述したように、本手法で検出した和音進行は、わずかな曲調の変化に対して感度が低く、繰り返し部分の和音進行は概ね共通である場合が多い。しかしながら、繰り返し部分でも部分的な編曲が異なる場合には、和音進行が完全には一致しない。そこで、和音差分値列を比較する際に、同一変化点の距離を演算すると同時に、1～2つ先の変化点との距離も計算する。そして、現時点の距離よりも先の和音差分値との距離が近い場合には、比較位置を先に移動し、それ以降は新たな比較位置を基準に、互いの和音差分値列の距離計算を継続する。

以上の工夫によって、転調前後や、部分的に編曲が異なる繰り返し区間も容易に検出することができる。

最多繰り返し区間を検出するには、まず図3における横軸に沿って、各々に対する相関演算結果を縦軸の方向に加算し、最も相関度の高い位置(和音単位)を検出する。その後、当該位置に対する相関演算結果を縦軸の方向に極大値を検索することで、最多繰り返し区間のすべての開始位置を検出することができる。

2.4 ハイライト位置の選択

図1のC群に相当する最多繰り返し区間を検出した後、その内の一つをハイライト位置として選択する。音楽のサビは複数存在するという本手法の前提から考えると、最多繰り返し区間に優劣を付加する必要はない。しかしながら、多くの音楽を短時間で連続的に試聴するという実用上の目的から、複数のハイライト候補から最も視聴に適した部分を選択する必要がある。そこで本手法では、前項で検出した最多繰り返し区間の各開始位置から、約15秒の範囲に対して各区間の信号パワーの平均値を求め、それが最も大きい区間の開始位置をハイライト位置として選択する。この処理は、複数のサビの中から、最も盛り上がりを感じられる部分を選択することに他ならない。

3. 組み込み機器への実装

3.1 概要

本章では、本手法の組み込み機器への実装方法について説明する。対象とする機器には、大容量のHDDが搭載されており、多くの音楽を録音することができる。実装方法としては、各楽曲の楽音信号から時系列の和音候補を抽出する^{(3)・(4)}リアルタイム処理を16ビット固定小数点のDSPに、そして補正処理によって和音進行を生成する処理^{(3)・(4)}と、ハイライト検出処理を動作周波数400MHzの組み込みCPUに実装した(図4)。

3.2 リアルタイム処理

時系列の和音候補を抽出する処理は、音楽の録音と同時にされる。まず、録音される楽音信号に対してモノラル変換とダウンサンプリング処理を行う。次に、同楽音信号に対して離散フーリエ変換を行い、平均律音程に該当する周波数パワーを演算する。その後、各音程の周波数パワーの大きさから、時系列の和音候補群を演算する。なお、各和音候補は3つの音程で構成される3和音で表現されている^{(3)・(4)}。

3.3 和音進行生成処理とハイライト検出処理

各楽曲が録音されるごとに、時系列の和音候補群に対する補正処理⁽⁵⁾を施し、和音進行を生成する。その後、和音進行の変化点のみの情報を記憶し、2.3および2.4に記述したハイライト検出処理を実行する。

4. 性能評価

4.1 概要

本手法の有効性を確認するために、3章に記述した方法で組み込み機器に実装し、動作パフォーマンスおよびハイライト検出性能の点から性能評価を行った。以下、各々について報告する。

4.2 動作パフォーマンス

3.2に記述したリアルタイム処理では、ダウンサンプリング処理などの信号処理に対して約3MIPS、時系列の和音候補検出処理に対して約1.1MIPSで計約4.1MIPSを要した。対象とした楽曲フォーマットは、サンプリング周波数44.1kHz、リニアPCM形式であり、所要コードサイズは約9kバイト、ワークメモリは約17kバイトである。

また、3.3に記述した和音進行の生成処理およびハイライト位置の検出処理では、CPU占有率100%の状態において、3分の楽曲に対して平均150msec、6分の楽曲に対して平均400msecの処理時間を要した。

DSPとCPUの性能も飛躍的に向上しており、本結果は十分実用レベルであるといえる。ただし、実際のシステムにおいて他の優先度の高い機能が動作している場合には、上記所要時間が3~5倍になることも十分に考えられる。そのため、システム全体を安定に動作させるための設計が重要である。

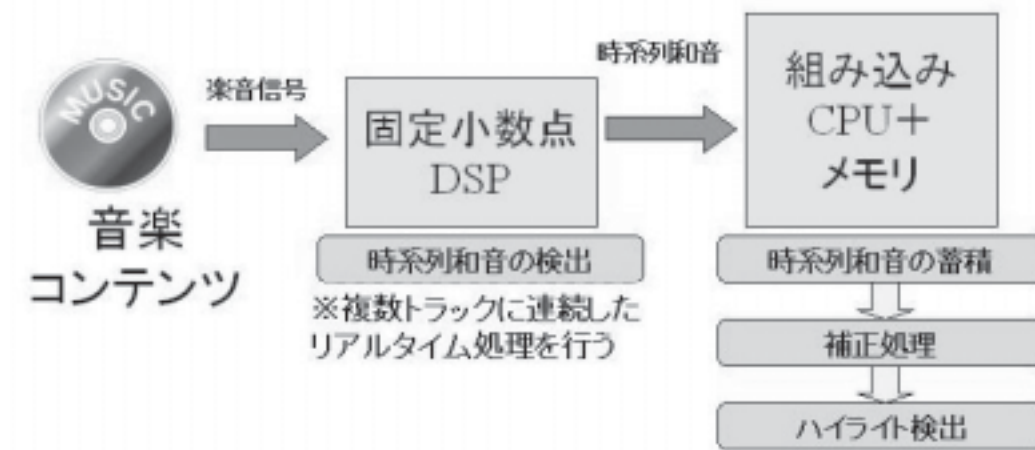


図4 組み込み機器への実装方法

4.3 ハイライト検出性能

ハイライト検出の評価楽曲には340曲の日本のポップス音楽(JPOP)を用いた。本手法の実用上の目的は、操作手段の限られた組み込み機器において、録音を行った多くの楽曲を簡単に試聴することである。そのためには、録音した各楽曲のハイライトを15～20秒程度再生できることが望ましい。したがって評価方法としては、ハイライト検出位置から5秒以内で各楽曲のサビを再生できた場合を正当とし、その曲数をカウントした。結果として308曲は良好な結果が得られ、約9割の正当率を確保した。

検出できなかった音楽の特徴として、サビ以上に多く繰り返される類似区間が存在する、あるいは楽曲自体が明確な繰返し構造をもっていない場合があげられる。また、元々ハーモニーに乏しい楽曲に対しては、和音進行の検出精度が悪化するため、結果的にハイライト検出精度の低下に影響していると考えられる。

5. まとめ

和音進行を用いた音楽のハイライト検出手法を提案し、組み込み機器への実装と評価を行った。組み込み機器に適用する場合には、演算速度とメモリ量の制約があり、従来手法では性能を確保することが難しかったが、本手法を実装した結果、十分な動作速度と性能が得られた。

本手法は、和音進行の類似性を用いてハイライトの検出を行っているため、人間が判別できるサビの開始位置を厳密に検出することはできない。しかしながら、組み込み機器における安定動作と、十数秒間のハイライト再生機能の実現という実用性を考慮すると、現状は十分な性能といえる。今後のDSPやCPUの性能向上に伴い、同課題を解決できる余地はある。

なお、本手法は実際にカーナビゲーション製品に搭載されており、楽曲のハイライト再生機能に用いられている。

6. 謝辞

本研究を進めるにあたり、協力頂いたモバイルシステム開発センターおよびモバイルエンターテインメントカンパニーの各位に感謝する。

参 考 文 献

- (1) 後藤真孝：“リアルタイム音楽情報記述システム：サビ区間検出手法”，情報処理研究報告，2002-MUS-47-6, Vol. 2002, No. 100, pp. 27-34, October 2002
- (2) 後藤真孝：“SmartMusicKIOSK：サビ出し機能付き音楽試聴機”，情報処理学会論文誌，Vol. 44, No. 11, pp. 2737-2747, November 2003
- (3) 莪山真一：“類似楽曲検索を目的とした楽音からの和音進行抽出手法”，FIT2003，一般講演論文集，Vol. 2, p245-246, 2003
- (4) 莪山真一：“楽音信号からの和音進行抽出手法と類似楽曲検索への応用”，パイオニアR&D, No. 2, p1-7, 2004
- (5) 林哲司：“新ポップス作曲法 Creators Handbook”，リットーミュージック出版，1999

筆 者

莪 山 真 一 (がやま しんいち)

所属：研究開発本部 総合研究所 情報メディア技術研究部

入社年月：1991年4月

主な経歴：アクティブノイズキャンセラの開発，デジタル無線変復調技術の開発，エージェント技術，音楽構造化技術の研究を経て，現在に至る。

得意分野，技術：ソフトウェアによるアルゴリズム研究とアプリケーション開発。特に，マルチメディア情報の特徴抽出と検索技術に興味を持つ。