

## 標準MIDIファイルからのメロディの自動抽出法

鷲坂光一

wasisaka@rouge.ntt.jp

Ⓞ NTT 基礎研究所

音楽データベースでの意味検索を実現するために、音楽データとして標準MIDIファイル(Standard MIDI File)を使用し、音楽の特徴を表すメロディを含むチャンネルを、簡単な分野知識を用いて自動的に選び出す方法について述べる。本手法では、メロディを特徴づけるパラメータを用いて各チャンネルを表し、メロディとして最適なチャンネルを選出する。この手法を224個の標準MIDIファイルに適用したところ、181個(80.8%)の標準MIDIファイルでメロディを含んだチャンネルを正しく選び出すことができた。

## Automatic Merody Extraction from Standard MIDI File

WASHISAKA Mitsukazu

NTT Basic Research Laboratories

Midori-Cho 3-9-11, Musashino-Shi, Tokyo 180, JAPAN

To implement a semantic information retrieval from music database, melody that characterizes the music data must be extracted from it. In this paper, we describe a melody extraction method from Standard MIDI File(SMF) music data using some simple domain knowledge. This method characterizes each channel by the parameters such as pitch extent and selects most suitable channel. Applying this method to 224 SMFs, we can get 181(80.8%) correct melody channels.

## 1 はじめに

計算機上で画像や音楽等のさまざまなマルチメディアデータが利用可能になり、オーサリングシステムでデモやプレゼンテーションを作成するために、これらのデータを利用する機会も増えつつある。このような場合に、必要なマルチメディアデータを検索して利用するには、高レベルで内容検索が行なえ、かつ再利用可能な形でデータを取り出せるマルチメディアデータベースが望まれる。

現在、マルチメディアデータベースとしては、絵や図形を検索対象にした画像データベースが主に研究されているが、マルチメディアを対象とした情報処理では絵や図形といった静止画像だけでなく、アニメーションやムービー等の動画像、音声や音楽等の音もデータとして扱える必要がある。また、検索が容易に行なえるように、高いレベルで検索内容を指定できる枠組みも必要である。

このため、我々は意味検索可能なマルチメディアデータベースを構築するための第一歩として、検索データとして音やメロディを対象にした音楽データベースの研究を展開している。

## 2 意味検索

マルチメディアデータベースの内容検索方法としては、画像データベースで画像の輪郭を指定したり [1]、音楽データベースでメロディを指定して検索を行なう [2] 類似検索や、「楽しい」「悲しい」等の感性的なパラメータの度合いをデータから抽出する [3]、またはあらかじめマルチメディアデータに付与しておき、その感性表現を指定して検索する感性検索がある。

これらの検索よりもレベルの高い検索方法として、画像データベースで「サッカーのシーン」を指定したり [4]、音楽データベースで「あるメロディをモチーフにした曲」といった指定をして検索する方法がある。このように、検索データおよび質問内容からその意味を抽出し、その意味世界で類似検索を行なう方法を意味検索 (Semantic Retrieval)

と呼ぶ。

意味検索を行なうには、データおよび質問内容からその意味を抽出してシステムに与える必要がある。音楽データベースの場合、音楽データの意味を抽出するには、まず音楽データからその特徴を表す、メロディ (主旋律)、コード進行 (和音)、リズム等を取り出す必要がある。

## 3 関連研究

従来の音楽データベース [5] では、音楽分析を目的にしており、音楽データをテキストベースで検索していた。このため、実際の音楽からメロディだけを人間が選び出し、楽譜記述言語に変換する作業が必要であった。この方法には次の問題点がある。

- 大量の音楽データを独自に入力する必要があるため、大規模データベースの構築が容易でない。
- 演奏時間の記述が不明瞭である、あるいはメロディしか音楽データを持っていないため、リアルな演奏を再現できない。
- 独自の楽譜記述言語を使用しているため、検索の結果得られたデータを他のマルチメディアソフトに容易に取り込めない。
- 音楽をテキストに変換して検索条件を指定するため、音楽的知識以外に楽譜記述言語の知識が必要である。

本稿では、まずこれらの問題を解決するために、音楽データベースの音楽データとして世界標準規格である標準 MIDI ファイル (Standard MIDI File) を使用する。

また、音楽からのメロディの抽出に関しては、音響信号処理および音楽理論を用いて、あるいは人間がメロディをどのように認識しているかといった心理学的実験に基づいて研究が行なわれている

が、一般に多重音からメロディを抽出するアルゴリズムは確立されていない。

そこで本稿では、音楽データベースで意味検索を行なうために必要なメロディを、簡単な分野知識を用いて、標準 MIDI ファイルから自動的に抽出する方法について述べる。

## 4 音楽データベース

### 4.1 General MIDI と標準 MIDI ファイル

General MIDI (Musical Instrument Digital Interface)[6] は、さまざまな電子楽器を接続して演奏 / 制御するためのインタフェースの世界統一規格である。

鍵盤を押す (Note On)、鍵盤を離す (Note Off)、音色 (楽器) を変更する (Program Change) 等の MIDI メッセージが定義されている。使用可能な音色の数は 128 種類あり、16 個のチャンネルにより最大 16 の音色を同時に演奏することができる。

標準 MIDI ファイルは MIDI メッセージをまとめた音楽の実演奏データのファイルで、4 分音符の長さやテンポ等の演奏に必要な情報、送出する MIDI メッセージとその送出時間等が送出時間順に並べられている。

### 4.2 音楽データとしての標準 MIDI ファイル

音楽データベースの場合は、検索の結果得られた音楽データが目的のデータであるかどうかを実際に試聴してインタラクティブに確認できることが求められるため、音楽データとしては音楽記述言語で書かれたテキストよりも実演奏データの方が望ましい。

General MIDI の標準 MIDI ファイルは、音楽の実演奏データであり世界統一規格である、大量のデータが容易に入手 / 作成可能である、電子楽器で容易に演奏が可能である、アナログデータに比べてサイズが小さい、マルチメディアソフトにそのまま取り込むことができる等の理由から、音

楽データベースの音楽データとして使用するのが適切だと考えられる。

## 5 メロディの自動抽出法

音楽データベースで意味検索を行なうには、音楽データからその特徴を表すメロディを抽出する必要がある。また音楽データベースでは、音楽データの検索条件として、曲名、作曲家名を指定する以外にも、フレーズ (メロディの断片) を指定することも必要である。

このように意味検索や検索条件にメロディやフレーズを指定した検索を行なうためには、あらかじめデータベース中の音楽データからメロディやフレーズを抜き出しておく必要がある。しかし、大規模な音楽データベースでは音楽データのメロディ部分を人間が判断し抜き出すことは、非現実的である。

ここでは、標準 MIDI ファイルからメロディを自動的に取り出す方法について述べる。この際、メロディは複数のチャンネルに分散されているのではなく、16 チャンネルのうちの 1 つのチャンネルを占有していると仮定とする。また、メロディを取り出す標準 MIDI ファイルには、メロディの存在が明らかであるように歌曲曲 (人間の歌う部分がメロディ) の標準 MIDI ファイルを使用する。

### 5.1 メロディを演奏しない音色の除去

音楽データからメロディを取り出すためには、メロディを演奏する音色を特定できれば良いが、一般にはほとんどの音色がメロディを演奏する可能性があるため、音色からメロディを特定するのは不可能である。しかし逆に、メロディを演奏する可能性のない音色 (ドラムの音色や電話の音、ヘリコプタの音等の SFX の音色) は特定できる。

General MIDI では 128 種類の音色とその音色番号が定義されており、標準 MIDI ファイルにはチャンネルと音色の対応を指定する Program Change と呼ばれる MIDI メッセージが入っているため、

各チャンネルで使用されている音色がわかる。従って、実際の音楽でメロディを演奏する可能性のない楽器の知識を基に、標準 MIDI ファイル中の Program Change メッセージと音色番号から、メロディでないチャンネルを除外できる。

次に残ったチャンネルからメロディを演奏しているチャンネルを選び出すために、各チャンネルごとに次の知識を適用して、メロディを構成していると考えられる音を選び出す。

## 5.2 音符と打鍵時間の正規化

メロディを構成する音を決定するためには、どの音とどの音が同時に演奏されているかを知る必要がある。

標準 MIDI ファイルは楽譜を基に機械的に生成されるだけでなく、人間が電子楽器を演奏して生成されることも多いため、演奏者や制作者のアレンジにより打鍵(鍵盤を押す)/リリース(鍵盤を離す)タイミングが完全に楽譜と同じではなく、同時に押されるべきはずの音が標準 MIDI ファイルでは微妙にタイミングがずれている場合がある。

演奏者や制作者による演奏の違いを吸収し、メロディを構成する音を正しく選択するために、標準 MIDI ファイルに記述されている4分音符の長さを基に、打鍵/リリース時間から実際の楽譜上での音符と打鍵時間(音が鳴っている時間)を推定し、正規化を行なう。

## 5.3 メロディを構成する音の選択

正規化を行ない各チャンネルを調べた結果、ある時間に演奏されている音が1つ(単音)であるならば、その音がメロディを構成している音と考えられる。しかし、音楽によっては同じ音色でメロディと伴奏を演奏していることや和音を演奏していることがあるため、複数音が同時に演奏されていることがある。このような場合、伴奏部分を除外したり、同時に弾かれている複数音からメロディを構成している1音を選択するために以下の知識を利用する。

- 音色がピアノの場合、一般的に右手はメロディを左手は伴奏を弾くことが多い。さらに低音よりも高音の方が人間の印象に残るため、メロディの構成音になりやすい。従って、複数音が同時に演奏される場合は、最高音をメロディの構成音とする。
- 単音である場合でも、非常に低い低音や、直前に演奏された音と比べて音の高さが著しく跳躍(変化)する音は伴奏の一部やメロディを構成していない音と考えられるため、メロディの構成音からは除外する。
- 複数音が同時に演奏される場合に、ある音が同時に演奏されている他の音よりも非常に大きな音量で演奏されている場合は、音量の大きな音がメロディを構成している。

## 5.4 フレーズへの分割

一般にメロディは複数のフレーズから構成される。これらのフレーズは、楽譜上では休止符と休止符の間のメロディ断片を意味している場合が多い。このため、打鍵/リリース時間を基に休止符部分を推定し、得られたメロディをフレーズごとに小さく分割する。

メロディを構成するフレーズは適当な長さ(音数)と音域を持つため、分割して得られたフレーズのうちメロディの構成音を数音しか含まないものや音域の非常に小さいものは、前処理で除去しきれなかった伴奏の断片であるか、伴奏を演奏しているチャンネルから得られたフレーズであると推定できる。そこで、メロディの構成音としては疑わしい音として印をつける。

以上の処理によって各チャンネルごとにメロディを構成していると考えられる音とその中で疑わしい音を選択する。

## 5.5 メロディチャンネルの選択

各チャンネルごとにメロディ構成音として選択された音の情報を基に、メロディの演奏されている

チャンネルを選択する。

人間が音楽中のある音色の音列をメロディとして認識するのは、その音列がなんらかの特徴を持っているためと考えられる。その特徴として次の項目を用いる。

- [音符] 標準 MIDI ファイルとして歌謡曲を仮定しているため、人間が歌うことを考えるとメロディを構成する音符は、ほとんどが8分音符、4分音符程度の長さを持つ。しかし、標準 MIDI ファイル中に指定された4分音符の長さを基に楽譜上に正規化された時には16分音符であっても、演奏速度が関係するため実際に演奏される時間は別の楽譜の8分音符より長いような場合がある。ここでは、4分音符の長さを500msec(1分間に4分音符が120個)だと仮定して、各音符の割合を測定する。
- [演奏時間] ほとんど演奏時間のない音列がメロディとして認識されることは少ない。メロディはその音楽中で一定以上の割合の演奏時間を持つ。
- [音域と平均音価] 標準 MIDI ファイルを歌謡曲に限定したため、その音域は人間が歌うことのできる範囲でありある程度限定される。また、メロディはある程度の音域に渡って音分散している。
- [フレーズの長さ] メロディのフレーズはある程度の長さを持つ。また、弾き続けられるような伴奏とは異なり、歌うことを考慮すると、フレーズが極端に長くなることはない。
- [疑わしい音数] 構成音のうち疑わしいと思われる音の割合が、メロディの場合は伴奏に比べて少ない。

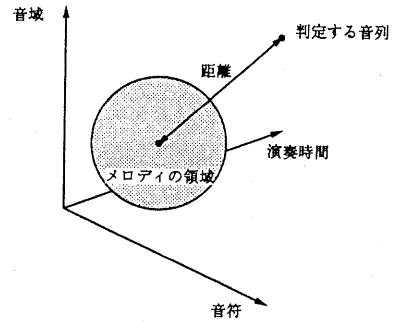


図1: メロディの領域と距離

## 6 チャンネルの選出方法

### 6.1 選出方法

各チャンネルごとに上記の項目を比較して、最も条件に適したチャンネルをメロディのチャンネルとして選び出した。

これらの項目を評価する場合に、不適切な項目があるチャンネルから順に除外していくという方法が考えられるが、1項目だけが不適切であるようなメロディがある場合や、最終的に1つのチャンネルだけに絞り込めない場合があるため、各項目ごとの総計を比較する方法をとった。

具体的にはメロディはこれらの項目をパラメータとする空間上に、メロディとして認識される領域を構成すると仮定した。その中心からの距離に相当する値を計算し、最も距離の小さいチャンネルをメロディのチャンネルとして選び出した{図1}。

基本となるメロディの領域を設定するために、あらかじめメロディのわかっている80個の標準 MIDI ファイルから、メロディが含まれているチャンネルに対して前述したメロディ構成音を選ぶ処理を行ない、各項目の平均と分散を求めてメロディの領域を設定した。

領域の中心からの距離は、次の式によって計算した。

$$\text{距離} = \sum_{i=1}^n w_i * ((s_i - v_i) / \sigma_i)^2$$

$v_i$  は距離を求めるチャンネルの項目  $i$  の値,  $s_i$  は項目  $i$  の平均値,  $\sigma_i$  は項目  $i$  の分散である. 分散で割ることで各項目の正規化を行なっている.

また,  $w_i$  は項目  $i$  の重みである. 例えば, 音域に関する項目よりも演奏時間の方がメロディの認知には関係が深いと考えられる. さらに, 平均より演奏時間が短い場合よりも長い場合の方がメロディである可能性が高いと考えられる. 重みはこのような項目間のあるいは項目内のメロディの認知に関係する要因を反映している.

## 6.2 適用例

本稿で述べた自動抽出法を用いて, 224 曲の標準 MIDI ファイルからメロディのチャンネルを選び出したところ, 181 曲 (80.8%) の標準 MIDI ファイルで正しくメロディのチャンネルを選び出すことができた. この際の, 各標準 MIDI ファイルの平均使用チャンネル数は 10.5 チャンネル (224 曲でのべ 2341 チャンネル使用) であった.

使用した標準 MIDI ファイルを歌謡曲に限定したため, 音楽理論をほとんど使用することなしに, 簡単な分野知識だけで, ある程度の正解率を上げることができた.

## 7 まとめ

General MIDI の標準 MIDI ファイルを使用して音楽データベースを構築することの利点を示すとともに, 音楽データベースの意味検索を行なうために必要なメロディが含まれたチャンネルを, 標準 MIDI ファイルから簡単な分野知識を用いて自動的に選び出す方法について述べた. 本手法を発展させることにより, 例えばデモやプレゼンテーションを作成するオーサリングシステムで, 標準 MIDI ファイル音楽データベースをインタラクティブに検索することで背景音楽等を選択することが可能となる.

本手法でのメロディの認識率を向上させる, あるいはメロディを正しくフレーズに分割するには,

ドミナント進行 [7] のようなコード進行の音楽理論に基づいた知識や, グローバルなメロディの繰り返し構造等の分析が必要であると考えられる. 今後は, このような音楽理論の組み入れと共に, 抽出されたメロディを意味空間にマッピングする方法 [8] について検討を加えていく予定である.

## 参考文献

- [1] 加藤, 栗田: 画像の内容検索 - 電子美術館への応用 -, 情報処理, Vol.33, No.5, pp.466-477, 1992.
- [2] Kageyama, T., Mochizuki, K., Takashima, Y.: Melody Retrieval with Humming, ICMC Proceedings, pp.349-351, 1993.
- [3] 金森他: コードおよびメロディー・パートからの感性情報の抽出, 情処研報, Vol.92, No.89, 92-HI-45, pp.75-79, 1992.
- [4] 坂内, 佐藤: 画像データベースにおけるモデル形成, 信学論 D-I, Vol.J74-D-I, No.8, pp.455-466, 1991.
- [5] 山本: 音楽データベース, 情報処理, Vol.29, No.6, pp.599-607, 1988.
- [6] MIDI 規格協議会: MIDI 1.0 規格, 1989.
- [7] 島岡: 音楽の理論と実習 I, 音楽之友社.
- [8] Balzano, G.J.: The Grouptheoretic Description of 12-Fold and Microtonal Pitch Systems, Computer Music Journal, Vol.4, No.4, pp.66-84, 1980.