

# 効率的な試聴のためのマルチレベル音楽要約

堀野 義博<sup>†</sup> 黒木 進<sup>†</sup> 北上 始<sup>†</sup>

<sup>†</sup> 広島市立大学大学院 情報科学研究科

〒 731-3194 広島市安佐南区大塚東 3-4-1

E-mail: †{horino,kuroki,kitakami}@db.its.hiroshima-cu.ac.jp

あらまし 楽曲を効率よく評価するためには重要なフレーズを聴取するのが便利であるので、われわれはユーザの求める詳細度に応じてダイジェストを生成するマルチレベル音楽要約法を提案する。この要約法は5段階の要約を提供する。これらの要約は、重要フレーズのうちのひとつを演奏する(レベル1)、すべての重要フレーズを一度ずつ演奏し、その他はスキップする(レベル2)、すべての重要フレーズを一度ずつ演奏し、その他は早送りする(レベル3)、すべての重要フレーズおよびそれらの類似フレーズを演奏し、その他は早送りする(レベル4)、曲全体を演奏する(レベル5)というものである。

キーワード 音楽DB, 楽曲要約

## Multi-Level Music Summarization for Efficient Listening

Yoshihiro HORINO<sup>†</sup>, Susumu KUROKI<sup>†</sup>, and Hajime KITAKAMI<sup>†</sup>

<sup>†</sup> Graduate School of Information Sciences, Hiroshima City University

3-4-1, Ozuka-higashi, Asaminami-ku, Hiroshima-shi, Hiroshima, 731-3194, Japan

E-mail: †{horino,kuroki,kitakami}@db.its.hiroshima-cu.ac.jp

**Abstract** Because it is convenient to listen to important phrases to evaluate them efficiently, we propose a multi-level music summarization method which generates a digest according to user's level of detail. The summarization method presents five levels of summarization. They play one of important phrases (level 1), play all the important phrases once and skip the rest (level 2), play all the important phrases once and fast-forward the rest (level 3), play all the important phrases and their similar phrases once and fast-forward the rest (level 4), play overall pieces (level 5).

**Key words** Music Database, Music Summarization

### 1. はじめに

現在の楽曲検索システムとしては、要求に合う楽曲を検索し、それらを列挙するものが一般的である。これらのシステムでは条件によって楽曲を絞り込むことができるが、利用者のイメージに最も合致する楽曲を1つだけ出力することは多くの場合困難である。そのため、出力された楽曲をそれぞれ評価し、最良の1曲を選択する必要がある。

楽曲について評価を行うには、作曲者などの書誌情報や楽曲の印象を表す感性語などが参考になるが[1][2][3]、最終的には楽曲を聴取することが不可欠である。楽曲の聴取は1曲あたり数分の時間が必要となり、さらに曲数が増えた場合などを考え

ると時間的な負担という問題を無視できなくなる。この問題点を解決するために、楽曲中から重要な部分を抜き出すような楽曲要約システムがいくつか提案されている[4][5][6]。

平田ら[4]はGTTMによる楽曲の構造化を行い、対話的に楽曲要約を行うシステムを提案している。このシステムは非常に正確な要約曲を生成できるが、4小節の要約を行うのに約3時間の時間がかかると報告されており、大規模な音楽データベースに対応することができない。

それに対してわれわれ[5][6]は小節、フレーズの出現頻度を基にした要約法を提案している。この方法は要約の正確さよりも簡易で高速な要約を目指している。この方法では1つの楽曲の要約を数秒で生成できる。

これらの要約法では、楽曲内の重要なフレーズを列挙もしくは連結して聴取するような方法をとっている。これらの提示方法を元に、本稿では重要フレーズの聴取をさらに効率化することを目指す。まずはじめに列挙された重要フレーズから聴取するものを選択する時点で、各フレーズについての付加情報を提示し、フレーズ選択を支援することを考える。次に、フレーズの提示方法を詳細度に応じてレベル分けすることで、段階的な楽曲評価や利用者の時間的余裕に応じた詳細度の選択が可能となる方法について提案する。

## 2. 重要フレーズの抽出

楽曲を短時間で試聴するためには、楽曲内部から重要であるフレーズを取り出し、それらを聴くようにすればよい。われわれは楽曲内部でのフレーズの出現頻度とフレーズの重要性との間に相関関係があると考え、出現頻度の高いフレーズを重要フレーズとして抽出する方法を提案した[5][6](図1)。

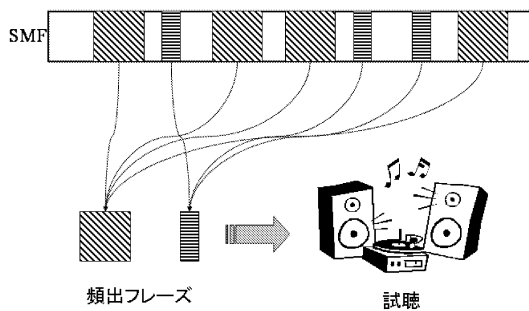


図1 重要フレーズの抽出・試聴

提案した重要フレーズ抽出の流れはおおまかに以下のようになる。

- (1) 重要フレーズ抽出を行う楽曲の標準 MIDI ファイル (Standard MIDI File, SMF) を用意する。
- (2) 楽曲の SMF から旋律に該当する部分を抽出し、それを新しい SMF として保存する。
- (3) 旋律のみの SMF について、すべての小節の楽曲内での出現頻度を求める。
- (4) 出現頻度の高い小節から順に、その小節を核とした区間の伸張を行い、頻出フレーズ区間を決定する。
- (5) 求められた頻出フレーズ区間の開始時刻、終了時刻をファイルに保存する。
- (6) 開始時刻、終了時刻を格納したファイルを参照しながら、元の SMF より頻出フレーズ部分のみを抽出した SMF をフレーズごとに出力する。

この手法によって3~5つのキーフレーズが抽出され、各キーフレーズは全体の1割程度の時間で聴取できる。その結果、

すべてのキーフレーズを聴取するには元の楽曲の3~4割程度の時間でよい。また、さらに聴取するフレーズを取捨選択することで聴取時間の短縮が期待できる。

## 3. 付加情報によるフレーズ選択支援

われわれが[5][6]で提案した手法では抽出した重要フレーズを単に列挙するだけであり、各フレーズの重要度の違いや内容に関する情報を全く提示していない。そのため、提示された重要フレーズから適当なものを選択することが困難である。そこで、重要フレーズを列挙するのと同時に、各フレーズに関する情報を提示するようにする。これにより重要フレーズを取捨選択することが容易になり、フレーズの試聴がより効率的に行える。

重要フレーズの付加情報として以下のようなものが考えられる。

- (1) フレーズの演奏時間
- (2) フレーズの楽曲内での出現確率
- (3) 楽曲内と DB 内の他の楽曲との出現確率の比率
- (4) 類似するフレーズを含む楽曲リスト
- (5) フレーズの特徴を表現するアイコンの表示

(1) は、フレーズを試聴するのにかかる時間である。(2) は、フレーズの核となる小節の楽曲内での出現頻度を楽曲の全小節数で割ったものである。(3) を求めるには、あらかじめ DB 内のすべて楽曲について、小節単位で音列パターンの辞書を作成しておく。現段階ではパターン辞書の作成には Web 上から収集した152曲のクラシック曲を用い、辞書にはパターンの出現回数、出現曲数、パターン情報(音列の各音の相対時刻と音高)を格納している。このようなパターン辞書の内部で各パターンがどれくらいの頻度で出現するかを求める。この出現頻度を辞書内の登録パターン数で割ったものと、前述の楽曲内での出現確率との比率を求める。これにより楽曲内で頻出するフレーズが他の楽曲ではどれくらい出現するかを評価でき、他の楽曲には現れないこの楽曲特有のフレーズであるかどうかを表すことができる(図2)。

(4) については、パターン辞書を作成する段階で、同時にパターンの出所である楽曲名を格納しておくことで、類似するパターンから楽曲名を逆に探索し、それらを提示できるようになる(図3)。

(5) はビデオ映像の注釈づけをするインターフェースとして MediaStream が提案する“Icon Space”を参考にした方法である[7]。“Icon Space”は映像の断片の内容をアイコンで表現する。これを音楽に応用し、フレーズの内容を表現するようなアイコンを提示することでフレーズを選択を支援する。アイコンで提示する内容としては、フレーズ内の音列の音高推移の形状、フレーズを演奏する楽器、フレーズの印象を表す顔などの画像が考えられる(図4)。

このようにフレーズについての付加情報を提示することで、



図 2 付加情報の提示 (1)~(3)

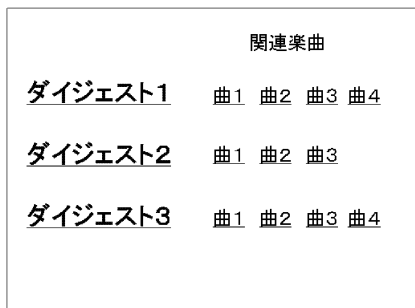


図 3 付加情報の提示 (4)



図 4 付加情報の提示 (5)

それらを手がかりに試聴するフレーズを効率的に選択することができる。

#### 4. 詳細度別マルチレベル要約

従来の研究 [5][6] では重要フレーズを列挙し、そのなかから試聴するフレーズを選択する方法をとっている。ここからさらに詳細な評価をしたいと考えた場合、現段階ではただちに曲全体の試聴をすることになる。しかし、フレーズの個別聴取と曲全体の聴取とは詳細度があまりに違いすぎるため、以下のような問題が発生する。

- 詳細度による段階的な評価ができない。
- 利用者の時間的余裕に合わせた評価ができない。

そこでわれわれは重要フレーズの試聴から曲全体の試聴までの間にいくつかの詳細度レベルを設定し、これによって段階的に詳細度をあげられると同時に、時間的余裕に合わせた要約法を選択できるようなマルチレベル要約について考える。

##### 4.1 詳細度レベル

現在の案として、つぎのような5つの詳細度レベルを設定することを考えている。

**レベル1** 出力された重要フレーズから選択した個々のフレーズを聴取する(図5)。従来の要約法 [5][6] はこのレベルに該当する。ここで3.章で述べた付加情報を同時に提示することで、個別聴取の効率をさらに向上することができる。

**レベル2** 出力された重要フレーズすべてを出現頻度順もしくは出現時刻順に連続して聴取する(図6)。レベル1での重要フレーズの個別試聴とは違い、出力された重要フレーズを一望するような方法である。

**レベル3** 楽曲のうち出力されたダイジェストに該当する部分はそのまま演奏し、それ以外の部分は早送りする(図7)。これはビデオ映像などの例を参考にした手法であり、重要フレーズ以外の部分についても早送りで聴くことでおおまかな特徴を捉えられるようにする。

**レベル4** 楽曲のうち出力されたダイジェストに該当する部分だけでなく、それに類似する部分もそのまま演奏し、それ以外の部分は早送りする(図8)。

**レベル5** 楽曲全体を聴取する(図9)。

利用者は各重要フレーズを個別に聴取するレベル1から始め、必要に応じて詳細度レベルを上げていき、最終的に楽曲全体(レベル5)を得るという段階的な評価をすることができる。また、利用者の時間的な制約に合わせて詳細度を選択するということもできる。

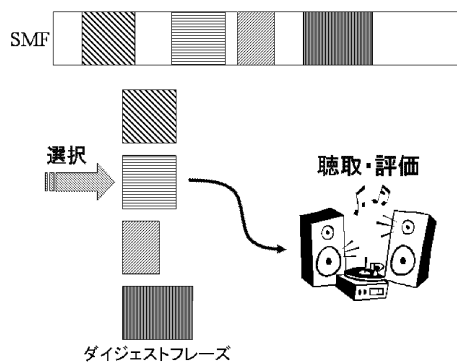


図 5 詳細度レベル 1

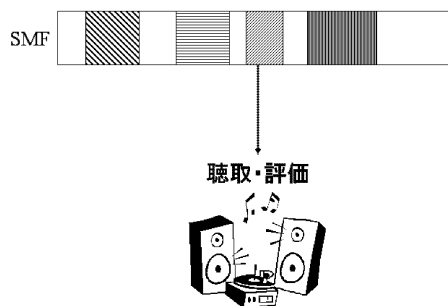


図 9 詳細度レベル 5

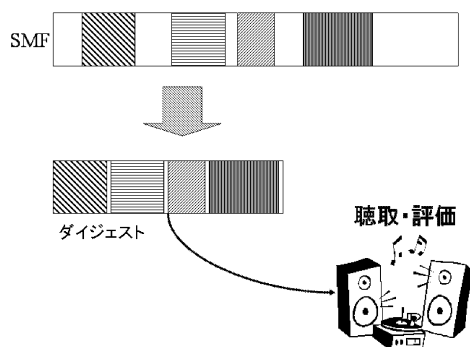


図 6 詳細度レベル 2

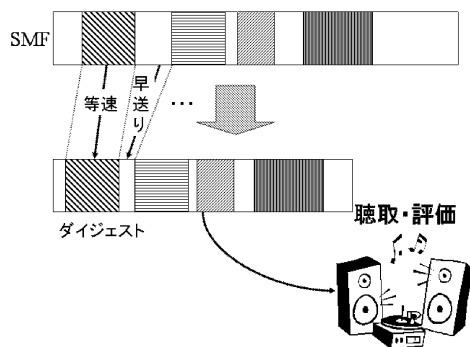


図 7 詳細度レベル 3

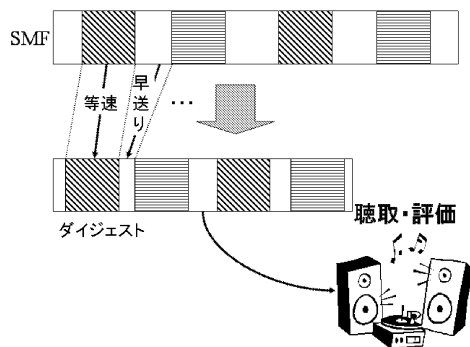


図 8 詳細度レベル 4

#### 4.2 各詳細度レベルの予測聴取時間

各詳細度レベルの聴取時間は図 10 のように予測できる。

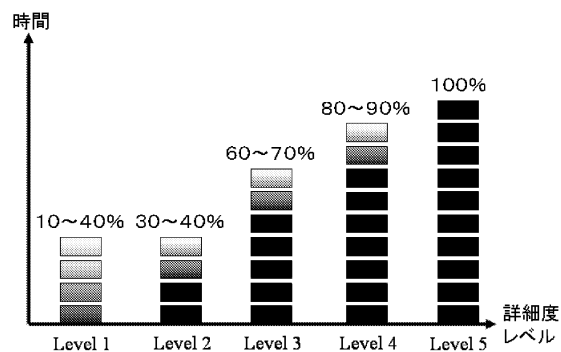


図 10 詳細度ごとの聴取時間の予測値

レベル 1 の場合、聴取するフレーズを任意に選択するので、選択したフレーズ数によって聴取にかかる時間が変わってくる。フレーズ一つあたりの聴取時間は全体の 1 割程度となる。すべて聴取した場合はレベル 2 と同じ時間が必要となる。

レベル 2 に必要な聴取時間は 2. 章で述べたように曲全体の 3 ~ 4 割程度である。

レベル 3 の場合、重要フレーズ部分についてはレベル 2 と同じ時間がかかり、それ以外の部分は早送りとなるので、早送りの速度を倍速と考えれば、レベル 2 の 3 割増し程度の聴取時間になると予測できる。

レベル 4 については、類似フレーズの数によって聴取時間が変わるので容易に予測はできないが、レベル 3 より長く、レベル 5 より短い時間で聴取が可能であることは確かである。

レベル 5 は曲全体の聴取なので、聴取時間はそのままである。

#### 4.3 マルチレベル要約のインタフェース

以上のようなマルチレベル要約を実現するインタフェースを図 11 のように実装した。

各詳細度レベルでのダイジェストを上から順にレベル 1 (キーフレーズ 1 ~ 3)、レベル 2 (連続試聴)、レベル 3 (等速、早送り)、レベル 5 (全曲試聴) と並んでいる。レベル 4 は現時

点では未実装である。さらに、各キーフレーズの横にはキーフレーズの特徴を示す付加情報を提示している。

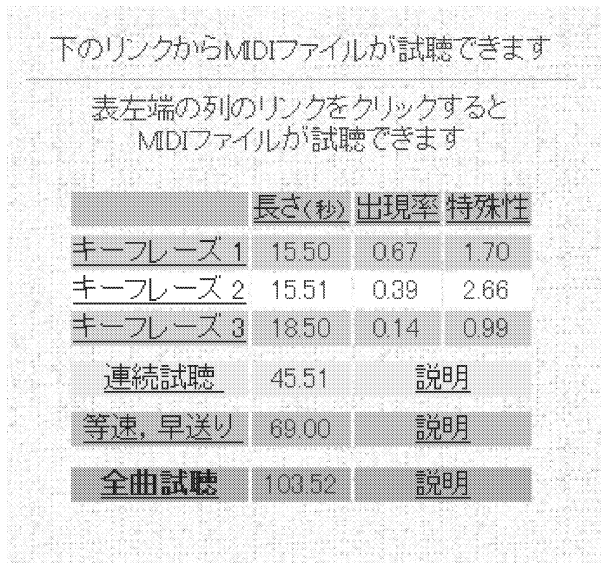


図 11 マルチレベル要約のインタフェース

図 11 ではリンクを列挙するにとどまっているが、今後の課題としては、図 12 のように楽曲内の構造などを図式化し、現在どの部分を聴いているかがわかりやすくなるようなインタフェースを作ることが挙げられる。

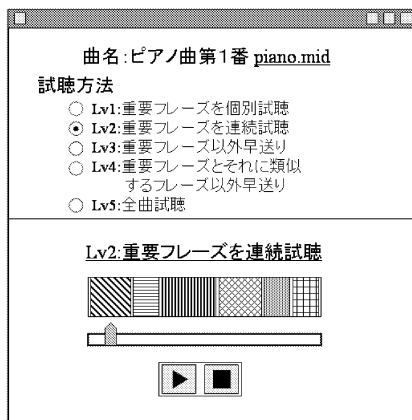


図 12 楽曲構造を図式化して提示するインタフェース

以上のような詳細度別のマルチレベル要約により、利用者の要求に合わせた効率的な楽曲評価が可能であると期待できる。

## 5. おわりに

本稿では楽曲の重要フレーズを列挙する要約法に対して、同時にフレーズの特徴を表す情報を提示することで聴取するフレーズの選択を支援する方法を提案した。また、楽曲の試聴方法を詳細度によりレベル分けし、利用者の要求に応じて試聴する詳細度を選択できるシステムについてのアイデアを提示した。この方法によって楽曲を詳細度別に段階的に評価でき、また一方では利用者の時間的余裕に合わせた詳細度の選択が可能

となる。

今後の研究ではマルチレベル要約の具体的な方法について検討し、これを要約システムの出カインタフェースとして実装することが課題である。また、要約の詳細度レベルとして、曲の1コーラス（歌謡曲でいう1番、2番）を単位とした提示法もひとつの案として考えられる。抽出した重要フレーズについての情報を構造化し、XML等で保存しておくことで、これらの情報を再利用できるようにすることも重要な課題であると考えられる。

## 文 献

- [1] 池添 剛, 梶川 嘉延, 野村 康雄: "音楽感性空間を用いた感性語による音楽データベース検索システム", 情報処理学会論文誌 Vol.42 No.12, pp.3201-3212, 2001.
- [2] 佐藤 聡, 菊池 幸平, 北上 始: "音楽データを対象としたイメージ検索のための感情価の自動生成", 情報処理学会研究報告「データベース」 Vol.118, pp.57-64, 1999.
- [3] 佐藤 聡, 小川 潤, 堀野 義博, 北上 始: "感情に基づく音楽作品検索システムの実現に向けての検討", 日本音響学会 音楽音響研究会資料 Vol.19, No.7, pp.51-56, 2001.
- [4] 平田 圭二, 松田 周: "パピブーン:GTTM に基づく音楽要約システム", 情報処理学会研究報告 MUS46-5, 2002.
- [5] 堀野 義博, 黒木 進, 北上 始: "クラシック曲のダイジェスト生成法の提案", 第1回情報科学技術フォーラム (FIT2002), pp.243-244, 2002.
- [6] 堀野 義博, 黒木 進, 北上 始: "フレーズの出現頻度に基づいたクラシック曲のダイジェスト生成法", データベースと Web 情報システムに関するシンポジウム (DBWeb2002), 情報処理学会データベースシステム研究会, pp.137-144, 2002.
- [7] 西尾 章治郎, 田中 克己, 上原 邦昭, 有木 康雄, 加藤 俊一, 河野 浩之: "情報の構造化と検索", 岩波書店, 2000.