

## 推薦論文

## SmartMusicKIOSK：サビ出し機能付き音楽試聴機

後藤 真孝<sup>†,††</sup>

本論文では、試聴のための新たな音楽再生インタフェース SmartMusicKIOSK を提案する。CD 販売店の店頭で音楽を短時間試聴する際には、通常の音楽鑑賞における受動的な聴き方と異なり、試聴者は早送りを何度も繰り返しながらサビを探すことが多い。しかし、こうした聴き方に対する支援は従来なかった。本研究では、サビの区間や楽曲中で繰り返される区間の先頭へジャンプする機能と、それらの区間の楽曲中での配置を視覚化する機能を提供する。これにより、試聴者が手探りでサビを見つける煩わしい作業を不要にし、試聴者が能動的に聴きたい場所を探す作業を容易にする。このような、インタラクティブに楽曲中の再生位置を変更しながら所望の箇所を見つけられるインタフェースは、試聴に限らず、音楽を選んで利用する一般的な目的で有用である。上記の機能を実現するために自動サビ区間検出手法を提案し、試聴機として実装・運用した結果、検出手法と試聴機の両者の有効性を確認した。従来の音楽再生インタフェースでは、楽曲単位でしか興味のない音楽を飛ばせなかったのに対し、SmartMusicKIOSK によって初めて、楽曲内部の興味のない箇所も容易に飛ばすことが可能になったといえる。

SmartMusicKIOSK:  
Music Listening Station with Chorus-search FunctionMASATAKA GOTO<sup>†,††</sup>

This paper describes *SmartMusicKIOSK*, a new music-playback interface for trial listening. In stores that sell music compact discs, short periods of trial listening of the CD music usually do not represent a passive appreciation of music — customers often search out the chorus or “hook” of a song by repeatedly pressing the fast-forward button. Listening of this type, however, has not been traditionally supported. Through our research, we have developed a function for jumping to the chorus section and other key parts of a song, plus a function for visualizing the song structure. These functions eliminate the hassle of searching for the chorus and make it easier for a listener to find desired parts of a song, thereby facilitating an active listening experience. This interface, which enables a listener to look for a section of interest by interactively changing the playback position, is useful not only for trial listening but also for more general purposes in selecting and using music. The proposed functions are achieved through an automatic chorus-section detection method, and the results of implementing these functions in a listening station have demonstrated their usefulness. While entire songs of no interest to the listener can be skipped on conventional music-playback interfaces, SmartMusicKIOSK is the first interface that allows the user to easily skip sections of no interest even within a song.

## 1. はじめに

店頭でコンパクトディスク (CD) 等に記録された音楽を「試聴」するときには、通常の楽曲全体の受動

的な鑑賞とは異なり、聴き手は、楽曲の再生に能動的に介入して聴きたい部分だけを選びだすという、音楽との新たなインタラクションを行っている。近年、CD 販売店の店頭には、購入の判断等を目的として CD を試聴できる音楽試聴機が設置されることが多い。通常、音楽を聴くときには、その鑑賞が主な目的であるため、各楽曲を先頭から最後まで通して再生する。ところが

<sup>†</sup> 科学技術振興事業団さきかけ研究 21「情報と知」領域  
“Information and Human Activity,” PRESTO, Japan  
Science and Technology Corporation (JST)

<sup>††</sup> 独立行政法人産業技術総合研究所  
National Institute of Advanced Industrial Science and  
Technology (AIST)

本論文の内容は 2003 年 2 月のインタラクション 2003 にて報告され、インタラクション 2003 プログラム委員長により情報処理学会論文誌への掲載が推薦された論文である。

試聴では、自分の探していた楽曲、好みの楽曲であるかどうかを短時間で判断することが目的であり、時間的な制約からこうした聴き方をすることは少ない。たとえばポピュラー音楽の場合、楽曲中で一番代表的な盛り上がる主題の部分であるサビ (chorus, refrain) を試聴して判断したいと考えることが多い。そこで試聴者は、イントロを少し聴いた後に、サビを探しながら早送りボタンを何度も押して途中を飛ばし、サビを再生するというような特殊な聴き方をする。

しかし、従来の音楽 CD の試聴機には、このような試聴固有の聴き方を支援する機能はなかった。試聴機は通常の CD プレーヤー相当の再生操作ボタンを持つが、その中で、早送りと早戻しのボタンしか、サビの部分を探すために利用できなかった。一方、最近 CD 販売店に導入され始めたデジタル試聴機では、MP3 等の圧縮形式で蓄積されている数十万曲の中から、ハードディスクあるいはネットワーク経由で再生することができる。しかし、楽曲先頭の短い区間 (通常 45 秒) だけが機械的に切り出されて収録されているため、試聴者はサビの部分必ずしも聴くことはできなかった。近年、日本のポピュラー音楽ではサビから始まる楽曲構造を持つ曲が増えているとはいえ、我々の調査では、日本のポピュラー音楽のヒットチャート (2001 年 1 月 ~ 12 月の週間ランキングのシングル上位 20 曲) の楽曲中、楽曲開始後 40 秒以内にサビが始まる曲は約 20% しかなかった。

そこで本研究では、「サビ出し」機能を搭載した音楽試聴機 SmartMusicKIOSK を提案する。試聴者はこの機能のボタンを押すだけで、サビの先頭へジャンプする (瞬時に早送りする) ことができ、自分でサビを探す煩わしい作業から解放される。さらに、サビ以外の楽曲中の繰返し構造も事前に推定あるいは用意しておくことで、次の楽曲構造の先頭へとジャンプする機能も提供する。たとえば、「イントロ ⇒ (A メロ ⇒ B メロ ⇒ サビ) × 2 ⇒ サビ」のような楽曲構造の場合、この機能を使うと、複数の A メロやサビの区間の先頭を自由に行き来することができるようになる。

従来、音楽情報処理の研究分野では楽曲の検索<sup>1),2)</sup> や音楽理解<sup>3)~9)</sup> に関する研究は多かったものの、音楽の試聴に着目した研究はなされていなかった。人間と音楽とのインタラクションでは、主に、発信・能動側 (作曲, 演奏等) と受信・受動側 (鑑賞, BGM 聴取等) の 2 つの形態があるが、試聴は、後者の鑑賞と異なり、

楽曲再生に能動的に介入しながら聴く行為であり、新しく興味深い研究対象であると我々は考えている。

本論文では以下、2 章で過去の音楽再生におけるインタラクションの形態について議論した後、3 章で音楽試聴機 SmartMusicKIOSK の全体構成を述べる。次に 4 章で、「サビ出し」機能を実現するための、音楽音響信号に対する自動サビ区間検出手法を提案し、具体的な実現方法を説明する。そして、5 章では、サビ区間検出手法の評価結果と、音楽試聴機の実装と運用結果を述べる。最後に、6 章で関連研究や応用に関して議論し、7 章でまとめを述べる。

## 2. 従来の音楽再生におけるインタラクション

楽曲の再生位置を容易に変更し、インタラクティブに音楽の再生に介入できるようになったのは、音楽の歴史の中で比較的最近のことである。音楽音響信号の記録が可能となる以前は、聴き手は生演奏された楽曲をその場で聴くことしかできなかった。その後、レコードやテープへの記録が可能になると、楽曲単位での再生位置の変更等が可能となったが、手間や時間がかかるために非リアルタイムな介入であった。これに対し、聴き手がインタラクティブに介入可能となったのは、CD 等の光磁気メディアへの記録が始まってからである。ボタンひと押しで再生位置のほぼ瞬時的な移動ができ、音楽を聴きながら楽曲単位でジャンプすることが容易になった。

しかし、楽曲間での移動は容易でも、試聴で求められるような、ある楽曲内で再生位置をインタラクティブに変更する支援は不十分であった。音楽試聴機や CD プレーヤー等の光磁気メディア再生機を持つ、典型的な再生操作ボタンは、再生、一時停止、停止、早送り、早戻し、次の曲へ頭出し、前の曲へ頭出しである (同一ボタンが兼用されることもある)。この中で、早送り、早戻しの 2 つにより、楽曲内での再生位置を変更できる。しかし、望みどおりの位置を見つける手掛かりとして、聴き手は主に以下の 3 つのフィードバックしか得られなかった。

- (1) ボタンを押している最中に早回しのように聞こえる音
- (2) ボタンを放した後の音
- (3) 楽曲の先頭からの経過時間の表示

そのため、たとえばサビを聴きたいと思っても、何度もボタンを押したり放したりして、手探りで見つけなければならなかった。

これらは、ハードディスク等に蓄積されている楽曲を計算機上のメディアプレーヤーで聴く場合にも基本的

これには、ポピュラー音楽のヒットチャートが音楽番組で紹介されたり、CM 放送で音楽が使われたりする際に、主にサビの部分が再生されることも影響している。

に同じであるが、再生位置スライダが用意されることがある。スライダの全長は楽曲の長さに対応しており、聴き手は、楽曲全体の中で先頭から何割ぐらいの位置を再生しているかを把握し、スライダ操作により任意の位置へジャンプできる。しかし、的確な再生位置を手探りで見つけることに変わりはない。

### 3. かしこい音楽試聴機 : SmartMusicKIOSK

上記の問題を解決するために、インタラクティブな再生位置指定を支援する音楽試聴機 SmartMusicKIOSK を実現した。ここで解くべき問題は、本来は時間をかけて聴かなければ把握できない音楽に対し、それを聴く前にどのようにして的確な再生位置の変更を可能にするかである。そこで、ポピュラー音楽を主な対象と想定して、以下の2つの解決法を提案する。

- (1) 楽曲構造上意味を持つ区間の先頭へ自動ジャンプ(サビ出しボタン)  
事前に楽曲構造を解析しておき、その中で聴き手が関心を持ちそうな部分へ自動的にジャンプできる機能を提供する。具体的には、「サビの区間の頭出し」、「前の楽曲構造の区間の頭出し」、「次の楽曲構造の区間の頭出し」の機能を用意しておき、聴き手がサビの部分だけ聴いたり、前後の楽曲構造の先頭へとジャンプして聴いたりすることを可能とする。
- (2) 楽曲の内容を反映した視覚化(音楽地図)  
聴き手がどこへジャンプすればよいかを自分で判断できる手掛かりとして、楽曲の内容を視覚化する機能を提供する。具体的には、図1のように、サビ区間や楽曲中の繰り返し区間の構造を視覚化する。この視覚化された画面から、イントロ、Aメロ、Bメロ、サビ、間奏等の相互の位置関係が把握できることが多い。聴き手は、これを見ながら前述の自動ジャンプボタンで操作してもよいし、通常の早送りボタンや再生位置スライダを使って、望みの箇所へ移動してもよい。

SmartMusicKIOSK のインタフェースの画面表示を図1に示す。比較のために、新たな提案機能を削って、従来の典型的な7種類の再生操作ボタンだけを残したインタフェース画面も図2に示す。図2のウィンドウ中、左側から順に、停止、一時停止、再生、前の曲へ頭出し、早戻し、早送り、次の曲へ頭出しのボタンが並んでおり、慣例的な記号が描かれている。停止ボタンの上側には、楽曲の先頭からの経過時間が表示

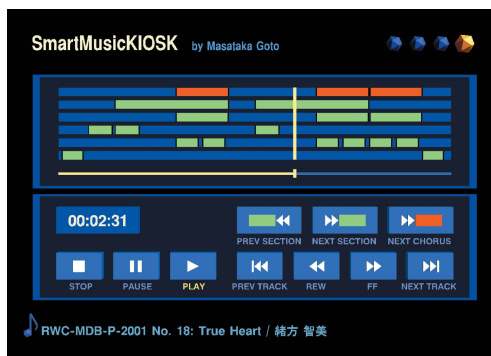


図1 SmartMusicKIOSKの画面表示:下側のウィンドウが再生操作ボタン群,上側のウィンドウが楽曲内容の視覚化結果(RWC研究用音楽データベース<sup>10</sup>) RWC-MDB-P-2001 No.18に対する自動サビ区間検出結果)を表す。上側のウィンドウの横軸は時間軸で楽曲全体を表示しており、最上段がサビ区間の一覧,その下の5段が繰り返し構造,その下の横棒が再生位置スライダを表す

Fig.1 SmartMusicKIOSK screen display. The lower window presents the playback-operation buttons and the upper window provides a visual representation of a song's contents (results of automatic chorus-section detection using RWC Music Database<sup>11</sup>) RWC-MDB-P-2001 No.18). The horizontal axis of the upper window is the time axis covering the entire song; the top row shows chorus sections, the five lower rows show other repeated sections, and the bottom horizontal bar is a playback slider.



図2 従来の典型的なメディアプレーヤ相当のインタフェースの画面表示:下側のウィンドウが再生操作ボタン群,上側の横棒が再生位置スライダを表す

Fig.2 Screen interface as used by a conventional media player. The lower window displays the playback-operation buttons and the upper horizontal bar is a playback slider.

されている。以下、図1の上下2つのウィンドウを説明する。

- 再生操作ウィンドウ(下側のウィンドウ)  
図2と比較して追加された3つの自動ジャンプボタンは、左から順に
  - 「前の楽曲構造の区間の頭出し」ボタン
  - 「次の楽曲構造の区間の頭出し」ボタン
  - 「サビの区間の頭出し」ボタン
 である。ボタン上の記号は新たにデザインした。

「サビの区間の頭出し」ボタンを押すと、現在の再生位置より後方の（後方になれば最初の）サビ区間を探索し、その開始点にジャンプする。通常、サビは楽曲中で複数回繰り返されるが、このボタンを押すたびにそれらの間を順にジャンプできる。他の2つのボタンを押すと、現在の再生位置の直後もしくは直前に位置する区間の開始点を探索し、その先頭にジャンプする。探索時には、区間の終了点は無視する。

- 楽曲構造表示ウィンドウ（上側のウィンドウ）最上段にサビ区間、その下（最大で5段）に繰り返し区間が視覚化されている。各段の中で、着色されている区間どうしが似ている（繰り返しである）ことを表している。最下段の細い横棒は、楽曲中の経過時間を知らせる再生位置スライダである。区間を直接クリック（タッチパネル使用時にはタッチ）して再生したり、再生位置スライダをクリックして位置変更したりすることが可能である。

以上のインタフェースと機能により、試聴者は、イントロを少し聴いた後に、ボタンひと押しでサビへジャンプして聴くような試聴が可能となる。また、楽曲全体の構造を視覚的に把握しながら、様々な箇所を選択的に試聴できるようになる。

#### 4. SmartMusicKIOSK の実現方法

SmartMusicKIOSK の自動ジャンプや視覚化の機能を実現するうえで、各楽曲のサビ区間と繰り返し構造の記述が必要である。そして、多数の楽曲に対して実用的に運用するには、これらを自動的に得ることが重要となる。しかし、実世界の複雑な音楽音響信号からこれらを得ることは容易ではなく、既存の手法では実現できない。

そこで、ポピュラー音楽を対象に、楽曲中のサビ区間や繰り返し区間の開始点と終了点の一覧を自動検出する手法 RefrainD (Refrain Detecting Method) を提案する。従来、楽曲の音響信号中に何度も出現するサビの中のどこか1カ所を、つねに指定した長さだけ切り出して提示する手法<sup>12)~14)</sup>はあったが、サビ区間の開始点と終了点は分からず、サビの転調も扱えなかった。

なお、再生操作ウィンドウに「前のサビの区間の頭出し」ボタンと「次のサビの区間の頭出し」ボタンの2つを用意してもよい。ここでは、以下の理由から1つのボタンとしている。(1)現状の「サビの区間の頭出し」ボタンを連打すれば、すべてのサビ区間を一巡した後にまた最初の区間に戻るため、短時間で望みの箇所へ移動できる。(2)瞬時に過去のサビ区間へ戻る必要がある場合には、楽曲構造表示ウィンドウで区間を直接クリックする手段が用意されている。

本手法は、様々な繰り返し区間の相互関係を調べることで、楽曲中で繰り返されるすべてのサビ区間を網羅的に検出し、それらの開始点と終了点を推定できる。また、転調後でも繰り返しと判断できる類似度を導入することで、転調をとまなうサビ区間も検出できる。

自動検出では検出結果に誤りが含まれることもあるが、実用上は、完全な精度でなくても聴き手が再生位置を見つかる手掛かりとなるため、従来の試聴機よりは便利となる。ただし、正確な記述が必要などときには、検出結果を手作業で修正できるとよい。そこで、サビ区間や楽曲構造を手動でラベリング・修正できるエディタも用意した。この手動ラベリングは、自動検出が困難な楽曲や他ジャンルの音楽に対しても有効である。

##### 4.1 サビ区間検出の実現上の課題

本研究ではポピュラー音楽の多様な楽曲を扱うため、サビ固有の音響的な特徴に関する事前情報をいっさい使わない、ロバストなサビ区間検出を実現したい。そこで、通常はサビ区間が楽曲中で最も多く繰り返されることに着目し、基本的には、ある区間の繰り返しを見つけ出し、最も出現頻度の高い区間を出力する戦略をとる。しかし、「繰り返し」といっても完全に一致する状態で繰り返されることはまれで、計算機にとっては判断が難しい。その際の主要な課題は、以下のようまとめられる。

##### 課題1：特徴量と類似度の検討

ある区間が別の区間の繰り返しであるかどうかを、各区間から求めた特徴量間の類似度に基づいて判断しなければならぬ。その際、繰り返すたびに伴奏のアレンジやメロディーライン等が多少変化しても、特徴量間の類似度は高い必要がある。単に、音響信号処理でよく用いられるパワースペクトルやMFCCを特徴量とすると、こうした要件を満たすのは困難となる。

##### 課題2：繰り返しの判断基準

類似度がどの程度高ければ繰り返しと見なせるかという基準は、楽曲に依存して変わる。この基準を、少数楽曲に特化して人間が手で設定するのは容易だが、幅広い楽曲に適用可能な手法とするためには、その基準を処理中の楽曲に基づいて自動的に変える必要がある。

課題3：繰り返し区間の端点（開始点と終了点）の推定  
様々な繰り返し区間の相互関係を調べることで、繰り返し区間の端点を推定する必要がある。たとえば、(A B C B C C)の構造を持つ楽曲に対し、単純に長い繰り返しを見つげると(B C)が求まってしまう。この場合、最後のCの繰り返し情報に基づい

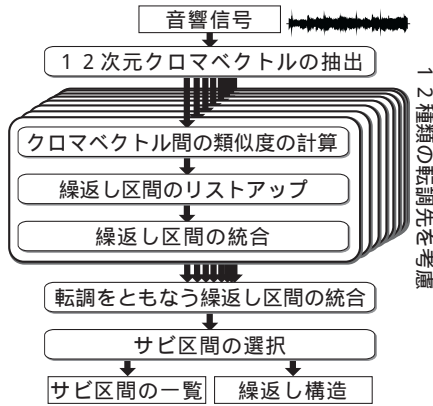


図3 サビ区間検出手法 RefraiD の処理の流れ

Fig. 3 Overview of chorus-section detecting method RefraiD.

て、(B C) の C の区間の端点を推定する、といった処理が求められる。

課題 4 : 転調をともなう繰返しの検出

転調後の区間は、一般に特徴量が大きく変わるために転調前の区間との類似度が低くなり、繰返しの判断が困難となる。特に、転調は曲の後半のサビの繰返しで起きることが多く、重要な課題である。

4.2 サビ区間検出手法 RefraiD の処理の概要

上記の課題を解決するサビ区間検出手法 RefraiD の処理の流れを図 3 に示す。なお、本手法が対象とするポピュラー音楽では、サビの区間はほぼ同じテンポで繰り返し演奏されることが多いため、本手法でも繰り返すたびに同じ箇所は同じテンポで演奏されることを仮定する(テンポは必ずしも一定でなくてよい)。

(1) 12次元クロマベクトルの抽出と類似度の計算  
 特徴量として、入力音響信号のパワースペクトル  $\Psi_p(f, t)$  (時刻  $t$ , 対数スケール周波数  $f$ , STFT 窓幅 256 ms, フレームシフト 80 ms) から、細部の変形の影響を受けにくい 12 次元クロマベクトル(chroma vector)  $\vec{v}(t)$  を求める。この計算の模式図を図 4 に示す。 $\vec{v}(t)$  の各次元  $v_c(t)$  は、12 音名の各音名  $c$  ( $c = 1, 2, \dots, 12$ ) の周波数のパワーを複数のオクターブ  $h$  にわたって加算したもので、

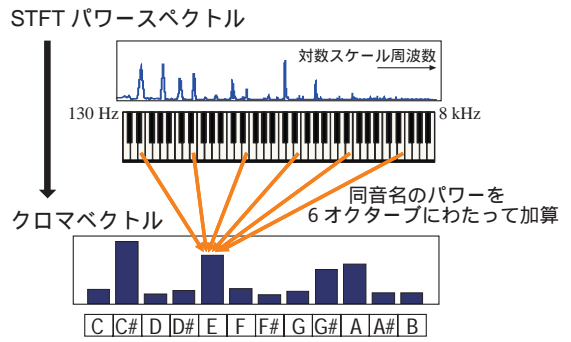


図 4 12次元クロマベクトルの抽出

Fig. 4 Overview of calculating 12-dimensional chroma vector.

$$v_c(t) = \sum_{h=Oct_L}^{Oct_H} \int_{-\infty}^{\infty} BPF_{c,h}(f) \Psi_p(f, t) df \quad (1)$$

と定義する。 $BPF_{c,h}(f)$  は、音名  $c$ , オクターブ  $h$  の位置のパワーを通過させるバンドパスフィルタで、 $Oct_L$  と  $Oct_H$  は、130 ~ 8,000 Hz の 6 オクターブにわたるように設定する。そして、時刻  $t$  の  $\vec{v}(t)$  とそれよりラグ (lag)  $l$  ( $0 \leq l \leq t$ ) だけ過去の  $\vec{v}(t-l)$  との類似度  $r(t, l)$  を、

$$r(t, l) = 1 - \frac{\left| \frac{\vec{v}(t)}{\max_c v_c(t)} - \frac{\vec{v}(t-l)}{\max_c v_c(t-l)} \right|}{\sqrt{12}} \quad (2)$$

により計算する(以上が課題 1 に対応)。

(2) 繰返し区間のリストアップ

判別基準に基づく自動閾値選定法<sup>16)</sup>によって、繰返しの判断基準(類似度に対する閾値)を楽曲ごとに自動的に変えながら、繰返し区間のペアをリストアップする(課題 2 に対応)。 $r(t, l)$  を、図 5 のように  $t-l$  平面に描画すると、繰返し区間に対応して、時間軸に平行な線分(類似度が連続して高い領域)が現れる。これを類似線分と呼び、ラグ  $L1$  の位置で、時刻  $T1 \sim T2$  の間に類似線分があるとき、区間  $T1 \sim T2$  と  $(T1 - L1) \sim (T2 - L1)$  が繰返しであることを意味する。よって、 $r(t, l)$  中の類似線分をすべて検出すれば、繰返し区間の一覧が得られる(実際には、時間軸に平行な成分を強調するフィルタを  $r(t, l)$  にかけてから、以下の処理を行う)。そこで、

$$R_{all}(t, l) = \int_l^t \frac{r(\tau, l)}{t-l} d\tau \quad (3)$$

の値が十分高い  $l$  の位置  $L1$  に類似線分が存在すると考え、類似度  $r(\tau, L1)$  が時間軸  $\tau$  方向に連続して十分高い領域を探索して類似線分とする(図 6)。この十分高いことを判断する閾値を、クラス分離度(ク

文献 15) によれば、音楽的な音高の知覚は上に昇る螺旋状の構造を持ち、螺旋を真上から見た円周上のクロマ(音名, chroma)と、横から見たときの縦方向のハイト(オクターブ位置, height)の 2 つの次元で表現することができる。クロマベクトルは、このクロマを周波数軸としてパワーの分布を表現した特徴量といえる。

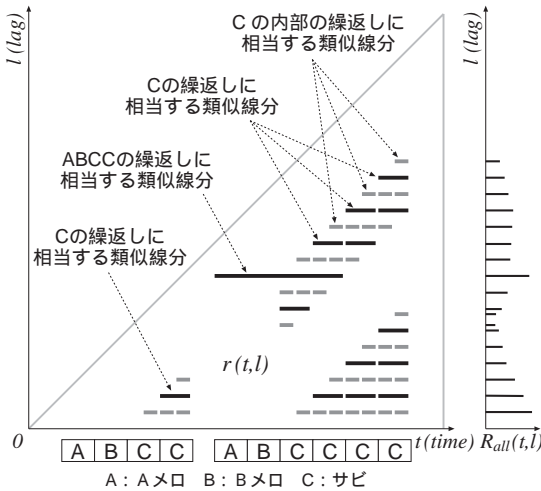


図5 ある楽曲に対する類似線分, 類似度  $r(t, l)$ , パラメータ空間  $R_{att}(t, l)$  の概念図:  $r(t, l)$  は, 右下半分の直角形内で定義される. 実際に得られる  $r(t, l)$  はノイズを多く含み, サビに関連しない類似線分も存在して曖昧なことが多い

Fig. 5 A plot of line segments, the similarity  $r(t, l)$ , and the possibility  $R_{att}(t, l)$  of containing line segments. The similarity  $r(t, l)$  is defined in the right-angled isosceles triangle in the lower right-hand corner. The actual  $r(t, l)$  is noisy and ambiguous and usually contains many line segments irrelevant to chorus sections.

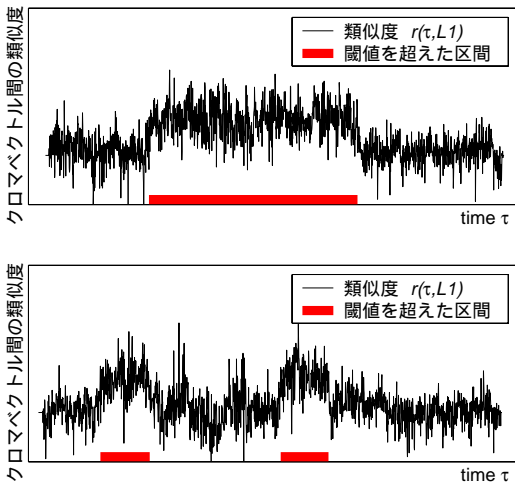


図6 ラグ  $L1$  の位置における類似度  $r(\tau, L1)$  の具体例: 一番下の横棒が, 自動調整された閾値を超えた区間(時間軸  $\tau$  方向に連続して十分高い領域)を表し, 類似線分に対応する

Fig. 6 Examples of the similarity  $r(\tau, L1)$  at lags  $L1$ . The bottom horizontal bars indicate the regions above an automatically adjusted threshold, which means they correspond to line segments.

ラス間分散)を最大とする判別基準に基づく自動閾値選定法<sup>16)</sup>により定める. 実際の閾値処理は, 類似度  $r(\tau, L1)$  を平滑化してから行う.

(3) 繰返し区間の統合

各類似線分は, 単に2つの区間が繰り返されていることを表すため, 共通区間を持つ類似線分どうしを1つのグループとして統合する. その際に, ボトムアップな検出でもれていた類似線分の再検出を, 他の類似線分の情報に基づいて行う. 具体的には, 各グループにおいて類似線分の両端を積分区間として式(3)に相当する積分をし, そのピークから類似線分を再検出する. たとえば, 図5のABCCの繰返しに相当する長い類似線分上で, Cの繰返しに相当する類似線分2カ所が得られていなくても, ここで検出されることが期待できる. こうして, 各グループごとに端点を適切に求め直す(課題3に対応). さらにここでは, 得られた類似線分が適切であるかも検証し, 不適切と判断されたものは削除する. たとえば, その上での  $r(t, l)$  の変動が大きすぎる類似線分や, ある区間が連続して繰り返される回数が多すぎるグループの類似線分は, 不適切と判断される.

(4) 転調をともなう繰返し区間の統合

ある演奏のクロマベクトルを  $\vec{v}(t)$  とし, それを半音  $\zeta$  個分上へ転調した演奏のクロマベクトルを  $\vec{v}(t)'$  とすると, それらの各次元は音名に対応しているため, 転調幅  $\zeta$  に応じて  $\vec{v}(t)'$  の次元間で値をシフトさせたものと, 転調前の  $\vec{v}(t)$  とは値が近くなる ( $\vec{v}(t) \approx S^\zeta \vec{v}(t)'$ ). シフトの操作は, 以下のシフト行列  $S$  で表現できる.

$$S = \begin{pmatrix} 0 & 1 & 0 & \dots & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 0 & 1 & 0 \\ 0 & \dots & \dots & \dots & 0 & 1 \\ 1 & 0 & \dots & \dots & \dots & 0 \end{pmatrix} \quad (4)$$

クロマベクトルのこの特長を利用し, 12種類の転調先を考慮して,  $\zeta$  ごとの12種類の類似度  $r_\zeta(t, l)$  を

$$r_\zeta(t, l) = 1 - \frac{\left| \frac{S^\zeta \vec{v}(t)}{\max_c v_c(t)} - \frac{\vec{v}(t-l)}{\max_c v_c(t-l)} \right|}{\sqrt{12}} \quad (5)$$

と定義し直す. これを出発点として, 上記の繰返し区間の検出処理も12種類分を行い, それらすべての繰返し区間を統合する(課題4に対応).

(5) サビ区間の選択

繰返し区間のグループ  $i$  ごとにサビらしさ  $\nu_i$  を評価し, 最も高いグループ  $m = \operatorname{argmax}_i \nu_i$  の区間をサ

$\zeta$  は  $0, 1, \dots, 11$  の12種類の値をとるものとする.  $\zeta = 0$  は転調しないことを意味し,  $\zeta = 10$  は半音10個分上か, 全音分下へ転調することを意味する.

ビ区間とする。サビらしさ  $\nu_i$  は、グループ  $i$  内の  $M_i$  個の各区間  $j$  ごとに求めた信頼度 (局所的なサビらしさ)  $\lambda_{ij}$  の和に、区間の長さ  $L_i$  に応じた重みをかけて

$$\nu_i = \left( \sum_{j=1}^{M_i} \lambda_{ij} \right) \log \frac{L_i}{D_{\text{len}}} \quad (6)$$

と定義する (定数  $D_{\text{len}} = 1.4 \text{ sec}$ )。  $\lambda_{ij}$  は、対応する類似線分における類似度  $r_c(t, l)$  の平均とし、さらに、以下の3つの仮定を満たすものが高くなるように修正する。

- (仮定1) サビには、許容される適切な長さの範囲 (現在の実装では 7.7 ~ 40 sec) がある。
- (仮定2) 「(A メロ  $\Rightarrow$  B メロ  $\Rightarrow$  サビ)  $\times 2$ 」に相当するような長い区間の繰返しがある場合、その末尾部分がサビである可能性が高い。
- (仮定3) ある繰返し区間に、その区間の半分程度の短い区間が繰返されている場合には、元の区間がサビである可能性が高い。

RefraiD は、こうして求めたサビ区間の一覧とともに、中間結果として得られた繰返し構造も出力する。

## 5. システムの実装と結果

3章で述べた SmartMusicKIOSK の機能をすべて実装したシステムを構築した。本システムのインタフェース部分は、サビ区間検出手法 RefraiD が出力するサビ区間と繰返し構造の記述を含むファイルを利用する形で動作する。自動検出結果としては様々な繰返し構造が求まるが、インタフェース部分でのジャンプや視覚化には、区間が長いものから上位5つだけを使用した。

本システムの GUI 部分、楽曲ファイル再生エンジン部分、音響出力デバイス制御部分は、拡張性が高くなるように、分散環境で動作する別々のプロセスとして実装した。そのために、音響信号や各種制御情報をネットワーク上で効率良く共有することを可能にするネットワークプロトコル RACP (Remote Audio Control Protocol) を設計し、それに基づいて実装した。RACP は、RMCP<sup>17)</sup>を音響信号の伝送用に拡張したプロトコルである。

本システムは移植性高く設計されており、すでに各種オペレーティングシステム (Linux 2.4, SGI IRIX 6.5, Microsoft Windows XP) 上に実装されている。図7に、ペンタブレット搭載ノートパソコン (Microsoft Windows XP Tablet PC Edition, Pentium III 933 MHz CPU) 上でスタンドアロンで動作



図7 タブレット PC 上に実装した SmartMusicKIOSK の写真  
Fig.7 Demonstration of SmartMusicKIOSK implemented on a tablet PC.

させた例を示す。写真中央のタブレット画面上をペンでタッチするだけでなく、右側の再生操作ボタン群が描かれた外付けのキーボードを押して操作することもできる。

以下、RefraiD の評価結果を述べた後に、実装した SmartMusicKIOSK のシステムの運用結果を述べる。

### 5.1 自動サビ区間検出手法の評価結果

RefraiD の有効性を確認する実験を行った。ここで評価対象はサビ区間の検出精度のみで、繰返し構造は評価対象外とした。評価には、「RWC 研究用音楽データベース: ポピュラー音楽」<sup>10)</sup>の100曲 (RWC-MDB-P-2001 No.1 ~ 100) を用いた。検出結果の正誤を判定するためには、基準となる正解のサビ区間を人間が手作業で指定し、その人間の判断と自動検出結果とを比較する必要がある。そこで、楽曲を分割してサビ区間等をラベリングできる、楽曲構造ラベリング用エディタを開発した。なお、このエディタにより、4章の冒頭で述べた手動ラベリングも可能となる。

こうして作成した正解に基づき、各曲に対する検出結果の区間と正解のサビ区間がどれくらい重なっているかを、再現率  $R$ 、適合率  $P$ 、および両者を統合した  $F$  値<sup>18)</sup>の観点から評価した。以下に定義を示す。

$$R = \frac{\text{正しく検出したサビ区間の長さの合計}}{\text{正解のサビ区間の長さの合計}}$$

$$P = \frac{\text{正しく検出したサビ区間の長さの合計}}{\text{検出した区間の長さの合計}}$$

$$F \text{ 値} = \frac{(\beta^2 + 1)PR}{\beta^2 P + R} \quad (\beta = 1 \text{ を使用})$$

ただし、転調をともなう場合には、相対的な調の移動幅が正解と一致したときだけ、正しく検出したと判断した。そして、 $F$  値が 0.75 以上のとき、その曲のサビ区間を正しく得られた (正答した) と判定した。

表 1 RefraiD の評価結果：4 つの条件下における 100 曲中のサビ正答曲数

Table 1 Results of evaluating RefraiD: the number of songs whose chorus sections were detected correctly under 4 sets of conditions.

|             | 条件 (使用: ○, 未使用: ×) |      |      |      |
|-------------|--------------------|------|------|------|
| 転調区間の検出     | ○                  | ×    | ○    | ×    |
| 仮定 2, 3 の使用 | ○                  | ○    | ×    | ×    |
| 正答曲数        | 80 曲               | 74 曲 | 72 曲 | 68 曲 |

評価結果として、100 曲中の正答曲数を表 1 に示す。通常の RefraiD の性能は一番左の 80 曲 (80 曲の平均 F 値は 0.938) である。誤検出は、サビの繰返しが他の箇所繰返しより多くなかったり、曲中ほとんどが類似伴奏の繰返しだったりしたのが主な原因だった。100 曲中には、サビに転調のある曲が 10 曲含まれているが、そのうち 9 曲は検出できていた。4.2 節 (4) の転調をともなう繰返しの検出をやめた場合、左から 2 番目のように性能が落ちた。一方、4.2 節 (5) で述べた仮定 2, 3 に基づく信頼度の修正をやめた場合は、右 2 つのように性能が落ちた。サビの繰返しで、伴奏やメロディーに大幅な変化をともなう曲は 22 曲あったが、そのうち 21 曲は検出できており、その中で、変化をともなうサビ自体は 16 曲で検出できていた。以上から、本手法が実世界の音響信号に対して有効であることが確認された。

## 5.2 SmartMusicKIOSK の運用結果

実装した SmartMusicKIOSK を、2 つの提案機能 (ジャンプボタンと楽曲構造表示) の有無に応じて、4 つの条件で運用した。試聴対象には、RWC 研究用音楽データベース (RWC-MDB-P-2001) の中から、試聴者が初めて聴く楽曲を選んだ。また、自動サビ区間検出手法によって正しく得られた記述を用いた。ここでは、条件間の比較のために、視覚化された楽曲構造上の区間を直接クリックして再生する機能は用いていない。

以下、4 つの条件とその運用結果を述べる。

条件 1: 提案機能がいない (図 2 相当)。

試聴者は、曲の頭の部分がサビでない場合、少し聴いては早送りボタンを押すという動作を、サビが出てくるまで 5~10 回程度繰り返した。少しずつ聴きながら早送りすることは、時間がかかって煩わしいものの、楽曲の雰囲気をつかみたいときには有効であった。

条件 2: ジャンプボタンがないが、楽曲構造表示はある。

どこまで再生位置を早送りすればよいかのフィー

ドバックが得られる点が効果的であり、試聴者は条件 1 よりも便利だと評価した。しかし、画面表示によって現在の再生位置よりも先が見えているだけに、そこまで飛ばしたいという欲求が強くなる傾向があった。

条件 3: ジャンプボタンはあるが、楽曲構造表示がない。

最初にイントロを聴いた後に、直接「サビの区間の頭出し」ボタンを押す聴き方と、「次の楽曲構造の区間の頭出し」ボタンを押しては少し聴くという操作を繰り返し、サビが出てきたらそこをじっくり聴くという聴き方が主だった。効率良く飛ばしながら聴ける点が評価され、条件 2 より好まれていた。

条件 4: 提案機能がすべて有効である (図 1 相当)。

条件 2, 3 の利点がともに得られる試聴方法であり、最も便利だと評価された。条件 3 のような聴き方に加え、楽曲構造上を自在に行き来しながら聴く傾向が強くなり、サビを聴いた後に A メロに戻って聴いたり、楽曲後半のサビの繰返しへ飛んで聴いたりしていた。

条件 3 は、通常のメディアプレーヤに 3 つのジャンプボタンを追加した場合に相当し、楽曲構造表示がなくても通常のプレーヤより便利であることが分かった。また条件 4 から、視覚化はさらにその操作を助け、楽曲の様々な箇所を聴くうえで有効であることが分かった。さらに、ボタンの機能や表示ウィンドウの中身に関していっさい説明を受けていない試聴者が、条件 4 で利用した場合でも、短時間の使用でこれらの意味を推測し、把握できた。

以上から、提案したインタフェースが機能し、試聴者は、楽曲構造表示の助けを得ながらジャンプボタンを押して、インタラクティブに楽曲再生に介入できることを確認した。また運用を通じて、提案したサビ出し機能や前後の区間の頭出し機能は、使用するのが容易で訓練は不要であり、直感的で使いやすいことが分かった。

## 6. 議 論

本研究は、音楽再生におけるインタラクションに取り組む従来にない研究であるが、音楽の視覚化や音楽要約に関しては様々な研究がなされてきた。以下では、それらを紹介するとともに、試聴という場面に限定しない音楽再生におけるインタラクションに関して考察する。また、提案手法がどのような場面に応用可能であるかも議論する。



## 6.1 関連研究

5.2 節で述べたように、楽曲構造表示のような視覚化は、試聴する際の再生位置変更の目安として有効である。音楽が持つ情報の視覚化に関しては、古くから、楽譜表示や MIDI データのピアノロール表示 等が用いられてきた。また、主にクラシック音楽の演奏の表情付けを分析する目的で、いくつかの視覚化手法が提案されている<sup>19)~23)</sup>。しかし、いずれも MIDI データに限定された手法であり、音楽音響信号に適用することはできなかった。また、本研究のようにサビ区間や楽曲中の繰返し区間の構造を表示するものはなく、ポピュラー音楽の楽曲全体を俯瞰して、聴きたい場所をインタラクティブに選ぶ目的には利用できなかった。

一方、音楽音響信号の視覚化として、音響信号の波形や周波数スペクトルを表示することは容易である。しかし、楽曲全体に対するこれらの表示を聴き手が見ても、音楽の構造を読み取れず、的確な再生位置の判断が困難な場合が多い。

本研究は、最近研究事例が報告され始めた音楽要約<sup>24)~26)</sup>とも関連がある。試聴の目的の1つに、音楽を短時間で聴くということがあるが、この目的は、楽曲の長さを短くする音楽要約と共通だからである。しかし音楽要約では、本研究のようなインタラクティブで能動的な聴取に関しては考慮されていなかった。試聴という観点からは、ユーザが自分の意志でインタラクティブに楽曲の聴きたい部分を選択できる意義は大きい。この能動的な聴取に関しては次節で引き続き議論する。

## 6.2 音楽を能動的に聴取できるインタフェース

近年音楽の利用場面は拡大しており、店頭での CD 試聴に限らず、携帯電話の着信音楽の選択や、ある場面に適切な BGM (バックグラウンドミュージック) の選択、WWW での音楽の利用等、自分の意志で音楽を選び、内容を確認、ときには部分的に切り出して利用する形態が増えている。それに対して、2 章で述べたように、音楽再生のためのインタフェースは、CD プレーヤや計算機上のメディアプレーヤの登場後も、典型的な再生操作ボタン群に固定されていた。これらは、受動的に鑑賞する目的では適切であったが、楽曲中で自分の望む部分をインタラクティブに見つける目的には不十分である。

一般的な音楽再生インタフェースとして SmartMusicKIOSK を考えると、従来の楽曲単位での操作体系

に対し、楽曲内部の区間単位での操作体系が追加されたものととらえることができる。従来は楽曲単位で興味の無い曲を飛ばせたが、楽曲内部の興味の無い箇所を飛ばすのは容易でなかった。本インタフェースにより、原曲の時系列に沿わずに、「好きなところを聴きたいように聴ける」ようになったメリットは大きい。この方向の延長線上には、楽曲単位での操作からの類推として、区間単位でのシャッフルプレイ機能の追加等が考えられる。

インタフェース構築時には予想しなかったことだが、試聴よりも長時間の利用が許される場面で、楽曲構造を見ながら音楽再生位置をインタラクティブに変更して聴くときには、従来の鑑賞に比べてより分析的に聴く傾向があった。たとえば、楽曲全体がどのような構造になっているかを確認、構成上の各区間を聴いたり、繰り返される区間どうしの比較をしたりする場面が見られた。また、楽曲構造の視覚化結果は、それを眺めながら音楽を聴くだけでも面白く有用であった。

試聴という観点からは、サビを聴きたいという要求の次に、楽曲の雰囲気把握したいという要求がある。SmartMusicKIOSK では、楽曲中の繰返し区間の先頭へジャンプすることで、ある程度短時間に雰囲気を把握することが可能であったが、繰り返されない区間(間奏・ギターソロ等)の先頭へはジャンプできない。こうした「どんな雰囲気の曲かが容易に分かるインタフェース」への発展は、今後の課題である。

## 6.3 応用

SmartMusicKIOSK の機能や自動サビ区間検出手法は、本論文での使用方法以外にも幅広い応用の可能性を持つ。以下では、主な応用例を紹介する。

### ● デジタル試聴機

1 章で紹介したデジタル試聴機では、機械的に切り出された楽曲の先頭しか収録されていなかった。自動サビ区間検出手法を用いることで、サビ区間を切り出して収録することが可能になる。将来的には、店頭のデジタル試聴機自身が、SmartMusicKIOSK のような機能を持つものに発展することが望まれる。

### ● 音楽サムネイル (music thumbnail)

多数の楽曲をブラウジングするときや、楽曲検索システムにおいて検索結果を提示するときに、サビの冒頭を短く再生 (プレビュー) できると便利である。これは、画像のサムネイルの音楽版と見なせ、自動サビ区間検出手法で切り出されたサビの冒頭を使用できる。

### ● 計算機上のメディアプレーヤ

横軸が時間、縦軸が MIDI ノートナンバ (通常は鍵盤表示) である二次元平面上で、発音中の部分に着色する表示方法。

近年のメディアプレーヤでは、プレーヤの外観(スキン)を変更する機能や、再生中に周波数帯域ごとのパワー等に基づいて幾何学図形のようなアニメーションを表示する機能が追加されているが、インタフェース自体に進歩は見られなかった。本研究の提案内容がメディアプレーヤに採用されるだけでなく、メディアプレーヤという音楽再生インタフェースの機能全体が、改めて再検討される流れが生まれることを期待したい。

## 7. おわりに

本論文では、インタラクティブな音楽試聴インタフェースについて検討し、聴き手が音楽再生に能動的に介入しやすい音楽試聴機 SmartMusicKIOSK を提案した。SmartMusicKIOSK により、試聴者は、サビ区間や楽曲中の繰返し区間の先頭へ自在にジャンプし、同時に楽曲全体を俯瞰してそれらの区間の配置を見ることが可能となった。また、上記区間を得るための自動サビ区間検出手法 RefraiD を提案し、RWC 研究用音楽データベース 100 曲を用いた評価実験により有効性を確認した。実際に、SmartMusicKIOSK を実装して運用した結果、試聴において便利だけでなく、今まで経験したことのないような音楽の聴き方ができる新たな音楽再生インタフェースとしても有用であることが分かった。

今後は、6 章で議論したような各種拡張を行っていく予定である。また、試聴以外も広く視野にのぞいた、人間と音楽とのインタラクションをより能動的で豊かにする研究の方向性も探求していきたい。

謝辞 本研究に対し有益な議論をしていただいた、麻生英樹氏(産業技術総合研究所)に感謝する。

## 参考文献

- 1) 園田智也, 後藤真孝, 村岡洋一: WWW 上での歌声による曲検索システム, 信学論(D-II), Vol.J82-D-II, No.4, pp.721-731 (1999).
- 2) 橋口博樹, 西村拓一, 張建新, 滝田順子, 岡隆一: モデル依存傾斜制限型の連続 DP を用いた鼻歌入力による楽曲信号のスポットニング検索, 信学論(D-II), Vol.J84-D-II, No.12, pp.2479-2488 (2001).
- 3) Desain, P. and Honing, H.: *Music, Mind and Machine: Studies in Computer Music, Music Cognition and Artificial Intelligence*, Thesis Publishers (1992).
- 4) Dannenberg, R.: Music Understanding by Computer, *IAKTA/LIST International Workshop on Knowledge Technology in the Arts Proc.*, pp.41-56 (1993).
- 5) 片寄晴弘: 自動採譜, コンピュータと音楽の世界, pp.74-88, 共立出版 (1998).
- 6) 柏野邦夫: 重なり合った音を聞き分ける—音源分離, コンピュータと音楽の世界, pp.89-99, 共立出版 (1998).
- 7) 後藤真孝: リアルタイム音楽情景記述システム: 全体構想と音高推定手法の拡張, 情報処理学会研究報告音楽情報科学 2000-MUS-37-2, pp.9-16 (2000).
- 8) 後藤真孝: 音楽音響信号を対象としたメロディーとベースの音高推定, 信学論(D-II), Vol.J84-D-II, No.1, pp.12-22 (2001).
- 9) Rowe, R.: *Machine Musicianship*, The MIT Press (2001).
- 10) 後藤真孝, 橋口博樹, 西村拓一, 岡隆一: RWC 研究用音楽データベース: ポピュラー音楽データベースと著作権切れ音楽データベース, 情報処理学会研究報告音楽情報科学 2001-MUS-42-6, pp.35-42 (2001).
- 11) Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R.: RWC Music Database: Popular, Classical, and Jazz Music Databases, *Proc. 3rd International Conference on Music Information Retrieval (ISMIR 2002)*, pp.287-288 (2002).
- 12) Logan, B. and Chu, S.: Music Summarization Using Key Phrases, *Proc. ICASSP 2000*, pp.II-749-752 (2000).
- 13) Bartsch, M.A. and Wakefield, G.H.: To Catch A Chorus: Using Chroma-based Representations for Audio Thumbnailing, *Proc. WAS-PAA '01*, pp.15-18 (2001).
- 14) Cooper, M. and Foote, J.: Automatic Music Summarization via Similarity Analysis, *Proc. ISMIR 2002*, pp.81-85 (2002).
- 15) Shepard, R.N.: Circularity in Judgments of Relative Pitch, *J. Acoust. Soc. Am.*, Vol.36, No.12, pp.2346-2353 (1964).
- 16) 大津展之: 判別および最小 2 乗規準に基づく自動しきい値選定法, 信学論(D), Vol.J63-D, No.4, pp.349-356 (1980).
- 17) 後藤真孝, 根山亮, 村岡洋一: RMCP: 遠隔音楽制御用プロトコルを中心とした音楽情報処理, 情報処理学会論文誌, Vol.40, No.3, pp.1335-1345 (1999).
- 18) van Rijsbergen, C.J.: *Information Retrieval*, 2nd edition, Butterworths (1979).
- 19) Hiraga, R., Igarashi, S. and Matsuura, Y.: Visualized music expression in an object-oriented environment, *Proc. Intl. Computer Music Conf.*, pp.483-486 (1996).
- 20) 平賀瑠美, 五十嵐滋, 松浦陽平: 統合演奏視覚化システム, 情報処理学会論文誌, Vol.38, No.11, pp.2391-2397 (1997).

- 21) Smith, S.M. and Williams, G.N.: A visualization of music, *Proc. IEEE Visualization '97*, pp.499-503 (1997).
- 22) Hiraga, R., Miyazaki, R. and Fujishiro, I.: Performance visualization — a new challenge to music through visualization, *Proc. ACM MultiMedia 2002*, pp.239-242 (2002).
- 23) Hiraga, R.: Case Study: A Look of Performance Expression, *Proc. IEEE Visualization 2002*, pp.501-504 (2002).
- 24) Peeters, G., Burthe, A.L. and Rodet, X.: Toward Automatic Music Audio Summary Generation from Signal Analysis, *Proc. ISMIR 2002*, pp.94-100 (2002).
- 25) Dannenberg, R.B. and Hu, N.: Pattern Discovery Techniques for Music Audio, *Proc. ISMIR 2002*, pp.63-70 (2002).
- 26) Hirata, K. and Matsuda, S.: Interactive Music Summarization based on GTTM, *Proc. ISMIR 2002*, pp.86-93 (2002).

(平成 15 年 4 月 11 日受付)

(平成 15 年 7 月 3 日採録)

## 推 薦 文

本論文では音楽のサビを音響信号解析に基づき自動検出し、視覚化することにより、ジャンプしながら試聴することを可能としたシステムを提案している。本論文はインタラクシオン 2003 プログラム委員会において、新規性、有用性ともにかわめて高い評価を受け

ており、推薦論文とするにふさわしいと考える。なお、本論文は、「インタラクシオン：理論、技術、応用、評価」特集号への推薦論文としたい。

(インタラクシオン 2003 プログラム委員長 暦本 純一)



後藤 真孝 (正会員)

1993 年早稲田大学理工学部電子通信学科卒業。1998 年同大学大学院理工学研究科博士後期課程修了。同年、電子技術総合研究所(2001 年に独立行政法人産業技術総合研究所に改組)

に入所し、現在に至る。2000 年から 2003 年まで科学技術振興事業団さきがけ研究 21 研究員を兼任。博士(工学)。音楽情報処理、音声言語情報処理等に興味を持つ。1992 年 jus 設立 10 周年記念 UNIX 国際シンポジウム論文賞, 1993 年 NICOGRAPH'93 CG 教育シンポジウム最優秀賞, 1997 年情報処理学会山下記念研究賞(音楽情報科学研究会), 1999 年平成 10 年電気関係学会関西支部連合大会奨励賞, 2000 年 WISS2000 論文賞・発表賞, 2001 年日本音響学会第 18 回粟屋潔学術奨励賞・第 5 回ポスター賞, 2002 年情報処理学会山下記念研究賞(音声言語情報処理研究会), 2002 年日本音楽知覚認知学会研究選奨, 2003 年インタラクシオン 2003 ベストペーパー賞各受賞。電子情報通信学会, 日本音響学会, 日本ソフトウェア科学会, 日本音楽知覚認知学会, ISCA 各会員。