

# SmartMusicKIOSK: サビ出し機能付き音楽試聴機

後藤 真孝

科学技術振興事業団さきがけ研究 21 「情報と知」領域 / 産業技術総合研究所

m.goto@aist.go.jp

あらまし 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を提案する。店頭で CD を短時間試聴する場合、通常の受動的な聴き方と異なり、早送りしながらサビを探ることが多い。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、サビや楽曲構造の先頭へジャンプする機能と、それらの構造の視覚化機能を実現し、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。提案機能を自動サビ区間検出手法により実現し、試聴機として実装・運用した結果、その有効性を確認した。

## SmartMusicKIOSK: Music Listening Station with Chorus-Search Function

Masataka Goto

“Information and Human Activity,” PRESTO, Japan Science and Technology Corporation (JST) /  
National Institute of Advanced Industrial Science and Technology (AIST)

**Abstract** This paper describes a new music-playback interface for trial listening. In music stores, short trial listening of CD music is not usually a passive experience — customers often search out the chorus or “hook” of a song using the fast-forward button. Listening of this type, however, has not been traditionally supported. This research achieves a function for jumping to the chorus section and other key parts of a song plus a function for visualizing song structure. These functions make it easier for a listener to find desired parts of a song and facilitate an active listening experience. The proposed functions are achieved by an automatic chorus-section detecting method, and the results of implementing them as a listening station have demonstrated their usefulness.

### 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オンをおこなっている。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。例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<sup>1</sup>。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、途中で飛ばし、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を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。

<sup>1</sup> 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。

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り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。例えば「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)~6)</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)の二つの形態があるが、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。ボタンひとつ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が持つ、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、早戻しの二つにより、楽曲内での再生位置を変更できる。しかし、望み通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三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、例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に同じであるが、再生位置スライダーが用意されることがある。スライダー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、スライダー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、以下の二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  
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

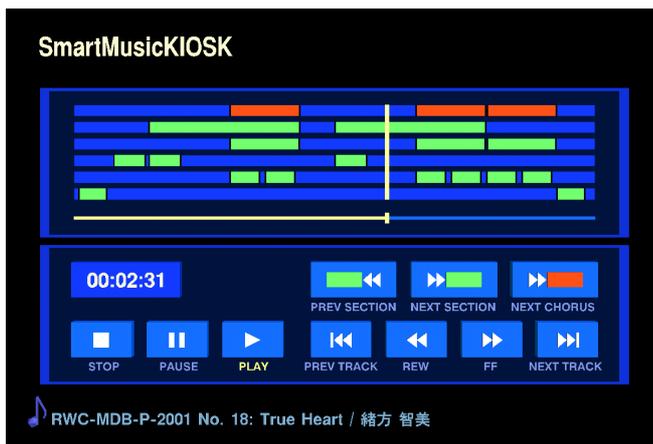


図 1: SmartMusicKIOSK の画面表示: 下側のウィンドウが再生操作ボタン群, 上側のウィンドウが楽曲内容の視覚化結果 (RWC 研究用音楽データベース<sup>7)</sup> RWC-MDB-P-2001 No. 18 に対する自動サビ区間検出結果) を表す. 上側のウィンドウの横軸は時間軸で楽曲全体を表示しており, 最上段がサビ区間の一覧, その下の 5 段が繰り返し構造, その下の横棒が再生位置スライダーを表す.



図 2: 従来の典型的なメディアプレーヤ相当のインター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ーを表す.

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트로, 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, 通常の前送りボタンや再生位置スライダーを使って, 望みの箇所へ移動してもよい.

SmartMusicKIOSK のインター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ーフェース画面も図 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されている. 以下, 図 1 の上下二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三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, このボタンを押す度にそれらの間を順にジャンプできる. 他の二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同士が似ている (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ーである.

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, 再生位置スライダーをクリックして位置変更したりすることが可能である.

以上のインターフェースと機能により, 試聴者は, イン트로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<sup>2</sup>. 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上で, 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

<sup>2</sup> 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」ボタンの二つを用意してもよい. ここでは, 以下の理由から一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も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.

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 RefraiD (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一箇所を、常に指定した長さだけ切り出して提示する手法<sup>8)~10)</sup>はあったが、サビ区間の開始点と終了点はわからず、サビの転調も扱えなかった。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、転調を伴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け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おときには、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一切使わない、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取る。しかし、「繰り返し」と言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、以下のようにまとめられる。

##### 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ない。その際、繰り返す度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み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

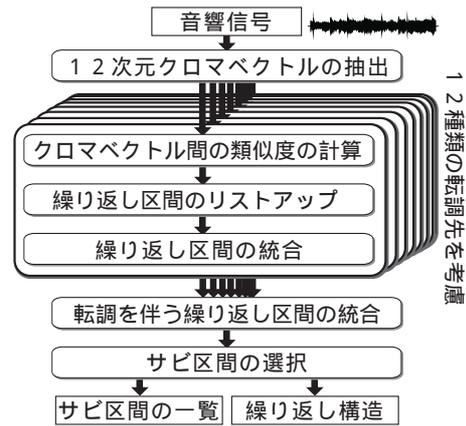


図 3: サビ区間検出手法 RefraiD の処理の流れ

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。例えば、(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けると(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: 転調を伴う繰り返しの検出

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、繰り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り返して起きることが多く、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。

##### 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 256ms, フレームシフト 80ms) 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。 $\vec{v}(t)$  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  $h$  に渡って加算したもので、

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

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 ~ 8000 Hz の6オクターブに渡るよ

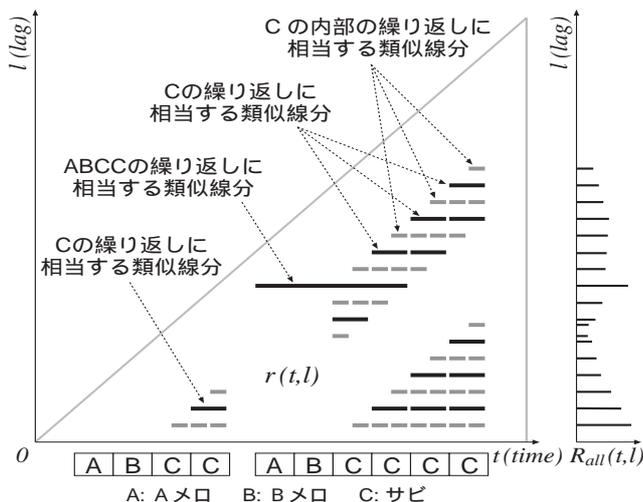


図 4: ある楽曲に対する類似線分, 類似度  $r(t, l)$ , パラメータ空間  $R_{all}(t, l)$  の概念図:  $r(t, l)$  は, 右下半分の三角形内で定義される. 実際に得られる  $r(t, l)$  はノイズを多く含み, サビに関連しない類似線分も存在して曖昧なことが多い.

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. 繰り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>11)</sup>によって, 繰り返しの判断基準 (類似度に対する閾値) を楽曲ごとに自動的に変えながら, 繰り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 4 のように  $t-l$  平面に描画すると, 繰り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  $L_1$  の位置で, 時刻  $T_1 \sim T_2$  の間に類似線分があるとき, 区間  $T_1 \sim T_2$  と  $(T_1 - L_1) \sim (T_2 - L_1)$  が繰り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, 繰り返し区間の一覧が得られる. 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  $l$  の位置に類似線分が存在すると考え, 類似度  $r(t, l)$  が時間軸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. この十分高いことを判断する閾値を, クラス分離度 (クラス間分散) を最大とする判別基準に基づく自動閾値選定法<sup>11)</sup>により定める.

## 3. 繰り返し区間の統合

各類似線分は, 単に二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同士を一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おこなう. 例えば, 図 4 の ABCC の繰り返しに相当する長い類似線分上で, C の繰り返しに相当する類似線分 2 箇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). また, 得られた類似線分が適切であるかも検証する.

## 4. 転調を伴う繰り返し区間の統合

ある演奏のクロマベクトルを  $\vec{v}(t)$  とし, それを半音  $tr$  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  $tr$  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 ( $\vec{v}(t) \doteq S^{tr} \vec{v}(t)'$ ). シフトの操作は, 以下のシフト行列  $S$  で表現できる.

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

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,  $tr$  ごとの 12 種類の類似度  $r_{tr}(t, l)$  を

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

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

## 5. サビ区間の選択

繰り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  $r_{tr}(t, l)$  の平均とし, さらに, 以下の三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り返しがある場合, その末尾部分がサビである可能性が高い. (仮定 3) ある繰り返し区間内に, その区間の半分程度の短い区間が繰り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共に、中間結果として得られた繰り返し構造も出力する。

## 5 システムの実装と結果

3章で述べた機能を実装したSmartMusicKIOSKのプロトタイプ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り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り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よ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>12)</sup>を音響信号の伝送用に拡張したプロトコルである。

以下、RefraiDの評価結果を述べた後に、SmartMusicKIOSKのプロトタイプ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おこなった。ここで評価対象はサビ区間の検出精度のみで、繰り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>7)</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、各曲に対する出力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、適合率、および両者を統合したF値<sup>13)</sup>の観点から評価した。以下に定義を示す。

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

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

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

ただし、転調を伴う場合には、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。

その結果、100曲中の正答曲数は80曲(80曲の平均F値は0.938)であった。誤検出は、サビの繰り返し以外の箇所の繰り返しより多くなかったり、曲中ほとんどが類似伴奏の繰り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。サビの繰り返しで、伴奏やメロディーに大幅な変化を伴う曲は22曲あったが、そのうち21曲は検出できており、その中で、変化を伴うサビ自体は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、二つの提案機能(ジャンプボタンと楽曲構造表示)の有無に応じて、四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、四つの条件とその運用結果を述べる。

条件1 提案機能が一切ない(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おきる傾向があった。

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。効率よ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の利点が共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り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三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わかった。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上で有効であることがわかった。さらに, ボタンの機能や表示ウィンドウの中身に関して一切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わかった。

## 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共に, 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<sup>3</sup>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>14)~17)</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り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

<sup>3</sup> 横軸が時間, 縦軸がMIDIノートナンバー(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。

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>18)~20)</sup>とも関連がある。試聴の目的の一つに, 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, 内容を確認め, 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捉えることができる。従来は楽曲単位で興味のない曲を飛ばせたが, 楽曲内部の興味のな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。例えば, 楽曲全体がどのような構造になっているかを確認め, 構成上の各区間を聴いたり, 繰り返される区間同士の比較をしたりする場面がみられた。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、楽曲中の繰り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わ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、店頭でのデジタル試聴機自体が、SmartMusicKIOSKのような機能を持つものに発展することが望まれる。

#### ● 音楽サムネール (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、画像のサムネールの音楽版とみなせ、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

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

中の繰り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わかった。

今後は、多数のユーザによる評価実験を実施するだけでなく、6章で議論したような各種拡張をおこなっていく予定である。また、試聴以外も広く視野にいった、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。

### 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), J82-D-II, 4, 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ティング検索, 信学論 (D-II), J84-D-II, 12, 2479-2488 (2001).
- [3] 片寄晴弘: 自動採譜, コンピュータと音楽の世界, 共立出版, 74-88 (1998).
- [4] 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, 共立出版, 89-99 (1998).
- [5] 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, 情処研報 音楽情報科学 2000-MUS-37-2, 9-16 (2000).
- [6] 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), J84-D-II, 1, 12-22 (2001).
- [7] 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, 情処研報音楽情報科学 2001-MUS-42-6, 35-42 (2001).
- [8] Logan, B. and Chu, S.: Music Summarization Using Key Phrases, *Proc. of ICASSP 2000*, II-749-752 (2000).
- [9] Bartsch, M. A. and Wakefield, G. H.: To Catch A Chorus: Using Chroma-based Representations for Audio Thumbnailing, *Proc. of WASPAA '01*, 15-18 (2001).
- [10] Cooper, M. and Foote, J.: Automatic Music Summarization via Similarity Analysis, *Proc. of ISMIR 2002*, 81-85 (2002).
- [11] 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), J63-D, 4, 349-356 (1980).
- [12] 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, 情処学論, 40, 3, 1335-1345 (1999).
- [13] van Rijsbergen, C. J.: *Information Retrieval*, Butterworths, second edition (1979).
- [14] 平賀瑠美, 五十嵐滋, 松浦陽平: 統合演奏視覚化システム, 情処学論, 38, 11, 2391-2397 (1997).
- [15] Smith, S. M. and Williams, G. N.: A visualization of music, *Proc. of IEEE Visualization '97*, 499-503 (1997).
- [16] 漆原めぐみ, 平賀瑠美, 五十嵐滋: 楽曲構造に基づく演奏の視覚化と分析, 情処研報音楽情報科学 2000-MUS-34-9, 53-57 (2000).
- [17] 宮崎麗子, 藤代一成: comp-i: MIDI データの 3 次元可視化, 情処研報 音楽情報科学 2002-MUS-46-9, 55-60 (2002).
- [18] Peeters, G., Burthe, A. L. and Rodet, X.: Toward Automatic Music Audio Summary Generation from Signal Analysis, *Proc. of ISMIR 2002*, 94-100 (2002).
- [19] Dannenberg, R. B. and Hu, N.: Pattern Discovery Techniques for Music Audio, *Proc. of ISMIR 2002*, 63-70 (2002).
- [20] Hirata, K. and Matsuda, S.: Interactive Music Summarization based on GTTM, *Proc. of ISMIR 2002*, 86-93 (2002).