

音楽におけるアノテーションとその応用 Musical Annotation and its Applications

梶 克彦[†]
Katsuhiko Kaji

長尾 確[‡]
Katashi Nagao

1. はじめに

近年、インターネット上に氾濫する Web ページやマルチメディアコンテンツを、単に視聴するだけでなく、高度に利用したいという欲求が高まっている。そのためのアプローチとして、現在それらのコンテンツに対してメタ的な情報を関連付けるアノテーションの研究が進められている。Semantic Web [8] やセマンティック・トランスコーディング [4] により、コンテンツとメタ情報を用いた Web コンテンツの高度な検索や、ユーザの特性に合わせた動的なコンテンツ変換などが可能になる。

マルチメディアコンテンツに対するアノテーション研究としては、MPEG-7 [1] が挙げられる。MPEG-7 では、ビデオコンテンツに潜在するシーンの内容、フレーム内のオブジェクトなどのメタ情報を明示的に記述する形式が定められている。MPEG-7 により、意味的な検索や、オブジェクト、シーンの抽出を実現することができる。

音楽も同様に、検索、分類、要約など高度なサービスを実現するためには、オリジナルの音楽コンテンツが持つ情報だけでは不十分な場合が多い。しかし現在、音楽に対するアノテーションの研究はまだ十分には行われていない。そこで、多方面で利用される音楽を高度に発展させるためのアノテーションシステムを実装した。

以下、2 章では本研究で実装した音楽アノテーションシステムの仕組みと機能について、続く 3 章ではアノテーションの応用システムについて述べる。

2. 音楽アノテーションシステム MiXA

本研究では、音楽に対しアノテーションを付与するシステムとして MiXA (MusicXML Annotator) (ミキサと発音する) を実装した。本システムにより、検索、再構成、補完、引用といった高度利用を実現する際に不足している情報を効率的に補うことができる。

以下で、本システムの機能と実装方法を示し、全体のシステム構成を説明する。

2.1 XML ベースのシステム

音楽コンテンツの記述形式として、楽譜を形成するために十分な情報を記述できる MusicXML [2] を採用した。また音楽コンテンツに対するアノテーションの内容も XML として記述できるよう設計した。以降、アノテーションを記述した XML をアノテーション XML と呼ぶ。

図 1 は曲の解説に関するアノテーションを表している。アノテーションが関連付けられている先の要素は、source タグ内に、XPath(XML Path Language) [6] の列として記述される。アノテーションの内容は information タグ内に記述される。dataType 属性の string とは、このアノテーションのデータ型が文字列型であることを表

```
<description id="description(kaji/score-partwise/
  identification/creator[1])/score-partwise/
  identification/creator[2]" x="428" y="83">
<source>
<group>
  <object>/score-partwise/identification/creator[1]</object>
  <object>/score-partwise/identification/creator[2]</object>
</group>
</source>
<information dataType="string">
<string>この二人で紅葉など多くの童謡を発表しています</string>
</information>
</description>
```

図 1: アノテーション XML の一部

```
<annotation id="part" type="reference">
<dataType>string</dataType>
<expression color="#F0D044">構成</expression>
<select>
  <item expression="イントロ" name="intro"/>
  <item expression="サビ" name="chorus"/>
  <item expression="間奏" name="bridge"/>
  <item expression="エンディング" name="ending"/>
  <item expression="Aメロ" name="verse-a"/>
  <item expression="Bメロ" name="verse-b"/>
  <item expression="Cメロ" name="verse-c"/>
</select>
<group>
<item>
  <object minOccurs="0" maxOccurs="unbounded">note</object>
  <object minOccurs="0" maxOccurs="unbounded">rest</object>
  <object minOccurs="0" maxOccurs="unbounded">lyric</object>
</item>
</group>
</annotation>
```

図 2: アノテーション定義 XML の一部

している。本システムでは、文字列型の他に、数値型やコード型といった複数のデータ型のアノテーションを作成することが可能である。データ型については次節において詳しく述べる。

2.2 サービスに応じた柔軟なアノテーション定義

従来の多くのアノテーションシステムは、収集するアノテーションの内容や形式が固定されている。しかし実現したいサービスによって最適なアノテーションの内容や構造が異なる場合が多い。そこでサービスに応じて柔軟に収集するアノテーションの形式を定義することができる仕組みを実装した。

XML ドキュメントに、どの要素に対して、どのような形式のアノテーションの関連付けを許すかといった定義を記述すると、システムは XML に記述されている情報を解析し、その記述どおりの形式のアノテーションを収集するようにメニューを動的に生成する。以降、この XML ドキュメントをアノテーション定義 XML と呼ぶ。

図 2 の XML は、曲の構成に関するアノテーションの定義を記述したものである。dataType タグではアノテーションのデータ型を、string(文字列型)、numeric(数値型)、boolean(真偽型)、chord(コード型) の 4 種類から選択する。コード型は、音楽における和音を記述するための型である。アノテーションの内容を選択式にする場合は、select タグ内に選択することのできるアノテーションを記述することができる。

[†]名古屋大学 大学院情報科学研究科, Graduate School of Information Science, Nagoya University

[‡]名古屋大学 情報メディア教育センター, Center for Information Media Studies, Nagoya University



図 3: MiXA の画面例

どの要素に対してアノテーションを許可するかを記述したのが group タグである。図 2 の場合、音符、休符、歌詞を同時に複数含む要素集合に対して関連付けることができる。また、アノテーションに対してアノテーションの関連付けを許すことも可能である。コード進行のように、一つの単純なアノテーションでは記述しきれないような情報でも、階層化したアノテーションの集合により記述することが可能である。

2.3 Web ブラウザを介したアノテーション

主観的な感想、意見など、楽曲の解釈に関する情報は音楽を意味的に理解するための重要な情報であるといえる。本システムは、このような主観的な情報を幅広いユーザから収集し、多くの応用システムに役立てるために、煩雑な手続きを必要としない Web ブラウザベースのシステムを採用した。

多人数のユーザを対象とする場合、アノテーションの信頼性を考慮する必要がある。そのため、アノテーションを行うユーザはあらかじめユーザ登録を通して個人プロフィールの XML を作成する。本システムを利用する際にはベーシック認証を通してログインする。アノテーション作成の際には、同時にユーザの ID も保存するため、誰がどのアノテーションを作成したかを識別することができる。このようにアノテータを特定する機能を持っているという事実により、間違った情報の書き込みが減り、アノテーションの信頼度を高めることができる。

2.4 楽譜に対するアノテーション

従来の音楽に対するアノテーションの研究として、楽曲に対して曲調を関連付ける研究 [9] などがある。しかし多くの楽曲は認識する箇所によって印象が異なるため、詳細に曲調検索を行う場合などには情報が足りない。そのため、楽曲を構成する詳細な要素に対してもアノテーションを関連付けることができる必要がある。そこで、本システムではユーザに楽譜を提示し、楽譜中に現れる音符、歌詞などの要素に対して情報を関連付けられるようにした。また必要に応じて実際に楽曲を演奏することで直感的に楽曲中の要素を理解できるように設計した。

ユーザに提示される楽譜には、図 3 のようにその音楽コンテンツに関連付けられているアノテーションを可視化したオブジェクトが重ね合わされて表示される。以降このオブジェクトをアノテーションオブジェクトと呼ぶ。

アノテーションの作成手順を説明する。まず要素を選択し、必要であれば選択部分を視聴して確認する。次に図 4 のようにメニューからアノテーションの種類を選択し、その後具体的な情報を記述する。このメニューは、



図 4: 選択した要素に対するアノテーション

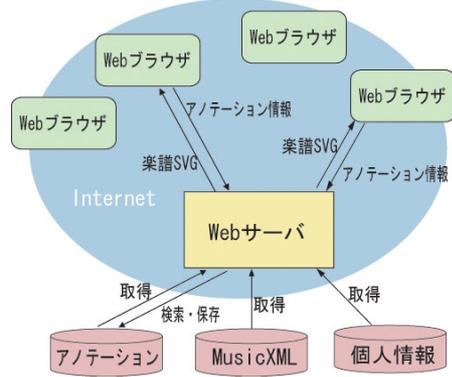


図 5: MiXA のシステム構成

あらかじめ記述されているアノテーション定義 XML に従い、選択済みのオブジェクト集合に対して関連付けを許すアノテーションの一覧を動的に生成している。

文字列型のアノテーションは、サブウィンドウのプロンプトから記入する。またアノテーション定義 XML には、指定した内容に対応する選択式のアノテーションを記述することができるが、その場合アノテーションは右クリックメニューからの選択により容易に作成することができる。一方、コード型のように複雑な構造を持つデータ型の場合は、サブウィンドウから詳細な内容を記入する。

2.5 システム構成

図 5 は、本システムの構成の概要である。

ユーザは Web ブラウザからベーシック認証を通して本システムにアクセスし、一覧から、または後述する検索システムから曲を選択する。Web サーバは要求された曲の MusicXML をデータベースから取り出し、ブラウザで表示できる SVG (Scalable Vector Graphics) [7] 形式の楽譜に変換する。各要素には、元の MusicXML のどの部分を指すオブジェクトかを指し示す XPath が付属されている。また楽譜には、該当する曲に関するアノテーションを視覚的に重ね合わせてブラウザに表示する。

アノテーションの保存の際には、アノテーションの内容や、楽譜上の座標などをサーバに送信する。サーバは、1 ユーザ、1 コンテンツにつき一個のアノテーション XML を生成し、XML データベースに保存する。

3. アノテーションの応用

本論文ではアノテーションの応用として検索・再構成システムを想定し、収集するアノテーションを意見・感想、解説、印象、コード、曲の構成の 5 種類に設定し、アノテーション定義 XML に記述した。また実際にこれらのアノテーションを作成した。以下に音楽検索・再構

成システムについて詳細を述べる。

3.1 音楽検索システム

本検索システムでは、MusicXML に記述されているタイトルや作詞・作曲者の情報に加えて、アノテーションを用いたキーワード検索、コード進行による検索、曲の構成による絞り込み検索を行い、算出された検索ランクに応じた結果一覧を作成する。この一覧からも MiXA のアノテーション画面や、後述する再構成システムへ移行することが可能である。以下にキーワード検索、コード進行による検索、曲の構成による絞り込み検索の詳細を述べる。

3.1.1 キーワード検索

ユーザはタイトル、作詞者、作曲者、意見・感想、解説の情報に関してブラウザのフォームからキーワード検索の要求を行うことができる。Web サーバはその検索要求を受け取り、以下のようにして検索を行う。

タイトル、作詞者、作曲者の情報に関しては、MusicXML が格納されているデータベースに対して検索を行い、該当する楽曲を得る。意見・感想、解説、印象については、まずアノテーションが格納されているデータベースに対して検索を行い、関連先の MusicXML を取得する。検索ランク $rank$ は同一の MusicXML に対するキーワードの出現数 n とする。

3.1.2 コード進行の検索

ユーザは検索フォームのサブウィンドウから順にコードを入力していくことで、コード進行による検索をすることができる。

サーバはまず、検索要求であるコード進行から相対的な音の変化を計算する。「C G/B Dm C」というコード進行の場合は、まず最初の音 C を 0 とし、この音と後続する音との相対差を計算する。C と G の相対値は、C から 5 つキーを下げると G になるため -5 である。上記のコード進行の相対変化は「0,-5/-1,2m,0」となる。

次に、各音楽コンテンツに対して、コードのアノテーションを取得する。そして、一曲分のコードに対し、相対的な変化を前から順に照合していく。コード進行の相対変化が一致した場合、検索要求のコードとの絶対的なキーの差を $diff$ とし、類似度 sim を計算する。 i 番目のアノテーションのコード進行と、検索要求のコード進行との類似度 sim_i は以下の式で表す。

$$sim_i = \frac{1}{(diff_i + 1)}$$

順に n 回コード進行の類似度計算を繰り返し、足し合わせたものが検索ランクとなる。

$$rank = \sum_{i=1}^n (sim_i \cdot base)$$

3.1.3 曲の構成による絞り込み検索

本検索システムでは、サビが悲しい曲といった曲の構成による絞り込み検索が可能である。この場合、「サビ」が絞り込みを行う曲の構成であり、検索対象は「悲しい」という印象のアノテーションである。

まず、曲の構成以外の検索要求である「悲しい」やコード進行について、前述の検索を行う。この時の検索ランクを $priorrank$ とする。

次に、絞り込む曲の構成「サビ」の部分に対して、どれだけ検索対象である「悲しい」というアノテーション

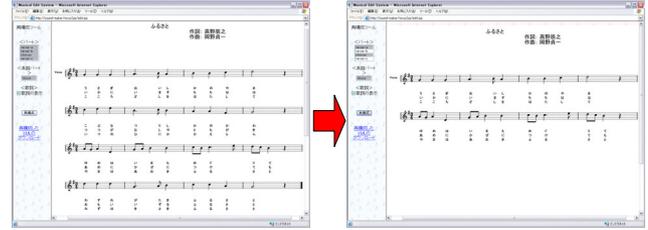


図 6: 再構成前 (左) と再構成後 (右)

が含まれているかという、含有率 $contains$ を計算する。「サビ」が関連付けられている要素は o_j から o_l までで、その要素数は $l - j$ であるとする。また、「悲しい」が関連付けられている要素は o_i から o_k までで、その要素数は $k - i$ であるとし、これら二つのアノテーションが共通して関連付けられている要素は、 o_j から o_k までで、その要素数は $k - j$ であるとする、含有率 $contains$ は以下の式で求められる。

$$contains = \max\left(\frac{k-j}{k-i}, \frac{k-j}{l-j}\right)$$

さらに、最終的な絞り込み検索の検索ランクは以下の式となる。

$$rank = contains \cdot priorrank$$

3.2 音楽再構成システム

再構成システムでは、歌詞、楽器パートと曲の構成から、ユーザが特に聴きたい、楽譜を表示したい部分を選択することにより、容易に曲を再構成することができる。例えば元の長い曲をイントロ サビ エンディングという構成に縮めるといったことが可能である。

ユーザはまず、MiXA の楽曲一覧ページや検索システムの検索結果一覧から、再構成を行いたい曲を選択する。図 6 の左側は再構成前の楽譜を表示した画像である。ブラウザの左側のフレームから、ユーザが取得したい曲の構成、歌詞の有無や楽器パートの選択を行い再構成の要求を行うと、サーバはユーザの再構成の要求を反映させた MusicXML を作成し、楽譜の SVG に変換して図 6 の右側のように再構成後の楽譜を表示する。再構成された MusicXML をダウンロードすることも可能である。

4. 実験

10 人の被験者に MiXA のマニュアルページを読んでもらい基本的なアノテーションの手順を習得してもらった。その後童謡 9 曲に対してアノテーションを行ってもらい、以下の項目について 5 段階評価をしてもらった。

- 質問 1: システム全体の使いやすさ
- 質問 2: 楽譜、アノテーションオブジェクトの表示
- 質問 3: 自分の思い通りにアノテーションできたか
- 質問 4: MiXA を用いてアノテーションをしたいか

今回の実験は、システムのユーザビリティに関して評価項目を設定した。アンケート結果を図 7 に示す。

質問 1 の評価の平均は 3.7 であった。保存を行わずブラウザを終了させてしまう場合が多くみられた。アノテーションの保存を促す手段を講じる必要がある。

質問 2 の評価の平均は 3.6 であった。現在の表示方法が、十分に適切であるとはいえない結果である。アノテーションをリストで表示するなど、楽譜に重ね合わせる以

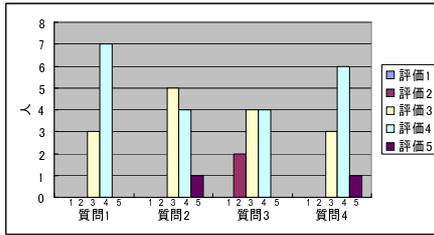


図 7: 評価実験の結果

外の表現手法を検討する必要がある。

質問3の評価の平均は3.2であった。質問と回答によりその曲に対する理解を深めていきたいという要求があった。アノテーションの項目についての検討が必要である。

質問4の評価の平均は3.9であり、高い評価結果となった。本システムによりユーザ同士のインタラクションが可能のため、情報収集の場となりうる。多くのユーザが集まれば、アノテーションの収集効率も高まるため、より応用サービスを高度なものにすることができる。

評価実験の結果、複数の問題点も挙げたが、本システムはアノテーションの作成支援システムとして有用性が高いと考えられる。また、本システムは、今回の評価実験によって明らかになった諸問題の多くについても今後対応できる拡張性を備えている。今後も研究を進め、これらの問題を解決していく予定である。

5. 関連研究

5.1 CDDB

CDDB [3] は音楽 CD に収録された楽曲情報を検索・認識するデータベースである。CDDB は音楽コンテンツそのものに対して、タイトルなどのアノテーションを付与することができる仕組みである。しかし、音楽コンテンツの詳細な部分要素に対してのアノテーションはこの仕組みでは実現することができない。楽曲の詳細部分に対してアノテーションを行うためには、本研究で実装したように、楽譜に現れるオブジェクトに対するアノテーションなどのアプローチを取る必要がある。

5.2 パピブーン

パピブーン [10] は、GTTM のタイムスパン簡約と演繹オブジェクト指向データベースを利用し、音楽を要約するシステムである。GTTM (Generative Theory of Tonal Music) [5] とは、音楽理論のグルーピング構造と拍節構造を分析する代表的な手法であり、タイムスパン簡約とは、これらの構造を用いて解析されるもっとも安定的なまとまりを形成する階層構造である。パピブーンによりユーザとのインタラクションを通じて、質の高い要約を生成することができる。要約の前処理として TS-Editor を用いて GTTM のタイムスパン簡約に基づく曲分析を行う。MiXA は楽譜を使用しているが、TS-Editor はピアノロールとツリー形式で楽曲を表現している。

6. まとめと今後の課題

本論文では、音楽コンテンツに対するアノテーションシステム MiXA について、新規性と意義、方法について述べた。本システムは楽譜を用いて楽曲の詳細要素に対してアノテーションを関連付けることができ、多くのユーザが非同期に利用することが可能である。さらに、

応用システムに応じて収集するアノテーションの形式を定義することができる機能を持つ。現在本システムを以下の URL で試験的に公開している。

<http://www.nagao.nuie.nagoya-u.ac.jp/demo/mixa/>

また検索・再構成システムの実装とユーザビリティに関する評価実験を行い、有効性を確認した。以下に今後の課題を挙げる。

6.1 メタ情報の自動生成とその編集

曲調やコード進行のように、ある程度機械的に自動解析できる情報については、曲を登録する際にそれらの情報を自動生成することが可能である。しかしその情報は十分な精度ではない場合が多い。ユーザがその情報を編集、訂正し、精度のよい解析結果に昇華することにより、さらに高度な応用につなげることができる。このように、機械が生成した粗い情報を、人手でユーザが編集、訂正していくという方式をとれば、アノテーションのコストをかなり削減することができる。

6.2 別の視点でのアノテーション

楽譜に対してのアノテーションは関連付ける要素を詳細に選択できるというメリットがある。しかしアノテーションを行うユーザの労力は比較的大きくなってしまふ。楽譜上でのアノテーションの他に、視聴時に簡単なボタンにより印象を関連付けていく手法や、楽曲の作成時の履歴を利用するといった手法を併用し、アノテーションのコストを下げる必要がある。

6.3 アノテーション応用システムの拡張

多くのユーザは「サビの明るい曲」という検索要求を行った場合、サビの部分のみを聴きたいと考えるだろう。そこで、検索に使用された曲の構成のアノテーションを用いて検索された曲を自動的に再構成し、検索結果一覧から再構成済みの曲のプレイリストを作成することができるシステムを実装する予定である。

また遠隔音楽教育について検討中である。教師と学習者がアノテーションシステムを介してインタラクションを行うことにより、楽曲に対する理解を深めると同時に学習者の演奏練習の質を高めるシステムである。

参考文献

- [1] MPEG-7. <http://www.mp7c.org/>, 2002.
- [2] Michael Good. MusicXML in practice: Issues in translation and analysis. *International Conference Musical Application using XML*, pp. 47-54, 2002.
- [3] Gracenote. CDDB. <http://www.gracenote.com/gn-products/cddb.html>, 2003.
- [4] Katashi Nagao. *Digital content annotation and transcoding*. Artech House Publishers, 2003.
- [5] Fred Lerdahl Ray Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, 1996.
- [6] W3C. XML Path Language. <http://www.w3.org/TR/xpath.html>, 1999.
- [7] W3C. Scalable Vector Graphics. <http://www.w3.org/Graphics/SVG/>, 2003.
- [8] W3C. The semantic web community portal. <http://www.semanticweb.org/>, 2003.
- [9] 吉野太智, 高木秀幸, 清木康, 北川高嗣. 楽曲データを対象としたメタデータ自動生成方式とその意味的連想検索への適用. 研究報告「データベースシステム」No.116 - 041, 2001.
- [10] 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. 研究報告「音楽情報科学」No.046 - 005, 2002.