MiXA: A Musical Annotation System

Katsuhiko Kaji
Graduate School of Information Science
Nagoya University
Furo-cho, Chikusa-ku,
Nagoya 464-8603, Japan
kaji@nagao.nuie.nagoya-u.ac.jp

Katashi Nagao
EcoTopia Science Institute
Nagoya University
Furo-cho, Chikusa-ku,
Nagoya 464-8603, Japan
nagao@nuie.nagoya-u.ac.jp

Abstract

We present a Web-based musical annotation system named MiXA (MusicXML Annotator). The system enables users to associate any element of musical content, such as a note, a lyric, and a title, with some additional information such as chords, comments, and impressions. MiXA can also handle annotations created by multiple users, and these annotations are managed separately for each annotator. This allows application systems for annotations to deal with various comments created by the different annotators. Since the content and form of annotation data depend on applications or services, the system allows application developers to define their own semantics of annotation data.

We also present several application systems, such as music retrieval and reproduction, based on annotations acquired through MiXA.

1 Introduction

Recently, the demand for advanced use of multimedia content has experienced rapid growth. To permit such advanced use, such as semantic-content-based retrieval and summarization, several studies have been performed on annotation that associates the multimedia content with metadata. For example, some researches on the Semantic Web (W3C, 2003b) and Semantic Transcoding (Nagao, 2003), which focus on semantic annotation of Web content, deal with pinpoint retrieval of target content and personalization/ adaptation of content according to our characteristics, which change dynamically.

MPEG-7 (MPEG-7 Consortium, 2002) is one of the recent hot activities on multimedia annotation. In MPEG-7, a data format is defined that describes various types of information related to audio-visual content (e.g. scene and object description). The MPEG-7 format enables semantic retrieval and automated analysis such as object tracking in frames of multimedia content.

Similarly, musical content works better when some additional information is accompanied by it. Such information will enable some advanced services such as retrieval, classification, summarization, etc. This should also lead to the development of some musical annotation systems. Musical annotation data that can be utilized in many applications should include the following information: information on the musical piece itself (genre, work year, etc.), and information about the detailed portions of a musical piece (a note, lyrics, etc.).

Since music is an art, interpretations of a musical piece are strongly dependent on people's subjectivity. We think that open global annotation systems can cover the diversity of the musical piece interpretation by collecting many users' subjective data through the Internet. Additionally, since the required information changes according to types of services, the annotation system should allow the administrator to customize the content and format of annotation data according to service requirements.

```
<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>
```

Figure 1: Example of Annotation XML (a part)

In this research, we implemented a musical annotation system that satisfies these requirements and can be adapted to various applications. The following sections describe the MiXA architecture and functionality and some application systems using annotation data collected through MiXA.

2 MiXA: A Musical Annotation System

We developed MiXA (MusicXML Annotator) (pronounced "mixer") as a musical annotation system. The system can assist in the construction of various musical application systems (e.g. retrieval, summarization, supplementing and citation) due to its capability of efficiently supplying required information. The functions and the implementation details of this system are described in the following sections.

2.1 XML-based system

We adopt MusicXML (Good, 2002) as the form for describing music, since it can describe information sufficiently well to form a score or play the music. The form of annotation is also described in XML. Hereafter, such a document is called "Annotation XML."

Fig. 1 illustrates an annotation of musical description. The object associated with the annotation is described as XPath (XML Path Language) (W3C, 1999) the in the "source" element, while the content of the annotation is in the "information" element. In the figure, "String" is described as a "dataType" attribute, which means that the annotation is written in string. MiXA prepares a number of data types such as "numeric" or "chord". Further details of data types are described in the following section.

2.2 Flexible definition of annotation according to services

In many conventional systems, the contents and form of annotations are fixed. Consequently, few application systems that employ such annotation have been realized. To solve this problem, we implemented a scheme that can define the form of annotation flexibly according to the service.

```
<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>
```

Figure 2: Example of Annotation Definition XML (a part)

For preparation, it is necessary to describe the definition (form and objects that permit association) of annotation in XML documents. The system then analyzes the document and collects annotations of the form as the description. Henceforth, this document is called "Annotation Definition XML."

Fig. 2 represents the definition of musical structure. The annotation data type of is chosen from four varieties: the string type; the numeric type; the boolean type; and the chord type. The chord type is that which describes harmony. When making the contents of annotation choose, annotation is enumerated in select element.

Objects which allow to be associated with is described in group element. In Fig. 2, annotation can associate with the object set that contains two or more notes, rests, and lyrics simultaneously. At the same time, it is also possible to allow correlation of one annotation to another. Like chord progression, there exists information that cannot be described in one simple annotation. In the case like that, such complicated information can be described as a set of simple annotations.

2.3 Annotation through Web browsers

Subjective information (affection, impression, etc.) is essential information for recognizing music semantically. To collect such important annotations from a broad spectrum of people, we designed the system as a Web service.

When aiming at a lot of people, it is necessary to take the reliability of annotations into consideration. For this reason, we created at user register for the system and created an XML profile beforehand. After registering, each user can log-in to the system through a basic authentication step. Each annotation includes information on the person who create it, and the system includes a function to specify the annotator, which helps in holding down the number of irresponsible annotations.

2.4 Music expression format

There is research which associates affection with a musical piece itself (Taichi Yoshino,Hideyuki Takagi, Kiyoki Yasushi, Kitagawa Takashi, 2001 in Japanese), though the impression changes with the part of the musical piece the listener recognizes. There is still a lack of data that would enable us to create a detailed retrieval system for music affection, To solve this problem, annotations should be associated with detailed objects of music. Consequently, we adopt the score as the form of music because users can easily select partial objects of music from a score. Moreover, a user can understand the selected object by performing that chosen part of the musical piece.

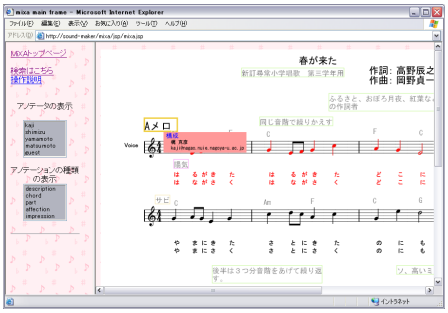

Figure 3: Example of score generated by MiXA

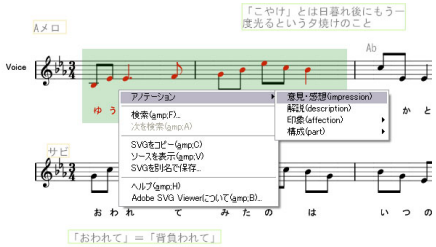

Figure 4: Annotation to selected objects

In the score, the annotation related to the musical objects is also displayed. Hereafter, these objects are called "Annotatoin Objects."

The following procedures create annotations. First, the user selects the objects that will be associated to annotation. Next, as in Fig. 4, the user selects the annotation type from the menu, after which he or she can describe concrete information. The annotation menu is dynamically generated according to the selected objects and the Annotation Definition XML.

If the annotation's data type is of the string or chord type, etc., that information is added via a sub-window. Besides, if the annotation is selective, information can be selected easily from the annotation menu.

2.5 System architecture

Fig. 5 shows an outline of MiXA's architecture.

The user logs into this system through basic authentication with a Web browser and selects one piece of music from music list or retrieval result. The Web server takes the MusicXML of the demanded music from the XML database and transforms it into score described by SVG (Scalable Vector Graphics) (W3C, 2003a). Each object has an XPath, along which each element of MusicXML is pointed. Anno-

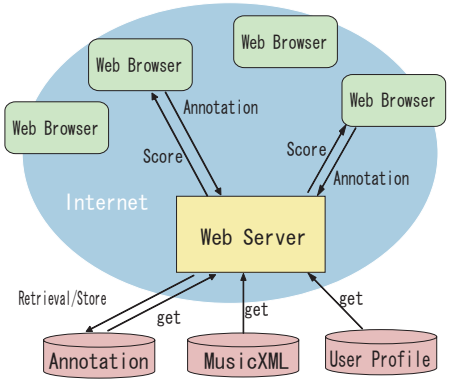

Figure 5: System architecture of MiXA

tation objects associated with the music are also displayed in the score.

To preserve the annotations, the contents of annotation and the coordinates of a score, etc. are transmitted to a Web server. One Annotation XML per user and one content is generated and stored.

3 Application Systems

In this paper, we suppose music retrieval / reconstruction as an application system based on annotations. To realize these applications, we have defined five types of annotation: "Impression"; "Affection"; "Description"; "Chord"; and "Structure". These types are described in Annotation Definition XML. Here, "Structure" means musical rough structure such as the intro, the chorus, the bridge and the ending.

Details of these systems are given as follows.

3.1 Music retrieval system

In this retrieval system, it is possible to retrieve data by chord progression or keywords using annotation in addition to the information on the title, lyrics and composer as described by MusicXML. Furthermore, the user can narrow down the search by using the music's rough structure. Because the system creates a list of results in the high order of the calculated retrieval rankings, it is possible to shift to the annotation phase and the reconstruction system (mentioned later) also from this list.

3.1.1 Retrieval by keyword

With retrieval by keyword, the user can input title, composer, lyricist, impression and description via a Web browser. The server receives the retrieval request and conducts its search as follows.

For information on title, lyricist, and composer, it searches the database on which MusicXML is stored, and the corresponding musical piece is obtained. For other information, the system searches for the musical piece under two time frames: First, it searches through the database in which annotations are stored to obtain the corresponding annotation(s). Second, the musical piece associated with the annotation is obtained.

The retrieval rank "rank" is the number of appearances n of the keyword for the same music.

3.1.2 Retrieval by Chord progression

To retrieve a particular chord progression, the user pushes the "chord progression" button for the retrieval form. Using the following method, the system provides a music list retrieved by chord progression. First, a relative value is calculated from the specified chord progression, which is a reference request. For example, "C G/B Dm C" is changed relative to the chord progression "0,-5/- 1,2m,0." Second, for each music content, annotation of a chord is acquired and a chord progression is created. From the required relative chord progression and the chord progression retrieved from the annotations, the distance between a request and the correct response is measured by DP-matching; the distance decreases with a more correct answer. The rank is a normalized value, which is the inverse of the distance.

Our retrieval algorithm for chord progression is based on the algorithm introduced in (Tomonari Sonoda, Masataka Goto and Yoichi Muraoka, 1998).

3.1.3 Filtering of search results based on musical structure

In this retrieval system, it is possible to filter by musical structure, such as with "music with a sad chorus."

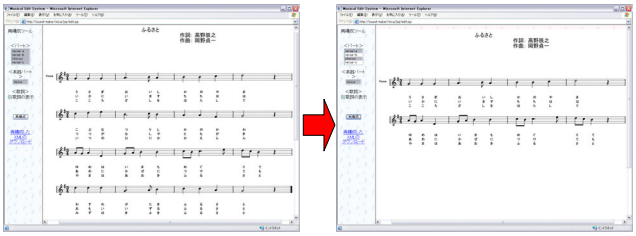

Figure 6: Original music (left), After reproduction (right)

The server receives user requests for filtering as well as keyword and chord progression requests. If the system receives "music with a sad chorus," then "chorus" is the musical structure used for filtering. Another retrieval request is "sad" affection.

First, the above-mentioned retrieval is performed for keywords or chord progressions, and the rank of the result is "priorrank". Each keyword or chord progression, "content" is calculated based on how much of it is contained in the filtering request.

The music object with which annotation of "chorus" is associated ranges from o_j to o_l, and the number of its objects is $l - j$. The music object with which annotation of "sad" is associated ranges from o_i to o_k, and the number of its objects is $k - i$. The object with which these two annotation is associated in common ranges from o_j to o_k, and the number of its objects is $k - j$. The "content" is then calculated in following formula.

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

The "content" become large when many objects associated with the chorus are contained in "sad," or when many objects associated with "sad" are contained in the chorus. The rank of the final filtering is derived by the following formula.

$$\text{rank} = \text{content} \cdot \text{priorrank}$$

3.2 Music reproduction system

We have also implemented a musical reproduction system. By using the lyrics, the instrumental part, and musical rough structure (such as the introduction, the chorus, the bridge, and the ending), the user can reconstruct original music to the music of a favorite composition. For example, if the original music is too long, we can shorten the music, like in an intro-chorus- ending.

First, the user selects one piece of music from the entire music list or retrieval result. The left-hand side of Fig. 6 shows the original music. The user can choose the musical structure, lyrics and instrumental part from the left frame of the browser. After requesting reproduction, the server creates a MusicXML constructed from the user's request and converts it to the SVG form (right-hand side of Fig. 6). In addition, the user can download the reconstructed MusicXML.

4 Experimentation

We evaluated MiXA using a series of experiments conducted with the help of 30 human subjects. This evaluation experiment was designed to determine whether the system can collect annotations efficiently from a network and whether annotations associated to the partial structure of a musical piece can be used for application services.

We implemented a retrieval system based on information associated with the musical piece itself such as genre, work

	average	standard deviation
Q1	3.87	0.89
Q2	3.93	0.99
Q3	3.80	0.98
Q4	3.24	0.64
Q5	3.82	0.70

Table 1: The result of subject experiments

year, and length of the music for these experiments. Hereafter, this system is called "R1".The retrieval system using annotation introduced in this paper is called "R2".

First, we had the 30 subjects read MiXA's manual page until they mastered the procedure of fundamental annotation. Then they created annotations to the music of nine traditional songs, and answered questions 1 and 2. Following that, they actually searched for the musical pieces by R1 and R2 and answered questions 3, 4, and 5. The questions are given as follows.

- (Q1) Could annotation be created smoothly?
- (Q2) Was it possible to intuitively select a partial object of a musical piece?
- (Q3)Do you want to use retrieval for the partial structure of a musical piece?
- (Q4) Ease of musical piece retrieval by R1.
- (Q5) Ease of musical piece retrieval by R2.

Table 1 shows a summary of the experimental results.

Here, Q1 and Q2 are evaluations of whether annotations could be collected accurately from a network. The average for Q1 was 3.87,with a standard deviation of 0.89. The average for Q2 was 3.93, with a standard deviation of 0.99. By using this system, it could be said that annotations are effectively collectable from a network.

Questions 3, 4, and 5 are evaluations of whether annotations for a partial object of a musical piece can be used effectively. The average for Q3 was 3.80, with a standard deviation of 0.98. From these results, we consider that many people were motivated to retrieve to partial structures of a musical piece. Moreover, the average for Q4 was 3.24 and that for Q5 was 3.82, giving a difference between them of 0.58.

We performed a paired t-test to verify whether there was any significant difference between R1 and R2; the result was 2.76. Consequently, at a 5% significance levels, we conclude that R2 was more effective than R1. This means that retrieval using annotations associated with a partial object of a musical piece is superior to retrieval from information related to the musical piece itself, such as genre. R2 was especially effective when the description of the music desired for retrieval was vague, or the user knew only some part of the music. Moreover, we received the comment that retrieval by chord progression will also be helpful for those who are studying a composer or musical theory.

In fact, it could be said that annotations collected by MiXA are effective in realizing an advanced service.

5 Related Research

5.1 Papipuun

Papipuun is a summarization system based on polyphony, which is itself based on time-span reduction in the generative theory of tonal music (GTTM) (Ray Jackendoff, 1996) and the deductive object-oriented database(DOOD). GTTM is a typical theory for analyzing musical structure. By using Papipuun, it is possible to generate a high-quality music summary through interaction with a user.

As for preparing a summary, TS-Editor helps the annotator to create a polyphony based on time-span reduction. For example, TS-Editor displays the musical piece in a piano roll.

Ideally, according to the kind of annotation associated with a musical object, the user should be shown the optimal form.

6 Summary and further work

This paper described the novelty, meaning, and the method of the musical annotation system MiXA. MiXA can associate annotations to detailed objects of a musical piece by using score. Another feature is that it can be used by many users asynchronously. Furthermore, it has function in which the form of an annotation, according to an application system, can be defined. We also implemented a retrieval/reproduction system using annotations. Evaluation experiments were conducted, and the system's validity was checked.

Some future research items are listed below.

6.1 Use of automatically-generated annotation

Although it is nessesary to employ information that can be used to carry out automatic analysis somewhat mechanically, like a music tone or chord progression, such information analyzed automatically is not sufficiently accurate in many cases. In such cases, however, a user's editing and correcting of the analysis result, can produce more accurate data.

Thus, it is possible to considerably reduce the cost of annotation if the system with which a user edits and corrects the coarse information analyzed automatically is adopted.

6.2 Another point of view for collecting annotation

Annotation to a score has the advantage that an object that is associated can be identified precisely; however, the labor of the annotator is likely to increase as a result.

As a result, it is necessary to integrate other techniques to make it easier to create a history for a musical piece while listening to it.

6.3 Extension of application systems

We are considering developing an online musical education system based on annotation to deepen students' understanding of musical pieces while interacting with their teachers. The aim is to improve the quality of students' practice performance.

References

Michael Good. 2002. Musicxml in practice: Issues in translation and analysis. *International Conference Musical Application using XML*, pages 47–54.

MPEG-7 Consortium. 2002. *MPEG-7*. http://www.mp7c.org/.

Katashi Nagao. 2003. *Digital content annotation and transcoding*. Artech House Publishers, London.

Fred Lerdahl Ray Jackendoff. 1996. *A Generative Theory of Tonal Music*. MIT Press.

Taichi Yoshino,Hideyuki Takagi, Kiyoki Yasushi, Kitagawa Takashi. 2001 (in Japanese). The metadata automatic generation system for musical data and its application to semantic associative reference. *Research report 'Database System' No.116 - 041.*

Tomonari Sonoda, Masataka Goto and Yoichi Muraoka. 1998. A www-based melody retrieval system. *ICMC'98 proceedings, pp.349-352*.

W3C. 1999. XML Path Language. *http://www.w3.org/TR/xpath.html*.

W3C. 2003a. Scalable Vector Graphics(SVG). *http://www.w3.org/Graphics/SVG/*.

W3C. 2003b. The semantic web community portal. *http://www.semanticweb.org/*.