Melody Retrieval with Humming

Tetsuya Kageyama  Kazuhiro Mochizuki  Yosuke Takashina
Information Technology Research Laboratories, NEC Corporation
NEC Scientific Information System Development Ltd.

Abstract
This paper describes a melody retrieval method which takes a hummed melody as a search clue for information retrieval from databases. The system using this retrieval method processes a query in the following manner. At first, the system transcribes a hummed melody fragment into melody sequence data. Next, the clue melody data is compared with each melody in the melody database to find the song which most closely matches the clue. If this retrieval, a user is able to retrieve song information, e.g. bibliographical data, from a hummed melody fragment based on a user's uncertain memory. The authors developed a prototype system with a 590 song database on a personal computer. The system replies to a query in 20 seconds on an average, and the answer accuracy is 80%.

1 Introduction
A musical information retrieval system should utilize melody information as well as bibliographical information so that the system may handle melody clues like bibliographical clues in database searching. Dillon and Hunter [1982] reported melody matching strategies for seeking melodic variants in folk music research [Dillon and Hunter, 1982]. Yamamoto [1988] built a music database system which can retrieve song information by using a melody pattern clue [Yamamoto, 1988]. These systems, however, expected users to be musical researchers and queries to be musically correct. The authors describe a melody retrieval method, which takes a solo melody as a search clue. The retrieval method is designed for musically untrained novice users, and is able to cope with inaccurately/ambiguously pitched melodies.

In the following, an overview of this method, the similarity definitions between two melodies, prototype system and evaluation results are described.

2 Melody Retrieval
The melody retrieval method proposed by the authors takes a hummed melody fragment as a retrieving clue.

The humming input provides the system with more friendly user interface than the ones offered by previously developed systems.

The system is capable as well of retrieving data with an ambiguous melody fragment input, enables a user to search song related information more flexibly than on other systems which require exact clues for

![Figure 1: Melody Retrieval System](image)

retrieval.

The system using the proposed method consists of three modules; a transcription module, a melody matching module and a melody database to be searched through (see Fig. 1).

Each module has the following functions.

1. Automatic transcription module transcribes a hummed, analog signal, to a note sequence by the following processes (a) through (d) (Tsuruta et al., 1988).

(a) Analog-to-digital conversion
(b) Pitch-Power extraction
(c) Note segmentation
(d) Pitch name determination

ICMC Proceedings 1993  349  2P.94
The automatic transcription is a preprocess for melody matching.

2. The matching module attempts to match every melody in the database with the clue, carrying out similarity calculations. The song with a melody which has the highest similarity is then chosen as an answer to the query.

3. The melody database contains all song data which are of searching target. The song data contains the entire melody of the solo part, and other bibliographical information regarding the song, e.g. the title, the name of the composer, etc. Individual melody data in the database is represented as a sequence of notes and rests. In a sequence, an individual note is represented by two attributes, pitch and duration of the note. An individual rest is represented by one attribute, rest time.

3 Similarity Calculation between Two Melodies

3.1 Clue Ambiguosness

In the melody matching module, the retrieval result is determined by the similarity calculation between a clue melody and the melody in the melody database. The input clues, however, sometimes are inaccurate, because they have ambiguousness based on a human memory. The similarity definition is required to take those clue inaccuracies into consideration.

The difference between a clue melody and the melody in the melody database can be classified into two categories: global and partial. The global difference is pitch-key difference between the clue melody and the song in the database to be retrieved. The partial difference (referred to as “the clue error” hereafter) is due to the following.

- Imperfect user memory
- Singer’s vocal organ limitations

What is still worse, the automatic transcription results contain errors, if the humming or singing manner is in the following cases.

- When two adjacent notes of the same pitch are hummed rather continuously, the system may fail to divide them.
- When the pitch of a note fluctuates over a semitone range, the system may divide a note into two shorter notes of different pitches.

These errors appear in the result of a transcription as note insertions/deletions and pitch/duration determination errors.

3.2 Normalization

Normalization is required to match two melodies whose pitch-keys are different. A typical normalization is a transposition of a melody to a certain key. However, this is not effective when the scale notes of a clue melody are altered by the partial errors. To restrain partial error influence, the authors convert the clue melody and every song melody in the database to the sequences of pitch differences between every two adjacent notes.

The following matching section uses those pitch interval sequences to calculate similarity between two melodies.

3.3 Simple DP algorithm

With the normalization described above, an error in the original humming is transcribed either to a note in insertions/deletions and pitch/duration differences.

The basic idea to define the melody similarity are as follows.

1. Melody data is represented as a one-dimensional pattern, whose minimum unit is a note.
2. The melody similarity is basically a sum of note matching scores between related notes of two melodies. The melody similarity is decreased according to note insertion/deletion, which disorders the note relations.

To find the best note relation, between two sequences with some note insertion/deletions, melody matching is carried out by using a Dynamic Programming (DP) matching technique, which is basically a time axis warping matching method and is thus suitable for matching between the normalized pitch sequences.

3. The note matching score is determined according as two pitch intervals between adjacent notes are same or in semitone difference or different over semitone.

The definition above is referred to as the “simple DP”.

3.4 Modified DP algorithm

Modification on the simple DP uses the note duration information additionally to improve the retrieval accuracy.

1. The melody similarity is defined as weighted sum of each note matching score, and note insertion/deletion penalty. The weight for note matching score is defined by the note duration.

2. Each note matching score is defined as the sum of pitch interval matching score and note duration matching score.

This similarity definition is referred to as the “modified DP”.

2P.04 350 ICMC Proceedings 1993
4 Prototype system

To make a test environment of this retrieval method, a prototype system has been developed on a personal computer (80286, 12MHz). The transcription module is based on a personal computer music system [Tsuruta et al., 1988]. A database, which is the searching target is built from 500 songs from five genres. Five genres are Japanese folk tunes, modern folk tunes, Japanese pop, western pop and Japanese rock. Each genre includes 100 songs.

All of the melody data are indexed in advance at points where a hummed clue melody could be started. This kind of indexing is useful in the case when a user only remembers a part of the theme melody, starting from somewhere in the middle of the desired song. The authors used two rules, which are based on subjective observation regarding human singing.

- Index at the beginning of the song
- Index at the next 4/4 after rest symbol

To retrieve song information on this system, a user hums a melody fragment at his/her own pitch/tempo, and the system answers the most similar 10 song titles as retrieved results in 20 seconds on an average.

5 Evaluation

To check the validity of the proposed retrieval method and proposed similarity definitions, the authors executed tests under the following conditions:

- The similarity definitions, Simple DP and Modified DP, are tested. SEQ-NCS and BP-NCS definitions, described below, are also tested for comparison.

- SEQ-NCS: Sequentially matching.
- No Consideration of Semitone pitch error
- No Consideration of Semitone pitch error
- Simple DP: DP matching, considering semitone pitch error
- Modified DP: DP matching, considering semitone pitch error, duration weighted

The prototype melody database, previously mentioned, is used. A song in the database can be identified with four exact notes on an average, or twelve exact notes in the worst case.

Clue melodies to test the performance are 100 hummed clue samples by 10 musically untrained persons. Each clue has 13.5 notes on an average.

Parameters to define each similarity algorithm are determined by using other clue samples.

Table 1: Examination Results (in 100 samples)

<table>
<thead>
<tr>
<th></th>
<th>Retrieved as Top 1</th>
<th>Retrieved in Top 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEQ-NCS</td>
<td>31</td>
<td>45</td>
</tr>
<tr>
<td>DP-NCS</td>
<td>40</td>
<td>63</td>
</tr>
<tr>
<td>Simple DP</td>
<td>53</td>
<td>66</td>
</tr>
<tr>
<td>Modified DP</td>
<td>65</td>
<td>81</td>
</tr>
</tbody>
</table>

By test runs (Table 1), the modified DP algorithm gives the best result. The result shows the validity of the proposed melody retrieval method.

6 Conclusion

In this paper, the authors described a melody retrieval method that takes a hummed melody as a search clue for information retrieval from databases. The system using the proposed method has an automatic transcription module, a melody matching module, and a melody database. By using an automatic transcription module, the system offers a user interface which is more friendly than those previously developed. Melody similarity definitions for melody matching are designed to have ease of handling capabilities.

As a result of the evaluation on a PC-based prototype system, the validity of the proposed melody retrieval method has been shown.

Acknowledgements

The authors would like to thank all testers who hummed melodies. The authors would also like to thank colleagues in NCS Corporation for useful discussions and suggestion.

References

