A QBSH SYSTEM BASED ON THREE-LEVEL MELODY REPRESENTATION

Xiao Wu and Ming Li
ThinkIT Speech Lab., Institute of Acoustics,
Chinese Academy of Sciences
{xwu,mli}@hccl.ioa.ac.cn

ABSTRACT

This extended abstract briefly describes the ThinkIT Speech Lab's submission to the query-by-singing/humming task of MIREX 2007. Our system adopts a three-level framework for retrieving the target song. In such framework, different searching stages employ different melody representations, including acoustic representation for melody scoring, symbolic representation for melody filtering and sentence representation for melody indexing.

1 TASK DESCRIPTION

The goal of query-by-singing/humming (QBSH) task is to evaluate the MIR systems which retrieve songs through human humming/singing. Two subtasks are proposed for the evaluation[4]. The first subtask is a classic QBSH evaluation, which is exactly the same as last year' subtask1. The second task, however, encourages the participants to submit separate algorithm modules instead of integrated one so that various combinations of transcription and matching could be evaluated. In this subtask, the acoustic approaches (which are pitch-based) and symbolic approaches (which are note-based) are also discriminated and compared. Mean reciprocal rank (MRR) of the ground truth is calculated over the top 20 candidates returned by the matchers.

As evaluation data, the first subtask adopts Jang's collection which consists of 2797 queries from 48 ground-truth MIDIs [5]. The second task adopts ThinkIT collection with 355 sung queries from 107 MIDIs [6]. Both tasks employ 2000 Essen MIDIs as noise data.

2 SYSTEM DESCRIPTION

Our system is improved from our submission of MIREX 2006 [3], and adopts the three level framework described in [2]. At acoustic level, melodies are treated as time series. The employed similarity measurement called recursive alignment (RA) features the top-down searching strategy with a predefined heuristic rule. Compared with other algorithms, it puts more stress upon the similarity of overall rhythm and overall contour, which corresponds

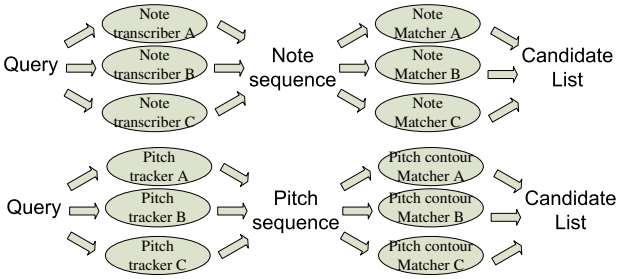

Figure 1. Combination of transcribers and matchers.

with the nature of the melody query. At symbolic level, note-based version of RA (SRA) acts as filter, which takes efficient symbolic representation at the the same time reserves RA's top-down fashion. At sentence level, we segment melodies and obtain the contour tendency and the phrase information for efficient melody indexing. These sentence information, which is largely neglected by previous QBH research, is thought to be robust against query errors as well as transcription errors, and is also complementary to the lower level features. As opposed to relying on single index term, we employ the trigram voting strategy so that multiple index anchors could be involved. Besides, considering the imperfectness of human singing, we introduce fuzzy technique to tolerate query errors. The details of these algorithms are described in our previous and upcoming publications. find in [1, 2, 3].

The above framework benefits from following advantages. First, both representation efficiency and representation accuracy are considered. Symbolic approaches are

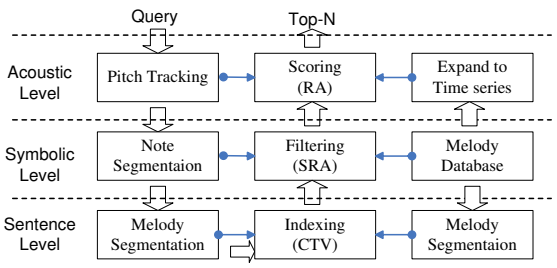

Figure 2. System Framework.

SYS	AU1(S)	AU2(S)	AU3(S)	CG(S)	NM(S)
MRR	0.240	0.093	0.110	0.477	0.576
SYS	FH(S)	RJ1(A)	RJ2(A)	XW(S)	XW(A)
MRR	0.355	0.704	0.872	**0.909**	**0.925**

Table 1. Results of subtask1 (MRR. "A" represents acoustic approach, and "S" represents symbolic approach.).

	CG	FH	NM	XW
XW	0.715	0.452	0.618	**0.917**

Table 2. Results of subtask2: symbolic approaches. The cross of row A and column B represents the combination of transcriber A and matcher B.

	RJ1	RJ2	XW2
RJ	0.345	0.536	**0.883**
XW	0.305	0.536	**0.937**

Table 3. Results of subtask2: acoustic approaches. The cross of row A and column B represents the combination of transcriber A and matcher B.

efficient for small problem size but may suffer from the error-prone audio-to-symbol conversion. On the other side, acoustic representation is accurate but leads to slow matching. Second, different level information can be integrated to rank candidate melodies. Higher level features such as melody tendency and breathing patterns reflect long distance information, which are complementary to the lower level representation.

Our submission involves four algorithm modules: a pitch tracker, a note transcriber, a pitch matcher and a note matcher. The pitch tracking and note segmentation algorithms are exactly same as our last year's submission. The note matcher employs SRA in the scoring stage. The pitch matcher follows the framework in Fig.2.

3 RESULTS

The results of the first subtask are presented in Table 1. It should be noticed that although our note matcher is based on symbolic representation, it works in a very similar way as the acoustic approaches[1, 2]. So, from these results we can conclude that the acoustic approaches generally perform better than the symbolic approaches. Our two algorithms give the best results, which indicate the validity of our proposed framework.

For the second subtask. It is regrettable that many participants do not provide transcription modules, so there is not many combinations (Table 2 and Table 3). In Table 3 we see interesting results, that is, the evaluated pitch matchers favor the suited pitch trackers.

4 ACKNOWLEDGEMENT

We are grateful to IMIRSEL's efforts for their great effort in organization and evaluation of the task. Thanks to Roger Jang for his contribution of test database. Also thanks to Rainer Typke for his wonderful idea of the module combination.

This work is partially supported by MOST (973 program2004CB318106), National Natural Science Foundation of China (10574140, 60535030), The National High Technology Research and Development Program of China (863 program, 2006AA0101022006AA01Z195).

5 REFERENCES

[1] X. Wu, M. Li, J. Liu, J. Yang, and Y. Yan, "A Top-down Approach to Melody Match in Pitch Contour for Query by Humming," Proceedings of International Symposium on Chinese Spoken Language Processing, pp.669–680, 2006.

[2] X. Wu, M. Li, J. Liu, and Y. Yan, "A three level framework for query by humming", unpublished.

[3] X. Wu, M. Li, "QBSH system for MIREX 2006", Extended abstract of MIREX06, 2006.

[4] http://www.music-ir.org/mirex2007/index.php/Query_by_Singing/Humming

[5] http://neural.cs.nthu.edu.tw/jang2/dataSet/childSong4public/QBSH-corpus/

[6] http://hccl.ioa.ac.cn/en/Thinkit.QBH.corpus.rar