Melodic Resolution in Music Retrieval

Timo Sorsa and Jyri Huopaniemi
Nokia Research Center,
Speech and Audio Systems Laboratory
P.O.Box 407, FIN-00045
Nokia Group, Finland

{timo.sorsa,
jyri.huopaniemi}@nokia.com

ABSTRACT
This poster considers the use of different levels of melodic resolution in acoustically driven music retrieval systems from the viewpoint of search-key lengths. A query-by-humming application was constructed to evaluate the dependencies between the melodic resolution, database size and the search-key length in order to consider the optimal level of melodic resolution in music retrieval applications.

1. INTRODUCTION
Acoustically driven retrieval systems (often referred to as query-by-humming applications[1]-[3]) are a recent approach for efficient and flexible music retrieval. These acoustically driven music retrieval systems use a hummed, whistled or played sample of a melody as a search-key to search matching database entries from a music database. Current efforts in standardization such as MPEG-7 [4] are a clear indication of the research and commercial interest on the topic.

In the general case, when the user generates the input to a melody retrieval process by humming, whistling or playing an instrument, the input is noisy. Noise, meaning errors in the input melody in relation to the database entries, affects the accuracy of the process. To overcome this some approximation can be introduced to the retrieval process.

It is customary to use two methods to introduce approximation to query-by-humming applications. First, approximate string matching algorithms are used. Second, different levels of melodic resolution are used. By lowering the resolution, i.e. using fewer intervals to represent melody-lines, the system can eliminate some of the interval errors included in the input signals. But it is a trade-off; the lower the resolution is the lower is the disparity between the database entries. In this poster the emphasis is on the use of different levels of melodic resolution.

2. SYSTEM ARCHITECTURE
A melody retrieval application was constructed for the purpose of evaluating the concept of query-by-humming in general and to test the effects of using different levels of melodic resolution in the retrieval process. In the following the two main functional blocks of the system are presented.

2.1 Acoustical front-end
The acoustical front-end transcribes the input melody into an inner representation (IRP)[5]. The transcribed melody is used as a search-key by the database engine.

The input signal is segmented into notes with amplitude-based segmentation. A normalized signal level is compared to two constant thresholds. A higher threshold is used for detecting note onsets and a lower one for detecting note offsets.

The fundamental frequencies of the frames within segmented notes are determined with an autocorrelation-based pitch tracker [6]. On a note level the pitch is estimated as the median of the pitch values of the frames within a note. The presented retrieval system does not use rhythmic information and therefore note duration is not detected.

The acoustical front-end is designed to accept input generated by humming, whistling and playing an acoustic instrument. The system also has an option to take typed search-keys as inputs. The user can type the names of the notes of the melody in question and that pattern is then used as a search-key.

2.2 Database engine
The core of the database engine has been implemented at the Department of Computer Science at the University of Helsinki. (See [7] for details.) A fast bit-parallel dynamic programming algorithm by Myers is used for approximate string matching [8]. Transcription invariance is assured by the use of intervals in the matching process.

3. MELODIC RESOLUTION
Melodic resolution refers to the accuracy of the representation of the melody-lines. Essentially different levels of melodic resolution are achieved by using different number of intervals to represent the melody.

The use of lower melodic resolution is motivated by the approximation that it offers for the user input. When using lower resolution the intervals in the input melody do not have to be as accurate as with higher resolution. On the other hand the use of lower resolution forces the user to use longer search-keys for successful retrieval.

In the developed system five different levels of melodic resolution can be used in the matching process. The highest resolution presents the melody with semitone accuracy using 25 intervals (12 up and down and the prime). The lowest resolution is the so-called sign-based contour representation that represents the melody with only three intervals (see for example [9]).

3.1 Required Length of the Search-key
From the users point of view the search-key length is a relevant parameter when considering the experienced quality of a retrieval system.

For the purpose of studying the effect of using different melodic resolutions the system was tested with correct typed inputs. That is, for every search-key used, there was an exact match in the database. The melody lines of 13 different tunes were used as search-keys. Typed input was used in this test so that the accuracy of the audio analysis would not have an effect on the test results.

The tests showed that on average, in order to find a unique match, the sign-based contour representation (3 intervals) requires about 1.7 times longer search-keys than the semitone resolution.

In *Figure 1* the relationship between the database size and the required search-key length is presented. *Figure 1* shows the average required search-key lengths for four different databases with five different levels of melodic resolution.

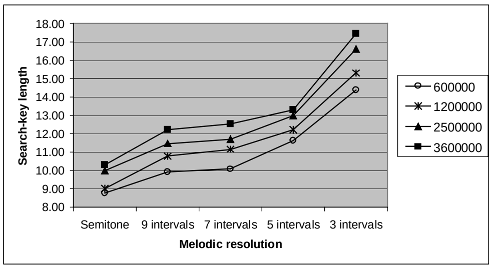

Figure 1. The average search-key lengths required for successful unique retrieval using different levels of melodic resolution in test MIDI databases of 0.6, 1.2, 2.5 and 3.6 million notes with correct search-keys.

The search-key length becomes even more relevant when the errors in the input signals are considered. This was studied with another test. In this test the same search-keys were used as in the previous test but one error was included in each of the keys by removing one random note from the pattern.

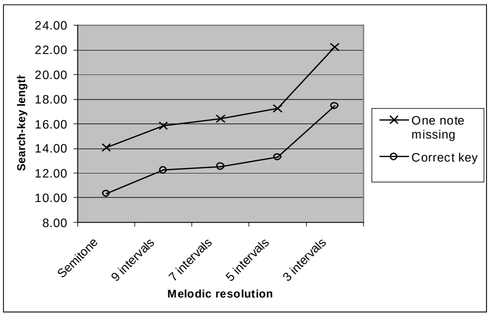

Figure 2. The average search-key lengths required for successful unique retrieval in a test MIDI database of 3.6 million notes. Lower values for correct search-keys and higher values for search-keys with one error (one random note missing).

In *Figure 2* the required search-key lengths for successful unique retrieval in a test database are presented for five different levels of melodic resolution. The search-key lengths are calculated for an optimal case where the search-key is exactly like the corresponding melody-line in the database and for a more realistic case where there is one error in the search-key.

4. CONCLUSIONS

The tests reported above imply that for large databases acoustically driven melody retrieval is not, in the general case, accurate enough. This is mainly due to the errors included in the search-keys generated by the users. This is not to say that query-by-humming type applications are not practical and user-friendly but it implies that the effective size of a large database has to be relatively small if query-by-humming type algorithms are applied.

The tests give an indication of the optimal level of melodic resolution from the point of view of search-key lengths. 9-, 7-, and 5-interval representations are relatively equal from this viewpoint, whereas the 3-interval representation requires significantly longer search keys than the other representations.

The concept of query-by-humming is in many ways a user friendly and efficient method for melody retrieval but the tests indicate that, from the database size point-of-view, it has some, relatively low upper limits beyond which the retrieval process becomes too inaccurate and impractical. When these upper limits are exceeded some other means for classifying the database entries has to be considered. One straightforward method of reducing the effective database size is the use of key words to identify relevant parts of the database in which the actual melody based retrieval process is executed.

5. REFERENCES

[1] McNab, R.J., Smith, L.A., Bainbridge, D., and Witten, I.H., "The New Zealand digital library MELody inDEX", *D-Lib Magazine, May 1997.*

[2] Sonoda, T. and Muraoka, Y., "A WWW-based melody-retrieval system- an indexing method for a large melody database", *Proceedings of ICMC 2000, pp. 170-173, Berlin, Germany, 2000.*

[3] Ghias, A., Logan, J., Chamberlin, D., and Smith, B.C., "Query by humming – musical information retrieval in an audio database", *ACM Multimedia '95, San Francisco, USA, 1995.*

[4] MPEG Requirements Group, "Overview of the MPEG-7 Standard", *Doc. ISO/MPEG N4031, MPEG Singapore Meeting, March 2001 edited by Martinez, J.M.,* http://www.cselt.it/mpeg/standards/mpeg-7/mpeg-7.htm.

[5] Lemström, K. and Laine, P., "Musical information retrieval using musical parameters" *Proceedings of ICMC'98, pp. 341-348, Michigan, USA, 1998.*

[6] Rabiner, L.R., Cheng, M.J., Rosenberg, A.E. and McGonegal C.A., "A comparative performance study of several pitch detection algorithms", *IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 24, no. 5, pp. 399-418, July 1976.*

[7] Lemström, K. and Perttu, S., "SEMEX - An efficient music retrieval prototype", *Proceedings of Music IR 2000, Plymouth, MA, USA, Oct. 2000.*

[8] Myers, G., "A fast bit-vector algorithm for approximate string matching based on dynamic programming", *Proceedings of the 9th Annual Symposium on Combinatorial Pattern Matching, Piscataway, USA, 1998.*

[9] Dowling, W.J., "Scale and contour: two components of a theory of memory and melodies", *Psychological Review, Vol. 85 no.4 pp. 341-354, 1978.*