International Journal of Computer Theory and Engineering, Vol. 2, No. 4, August, 2010
1793-8201

An Improved Melody Contour Feature Extraction for Query by Humming

Nattha Phiwma and Parinya Sanguansat

Abstract—**In this paper, we propose a new melody contour extraction technique and new normalization methods to improve Query-by-Humming. A critical issue of humming sound is noise interference from both environment and acquisition instruments. Furthermore, most users are not professional singers therefore they cause the other query problems about variation of pitch and timing. Advantage of the proposed technique can reduce noise whereas makes pitch smoothing. Our technique consists of four steps as follows: Firstly, the melody contour is extracted from humming sound by Subharmonic-to-Harmonic Ratio (SHR).Subsequently, the melody contour is filtered and smoothed by median filter and our propose technique. Afterwards, we used various normalization methods, including our new techniques, for scaling and noise robust. Finally, humming sound and melody sequences are different alignment methods such as Dynamic Time Warping (DTW), linear interpolations and nonlinear interpolations, before classification. Our technique offers several advantages: higher accuracy, lower complexity, faster query process and lower memory. In addition, the experimental results show that our proposed technique can perform more effective than other methods.**

Index Terms—**Query-by-Humming; melody contour; Dynamic Time Warping; pitch; Subharmonic-to-Harmonic Ratio**

I. Introduction

At present, the music is became part of our lives of most people both listening and singing for entertain and relax. They favor a new kind of entertainment in music which is called Karaoke. The prevalent of problem is users forget the name of the song, but they want to find a song for singing. However, users can retrieve song by only one way, which the user must type keywords (titles, singers, etc.). This search tool is not nsufficient and inconvenient for the user to retrieve the song. Nowadays, this system is known as a Query-by-Humming (QbH) system, which allows users to retrieve a song via simply humming a part of the song. QbH is especially active area of research in the MIR system. Normally, the user always remembers the melody or rhythm and can hum a part of the melody of the song into a microphone and let a QBH system to retrieve the song. Then QbH system will show the result by different names of songs, which users will find it easy and convenient. Outcome presented a list of song ordered by the similarity between humming sound and song in database. This can be used to return to the user a list of songs the system thinks they are humming, ordered by how likely that are the be the desired song. The QbH increase the usability of a music retrieval system meanwhile the user receive convenient and satisfy. Many researchers have focused on how to improve QbH for measuring similarity of humming sound. in particular, methods for detecting pitch and duration of music can be divided briefly into two categories; the time-domain based and the frequency-domain based. First of all, humming sound must be extracted to pitch by using many methods such as autocorrelation, maximum likelihood cepstrum analysis [1] or Subharmonic-to-Harmonic Ratio (SHR) [2]. Fundamental frequency normalization is necessary, therefore it is normalized by statistical approach. There are three frameworks of QbH, based on feature types: (1) the technique based on string matching [1], [3], [4]; (2) the technique based on continues pitch contour matching [5], [6], [7]; (3) the technique based on spectral [8], [9], [10], [11]. These techniques can be classified according to feature representations, i.e. string sequence, time and frequency, and spectrogram.

The first framework, most previous methods were focused on matching part of song retrieval systems. The technique based on string matching is used method of melody and song retrieval from a music database. As Dynamic Time Warping (DTW) can be used for measuring sound signals, it allows local flexibility in aligning time series [12], [13]. Pitch contour was used to represent music melodies. Probably the most prevalent method [1], [3], [4] of melodic representation in QbH systems, the three alphabets were used to display whether a note in sequence is up (U), down (D), or the same (S) as the previous note. But the pitch information alone is not enough to represent the melody. Then melodic representation will be analyzed by above technique. N-grams is another approach, which is widely used in text retrieval and applied to retrieve songs in music system [4], [14], [15], [16]. It is particularly effective for short queries and manual queries not for automatic queries [17]. In [14] considered the use of above method as a front end in a two-stage search in which a fast indexing algorithm based on n-grams narrows the search. In addition, string matching based on statistical models including Hidden Markov Models (HMMs) in [14], [18], [19]. This approach uses a combination of HMMs for sequence estimation and DTW for hierarchical clustering [20].

Subsequent to this technique is continuous pitch contour. From the above techniques, the discriminant information may be of lost and the changing of sounds is not different. We can look to probabilistic models being used in speech recognition and production as possible inspiration. Melody

Nattha Phiwma and Parinya Sanguansat are with the department of Information Technology Rangsit University, Pathumthani, Thailand (email:phewma@hotmail.com, sanguansat@yahoo.com).

contour or pitch contour used in [5], [6], [7], which is a time series of pitch values, represents melody content without using explicit music notes. In [5] present an approach of doing melody retrieval based on a continuous melody contour representation and created a melody alignment method and a new melody similarity metric for melody contour matching. This technique separates the melody alignment and melody similarity measure, difference the dynamic programming string matching methods which do it at the same time. A time series matching approach proposed in [6], [7], [21] has shown effectiveness for QbH in terms of robustness against note errors, since accurate note segmentation is not needed. The method above is based on time and frequency domain analysis which cannot be processed at the same time.

To the best of two domains, there is a technique that both domains possibly work together. According to time and frequency domain analysis, spectral features is the technique that we have classified. In some works, a feature extraction method of the sound recognition framework is used spectrum via spectral basis functions [8], [9], [10], [11]. In [22], compare the performance of spectrogram and a new variation of multiwindow (MW) spectrogram for various digital modulated signals. Spectrogram has been widely used as one of the method for time-varying spectral analysis which is important in many applications such as radar, sonar, speech, geophysics and biological signals [23]. In [10], present a new spectralbased approach to apply QBH efficiently on MP3 solo songs based on vocal part and this approach is to extract the feature descriptors from frequency spectral information from the data streams.

Pitch and fundamental frequency are important feature therefore it must be extracted pitch. A pitch determination algorithm (PDA) based on Subharmonic-to-Harmonic Ratio (SHR) is developed in the frequency domain and describe the amplitude ratio between subharmonics and harmonics [2]. In addition, pitch determination, SHR can be also used as a parameter for describing voice quality. For our system, we have implemented pitch tracking using SHR.

Median filter is well known for being able to remove impulse noise and smoothing signal [24], [25]. In [26] described desirable signal properties for signals used in it which if the real signal has added noise, then it may or may not be possible to remove the noise by filtering. It show how some types of noise can be removed the noised by median filtering and how other types cannot be removed. Median filter is adopted to generate smoother pitch sequence and it is used for smoothing pitch in QbH system [27]. Therefore, our system we decided to reduce noise a part of pitch by it.

Due to the variation of frequency rank, the normalization is needed to apply for reducing these influence. In [28], the fundamental frequency (F0) normalization methods are presented by statistical approach (min, max, mean, standard derivation, etc.). Furthermore, we proposed two new normalization techniques and compare with other normalization methods in [28].

In this paper, we found that appropriate process is as follow: Firstly, pitch tracking by SHR and then our proposed technique for feature extraction and normalization. Finally, DTW is used for signal alignment. The experimental results of this process achieve the highest accuracy, compare to other benchmarks.

This paper is organized as follows: Describing the concept of pitch tracking in Section II and Dynamic Time Warping in Section III. Melody Contour Extraction technique is proposed in Section IV. Pitch Normalization methods are presented, including our new techniques, in Section V. In Section VI, experimental results are presented. Finally, conclusion is in Section VII.

II. Pitch Tracking

In this section, the concept of pitch tracking is described how the system is converted into a sequence of relative pitch transitions. The concept of pitch is the fundamental frequency that matches what note we hear [1]. Notes can begin and end when pitches have been identified. The pitch detector decides based on the statistical information of pitch models. The detailed of each component of the pitch detector is given below.

Four pitch tracking methods: Autocorrelation, Maximum Likelihood, Cepstrum Analysis and SHR [1], [2]. The most of pitch detection autocorrelation is chosen for implementation pitch tracking [1]. In addition, a pitch determination algorithm (PDA) based on Subharmonic-to-Harmonic Ratio (SHR) is developed in the frequency domain and describe the amplitude ratio between subharmonics and harmonics [2], [29]. For our system, we have implemented pitch tracking using SHR. For each short-term signal, let $A(f)$ represents the amplitude spectrum, and let f_0 and f_{max} be the fundamental frequency and the maximum frequency of $A(f)$, respectively. Then the sum of harmonic amplitude is defined as

$$SH = \sum_{n=1}^{N} A\left(nf_0\right), \tag{1}$$

where N is the maximum number of harmonics contained in the spectrum, and $A(f) = 0$ if $f > f_{max}$. If the pitch search range is defined $[F0_{min}; F0_{max}]$, then N=floor(f_{max}=f_{min}) Assuming the lowest subharmonic frequency is one half of f_0, the sum of subharmonic amplitude is defined as

$$SS = \sum_{n=1}^{N} \left(\left(n - 1/2\right) f_0\right). \tag{2}$$

Let $LOGA(^2)$ denote the spectrum with log frequency scale, then we can represent SH and SS as

$$SH = \sum_{n=1}^{N} LOGA\left(\log\left(n\right) + \log\left(f_0\right)\right). \tag{3}$$

$$SS = \sum_{n=1}^{N} LOGA\left(\log\left(n - 1/2\right) + \log\left(f_0\right)\right). \tag{4}$$

To obtain SH, the spectrum is shifted leftward along the logarithmic frequency abscissa at even orders, i.e., log(2), log(4),...log(4N). These shifted spectra are added together and denoted by

$$SUMA(\log f)_{even} = \sum_{n=1}^{2N} LOGA\,(\log f + \log (2n)). \quad (5)$$

Similarly, by shifting the spectrum leftward at log(1), log(3), log(5),...log(4*N*-1), we have

$$SUMA(\log f)_{odd} = \sum_{n=1}^{2N} LOGA\,(\log f + \log (2n-1)). \quad (6)$$

Next, A difference function defines as

$$DA\,(\log f) = SUMA(\log f)_{even} - SUMA(\log f)_{odd} \quad (7)$$

In searching for the maximum value, the position of the global maximum is located and denoted as log (f_1). Then, starting from this point, the position of the next local maximum denoted as log (f_2) is selected in the range of [log (1:9375f_1) ; log (2:0625f_2)]. Equation of SHR is defined as

$$SHR = \frac{DA\,(\log f_1) - DA\,(\log f_2)}{DA\,(\log f_1) + DA\,(\log f_2)}. \quad (8)$$

In case of SHR is less than a certain threshold value, it indicates that subharmonics are weak, so that harmonics are preferred. Thus, f_2 is selected and the final pitch value is 2f_2. Otherwise, f_1 is selected and the pitch is 2f_1. In [2], SHR can be effectively used to pitch tracking.

III. Dynamic Time Warping

Due to the tempo variation of length of sequence, we cannot measure the similarity by any tradition distances. Dynamic Time Warping (DTW) is adopted to fill the gap caused by tempo variation between two sequences. For our system, we use DTW to compute the warping distance between the input melody contour and that of each song in database. Suppose that the input melody contour vector (or query vector) is represented by $t\,(i)$; $i = 1, \ldots, m,$ and the reference vector by $r\,(j)$; $j = 1, \ldots, n$. These two vectors are not necessarily of the same size. The distance in DTW is define as the minimum distance starting from the begin of the DTW table to the current position (*i; j*). According to the dynamic programming algorithm, the DTW table *D*(*i; j*) can be calculated by:

$$D\,(i,j) = d\,(i,j) + \min \begin{cases} D\,(i-2, j-1) \\ D\,(i-1, j-1) \\ D\,(i-1, j-2) \end{cases}, \quad (9)$$

where *D*(*i; j*) is the node cost associated with *t* (*i*) and *r* (*j*) and can be defined from the L1-norm as

$$d\,(i,j) = |t\,(i) - r\,(j)|\,. \quad (10)$$

The best path is the one with the least global distance, which is the sum of cells alone the path. This method exhibits good performance for word speech recognition and QbH in [21].

A warping path *W*, is a contiguous (in the sense stated below) set of matrix elements that defines a mapping between *t* and *r*. The k_{th} element of *W* is defined as $w_k = (i; j)_k$ so we have:

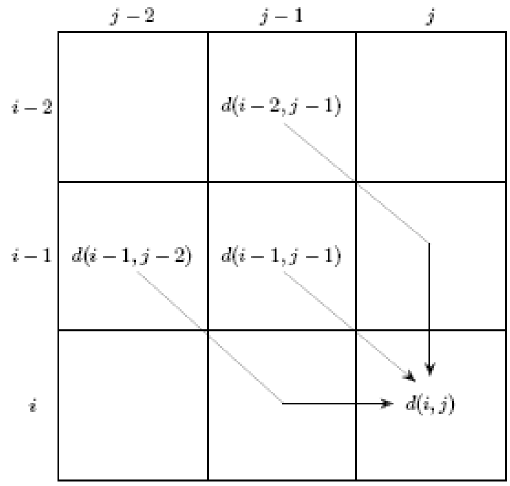

Figure 1. The calculation pattern for the dynamic time warping in the Melody Contour.

$$W = w_1, w_2, ..., w_k, ...w_K \quad (11)$$

where

$$\max\,(m,n) \le K \le m+n-1$$

The warping path is typically subject to several constraints as following [30].

Boundary conditions: $w_1 = (1;\ 1)$ and $w_K = (m;\ n)$ this requires the warping path to start and finish in diagonally opposite corner cells of the matrix.

Continuity: Given $w_k = (a; b)$ then $w_{k-1} = (a', b')$ where $a - d \le 1$ and $b - b' \le 1$. This restricts the allowable steps in the warping path to adjacent cells (including diagonally adjacent cells).

Monotonicity: Given $w_k = (a; b)$ then $w_{k-1} = (a', b')$ where $a - a' \ge 0$ and $b - b' \ge 0$. This forces the points in *W* to be monotonically spaced in time.

IV. Melody Contour Extraction

In this section, our proposed technique for feature extraction in Query-by-Humming (QbH) system is presented. The following algorithm describes how to extract pitch from humming sound to obtain the melody contour.

Let **m** represents melody contour and let **p** be the pitch. The variables of algorithm are describe as follows: *s* is the size of window for filtering, *g* is the gap of pitch difference, *T* is threshold of standard deviation and *v* is variance of pitch interval.

This algorithm was designed for feature extraction. The humming sound consists of pitch in several values and also has noise fused in the pitch as shown in Fig. 2(a). Normally, the humming sound is usually reduced noise by median filtering method which makes the signal is better smooth as shown in the Fig. 2(b). However, it usually makes the discriminant information of the signal be lost at the same time. It is also applied for filtering part of signals prior to further processing with small window. We can reduce noise meanwhile the information of the signal is still reserved by our method.

Algorithm 1 Melody Contour Extraction Algorithm

Require: $\mathbf{p}$, g, T, s
Ensure: $\mathbf{m}$

```
smoothing p by median filter.
initial m_1 ← p_1
N ← length of p
j ← 1
while t ≤ N do
    d = |p_t − p_{t−1}|
    Y ← {y_{t−v}, y_{t−v+1}, ..., y_{t+v−1}, y_{t+v}}
    S_Y ← Standard deviation of Y
    if d > g and S_Y < T then
        m_j ← p_t
    end if
    t ← t + s
    j ← j + 1
end while
return m
```

The first step of this method is taking pitch to pass the process of noise filter which uses the median filter in order to make the signal smooth. Then, find the different value of p by comparing with the defined g value by selecting only the value which different value exceed the g value. The value of s is determined in order to apply to find the range of signal that change a little for a while. In other words we discard the signal that change rapidly in short time comparing with this interval. There is the spread around the signal and we need the group of significant signal only. Hence, we find the range of signal which has a little value of the spread when comparing the threshold of standard deviation (*T*).

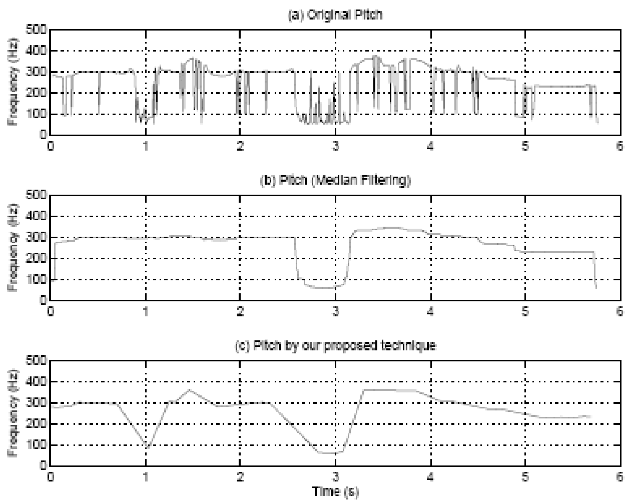

Figure 2. A graph is shown (a) Original Pitch, (b) Pitch (Median Filtering) and (c) Pitch by our proposed technique

From the Fig. 2(c), it can be seen that the pitch which is better smooth. The output of the algorithm melody contour contain significant pitch. Finally, when this technique is applied to retrieval task, it to do retrieval process, the result will be more correct than the traditional method.

V. Pitch Normalization Methods

In continuous speech, pitch contour of humming sound is affected by many factors. Therefore, pitch normalization is necessary. Let p (t) be the pitch and ¾log p(t) represents the standard deviation of logarithm of pitch. In this paper we proposed two new techniques for pitch normalization. For these techniques, logarithm of standard variation are used instead of standard variation of logarithm as shown in (12) besides in (13) logarithm of mean are used instead of mean of logarithm. The following pitch normalization methods are presented:

N1. Using mean and standard deviation value of pitch and normalizing this new value by logarithmic of each sequence.

$$\mathrm{p}(\mathrm{t}) = \frac{\log \mathrm{p}(\mathrm{t}) - \log \overline{\mathrm{p}(\mathrm{t})}}{\log \sigma_{\mathrm{p}(\mathrm{t})}} \tag{12}$$

N2. Using mean of pitch value and normalizing this logarithmic value of pitch by logarithm of each sequence.

$$\mathrm{p}(\mathrm{t}) = \frac{\log \mathrm{p}(\mathrm{t})}{\log \overline{\mathrm{p}(\mathrm{t})}} \tag{13}$$

N3. Using logarithm of pitch value and normalizing this logarithmic value of pitch by min and max of each sequence.

$$\mathrm{p}(\mathrm{t}) = \frac{\log \mathrm{p}(\mathrm{t}) - \min \log \mathrm{p}(\mathrm{t})}{\max \log \mathrm{p}(\mathrm{t}) - \min \log \mathrm{p}(\mathrm{t})} \tag{14}$$

N4. Pitch normalization by pitch mean of each sequence.

$$\mathrm{p}(\mathrm{t}) = \frac{\mathrm{p}(\mathrm{t})}{\overline{\mathrm{p}(\mathrm{t})}} \tag{15}$$

N5. Pitch normalization by min pitch and max pitch of each sequence.

$$\mathrm{p}(\mathrm{t}) = \frac{\mathrm{p}(\mathrm{t}) - \min \mathrm{p}(\mathrm{t})}{\max \mathrm{p}(\mathrm{t}) - \min \mathrm{p}(\mathrm{t})} \tag{16}$$

N6. Pitch normalization by mean and standard deviation of the pitch of each sequence.

$$\mathrm{p}(\mathrm{t}) = \frac{\mathrm{p}(\mathrm{t}) - \overline{\mathrm{p}(\mathrm{t})}}{\sigma_{\mathrm{p}(\mathrm{t})}} \tag{17}$$

N7. Using logarithmic value of pitch and normalizing this new value by mean and standard deviation of each sequence.

$$\mathrm{p}(\mathrm{t}) = \frac{\log \mathrm{p}(\mathrm{t}) - \overline{\log \mathrm{p}(\mathrm{t})}}{\sigma_{\log \mathrm{p}(\mathrm{t})}} \tag{18}$$

N8. Using logarithm of pitch value and normalizing this logarithmic value of pitch by mean of each sequence.

$$\mathrm{p}(\mathrm{t}) = \frac{\log \mathrm{p}(\mathrm{t})}{\overline{\log \mathrm{p}(\mathrm{t})}} \tag{19}$$

VI. Experimental Results

Experiments have shown the effectiveness of the system and according to the various conditions. For effectiveness of this system, the measures were setup to explore such as the variation of number of songs in database, normalization techniques, top-n rank and signal alignment techniques.

This section is organized as follows: Describing the dataset in subsection VI-A. The experimental results of variation of normalization are presented in subsection VI-B. Variation of alignment and variation of top-n rankings are presented, in subsection VI-C and VI-D. Finally, variation of feature extraction and denoising is in subsection VI-E.

A. Dataset

Our system, there are 100, 300 and 500 MIDI format songs in the database. The test query is humming sound which consists of tunes hummed with Da Da Da. We used 100 humming sounds from different people to test our system. The recording was done at 8 kHz sampling rate, mono and time duration 10 seconds, starting at the beginning of song. The result is showed that when the number of MIDI in database was smaller, the accuracy rate was higher. We used 100 test humming sound to queries in 100 MIDI songs in database, it has higher accuracy rate than 300 and 500 MIDI songs in database. For the example, Table I has higher accuracy rate than Table II and Table III with similarity alignment method and other tables are same.

B. Variation of normalization

Pitch of humming sounds are normalized by our new normalization techniques in (12) and (13). To compare with the normalized pitch by other methods i.e. N1-N8 normalization. The experimental results show that normalized pitch of each sequence by logarithm, mean and standard derivation gave better result than other methods. From Fig. 3 and Fig.4 show that the retrieval accuracies normalized pitch by N1 and N2 normalization, obtain higher accuracy rate compared with other normalization methods.

C. Variation of alignments

DTW is signal alignment method which is widely used in time series data. For experiment, DTW was used to alignment which the results are showed in Table I-Table III. Instead of using DTW, interpolations are used for signal alignment such as linear interpolation, piecewise cubic hermite interpolation polynomial and cubic spline interpolation. Interpolations are used to compare with DTW because they are simple and low complexity. We examined the alignment with different methods and it showed that DTW was the most effective method when we used our proposed technique with DTW alignment. It has higher accuracy rate than the alignment with linear interpolation and nonlinear interpolation. From Table I - Table III are alignment with DTW, accuracy rate is higher than other tables which alignment by other methods.

D. Variation of top-n rankings

Top-n rate was the rate of queries that retrieved correct music within top-n rank. In this paper, the performance evaluations include three measurements: top-1 rate, top-5 rate, and top-10 rate. In the experiments, top-10 rank has the accuracy rate higher than top-1 and top-5 as shown in Fig. 3-Fig. 10.

E. Variation of feature extraction and denoising

In this experiments, its method was using median filtering, the baseline noise reduction is described in detail [27] for comparing with our proposed technique. In our experiments, we set the values of variables such as s, g, and T to 5, 2, and 5 respectively. For median filter, we found that the optimal size of window is 53 to achieve the highest performance. Our propose technique used DTW for alignment and normalized with our new normalization methods can achieve highest accuracy, more 90% of top-10, as shown in Table I - Table III.

In Fig. 3-Fig. 10 shows the retrieval accuracies that retrieved 100 humming sounds from 500 MIDI songs database by varying the top-n rank from top-1 to top-25. In order to show the advantage of our proposed technique, the accuracy is better than use only median filter to reduce noise. Our new normalization techniques are higher accuracy rate when compare to other normalization techniques. Moreover, our technique can reduce the dimension of feature vector, which contains only the significant information. Thus in our experiments, the query time is faster than the conventional one around ten times.

VII. Conclusion

In this paper, we have proposed a new melody retrieval method by similarity matching of continuous melody contours and new normalization techniques. We have improved the process of feature extraction from various humming inputs. Furthermore, we used our technique for feature extraction and normalized pitch with our new normalization techniques. The experimental results show that the performance of our proposed techniques is better than other methods. Our technique offers several advantages: higher accuracy and low complexity. First of all, it can reduce noise meanwhile the discriminant information is extracted. That makes the accuracy improve as shown in our experimental results. Secondly, the query process is faster and consumes lower memory because the dimension of feature vector is smaller than traditional one.

Acknowledgment

This study is supported by Rangsit Univerity, Suan Dusit Rajabhbat University Foundation and we would like to thank students of Suan Dusit Rajabhbat University for their great help and also all people who fain hummed a lot of tunes for us. additionally, the invaluable recommendation and supervision from the anonymous reviewers are much appreciated.

References

[1] Asif Ghias, Jonathan Logan, David Chamberlin, and Brian C. Smith, "Query by humming: musical information retrieval in an audio database," in *MULTIMEDIA '95: Proceedings of the third ACM international conference on Multimedia*, New York, NY, USA, 1995, pp. 231–236, ACM.

[2] Xuejing Sun, "Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio," in *Proceedings of the IEEE*, 2002, pp. 333–336.

[3] Rodger J. McNab, Lloyd A. Smith, Ian H. Witten, Clare L. Henderson, and Sally Jo Cunningham, "Towards the digital music library: tune retrieval from acoustic input," in *DL '96: Proceedings of the first ACM international conference on Digital libraries*, New York, NY, USA, 1996, pp. 11–18, ACM.

[4] Alexandra Uitdenbogerd and Justin Zobel, "Melodic matching techniques for large music databases," in *MULTIMEDIA '99: Proceedings of the seventh ACM international conference on Multimedia (Part 1)*,New York, NY, USA, 1999, pp. 57–66, ACM.

[5] Yongwei Zhu and Mohan Kankanhalli, "Similarity matching of continuous melody contours for humming querying of melody databases," in *of Melody Databases, International Workshop on Multimedia Signal Processing, USVI*, 2002.
[6] Takuichi Nishimura, J. Xin Zhang, and Hiroki Hashiguchi, "Music signal spotting retrieval by a humming query using start frame feature dependent continuous dynamic programming," in *Continuous Dynamic Programming, Proc. 3 rd International Symposium on Music Information Retrieval*, 2001, pp. 211–218.
[7] Yongwei Zhu, Mohan S. Kankanhalli, and Changsheng Xu, "Pitch tracking and melody slope matching for song retrieval," in *PCM '01: Proceedings of the Second IEEE Pacific Rim Conference on Multimedia*, London, UK, 2001, pp. 530–537, Springer-Verlag.
[8] Jonathan Foote, Matthew L. Cooper, and Unjung Nam, "Audio retrieval by rhythmic similarity," in *ISMIR*, 2002.
[9] J. Foote and S. Uchihashi, "The beat spectrum: A new approach to rhythm analysis," in *Proc. International Conference on Multimedia and Expo 2001.*, 2001.
[10] Xiangyang Xue Leon Fu, "A new spectral-based approach to querybyhumming for mp3 songs database," in *World Academy of Science, Engineering and Technology 4 2005.*, 2005.
[11] John N. Gowdyl Sabri Gurbuz and Zekeriyu Tufekci, "Speech spectrogram based model adaptation for speaker identification," in *Proceedings of the IEEE*, 2000, pp. 110–115.
[12] Ada Wai-chee Fu, Eamonn Keogh, Leo Yung Hang Lau, and Chotirat Ann Ratanamahatana, "Scaling and time warping in time series querying," in *VLDB '05: Proceedings of the 31st international conference on Very large data bases*. 2005, pp. 649–660, VLDB Endowment.
[13] Yunyue Zhu and Dennis Shasha, "Warping indexes with envelope transforms for query by humming," in *SIGMOD '03: Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, New York, NY, USA, 2003, pp. 181–192, ACM.
[14] Roger B. Dannenberg, William P. Birmingham, Bryan Pardo, Ning Hu, Colin Meek, and George Tzanetakis, "A comparative evaluation of search techniques for query-by-humming using the musart testbed," *J. Am. Soc. Inf. Sci. Technol.*, vol. 58, no. 5, pp. 687–701, 2007.
[15] Stephen Downie and Michael Nelson, "Evaluation of a simple and effective music information retrieval method," in *SIGIR '00: Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, 2000, pp. 73–80, ACM.
[16] Yuen-Hsien Tseng, "Content-based retrieval for music collections," in *SIGIR '99: Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, 1999, pp. 176–182, ACM.
[17] Alexandra Uitdenbogerd and Justin Zobel, "Melodic matching techniques for large music databases," in *MULTIMEDIA '99: Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, New York, NY, USA, 1999, pp. 57–66, ACM.
[18] Hsuan-Huei Shih, S.S. Narayanan, and C.-C.J. Kuo, "An hmm-based approach to humming transcription," in *Multimedia and Expo, 2002. ICME '02. Proceedings. 2002 IEEE International Conference on*, 2002, vol. 1, pp. 337–340 vol.1.
[19] Hsuan-Huei Shih, S.S. Narayanan, and C.-C.J. Kuo, "A statistical multidimensional humming transcription using phone level hidden markov models for query by humming systems," in *Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on*, July 2003,\ vol. 1, pp. I–61–4 vol.1.
[20] Jianying Hu, Bonnie Ray, and Lanshan Han, "An interweaved hmm/dtw approach to robust time series clustering," *Pattern Recognition, International Conference on*, vol. 3, pp. 145–148, 2006.
[21] Jyh-Shing Roger Jang and Hong-Ru Lee, "Hierarchical filtering method for content-based music retrieval via acoustic input," in *MULTIMEDIA '01: Proceedings of the ninth ACM international conference on Multimedia*, New York, NY, USA, 2001, pp. 401–410, ACM.
[22] Tan Jo Lynn and A.Z. bin Sha'ameri, "Comparison between the performance of spectrogram and multi-window spectrogram in digital modulated communication signals," in *Telecommunications and Malaysia International Conference on Communications, 2007. ICTMICC 2007. IEEE International Conference on*, May 2007, pp. 97–101.
[23] L. Cohen, "Time-frequency distributions-a review," *Proceedings of the IEEE*, vol. 77, no. 7, pp. 941–981, Jul 1989.
[24] J. Astola, P. Haavisto, and Y. Neuvo, "Vector median filters," *Proceedings of the IEEE*, vol. 78, no. 4, pp. 678–689, Apr 1990.
[25] H.-M. Lin and Jr. Willson, A.N., "Median filters with adaptive length," *Circuits and Systems, IEEE Transactions on*, vol. 35, no. 6, pp. 675–690, Jun 1988.
[26] Jr. Gallagher, N. and G. Wise, "A theoretical analysis of the properties of median filters," *Acoustics, Speech and Signal Processing, IEEE* Transactions *on*, vol. 29, no. 6, pp. 1136–1141, Dec 1981.
[27] Lei Wang, Shen Huang, Sheng Hu, Jiaen Liang, and Bo Xu, "An effective and efficient method for query by humming system based on multi-similarity measurement fusion," in *Audio, Language and Image* Processing*, 2008. ICALIP 2008. International Conference on*, July 2008, pp. 471–475.
[28] Hong Quang Nguyen, P. Nocera, E. Castelli, and T. Van Loan, "Tone recognition of vietnamese continuous speech using hidden markov model," in *Communications and Electronics, 2008. ICCE 2008. Second International Conference on*, June 2008, pp. 235–239.
[29] Xuejing Sun, "A pitch determination algorithm based on subharmonicto- harmonic ratio," in *the 6th International Conference of Spoken* Language *Processing*, 2000, pp. 676–679.
[30] Eamonn Keogh, "Exact indexing of dynamic time warping," in *VLDB '02: Proceedings of the 28th international conference on Very Large Data Bases*. 2002, pp. 406–417, VLDB Endowment.

TABLE I
TEST RESULT OF EXPERIMENT WITH 100 TEST QUERIES AND 100 MIDI SONGS IN DATABASE USING DTW ALIGNMENT.

F0 Normalization	Method	Rank		
		Top-1	*Top-5*	*Top-10*
N1	Proposed technique	76	95	96
	Median Filtering	37	79	91
N2	Proposed technique	78	95	96
	Median Filtering	41	82	88
N3	Proposed technique	60	84	90
	Median Filtering	29	55	66
N4	Proposed technique	71	94	96
	Median Filtering	40	72	82
N5	Proposed technique	61	86	92
	Median Filtering	18	48	67
N6	Proposed technique	77	94	95
	Median Filtering	29	64	78
N7	Proposed technique	70	93	97
	Median Filtering	30	64	81
N8	Proposed technique	80	96	96
	Median Filtering	41	83	90

TABLE II
TEST RESULT OF EXPERIMENT WITH 100 TEST QUERIES AND 300 MIDI SONGS IN DATABASE USING DTW ALIGNMENT.

F0 Normalization	Method	Rank		
		Top-1	*Top-5*	*Top-10*
N1	Proposed technique	69	89	94
	Median Filtering	21	50	67
N2	Proposed technique	73	90	94
	Median Filtering	26	59	69
N3	Proposed technique	54	74	81
	Median Filtering	16	32	49
N4	Proposed technique	66	91	94
	Median Filtering	23	47	62
N5	Proposed technique	48	69	79
	Median Filtering	6	18	25
N6	Proposed technique	56	79	87
	Median Filtering	9	33	50
N7	Proposed technique	57	77	89
	Median Filtering	8	32	48
N8	Proposed technique	71	92	94
	Median Filtering	22	58	71

TABLE III
TEST RESULT OF EXPERIMENT WITH 100 TEST QUERIES AND 500 MIDI SONGS IN DATABASE USING DTW ALIGNMENT.

F0 Normalization	Method	Rank		
		Top-1	*Top-5*	*Top-10*
N1	Proposed technique	68	87	92
	Median Filtering	16	41	55
N2	Proposed technique	70	87	93
	Median Filtering	18	50	61
N3	Proposed technique	45	71	78
	Median Filtering	9	26	37
N4	Proposed technique	66	84	91
	Median Filtering	20	39	52
N5	Proposed technique	45	63	73
	Median Filtering	5	13	22
N6	Proposed technique	51	73	83
	Median Filtering	9	25	35
N7	Proposed technique	52	73	82
	Median Filtering	7	24	38
N8	Proposed technique	70	86	94
	Median Filtering	19	47	62

TABLE IV
TEST RESULT OF EXPERIMENT WITH 100 TEST QUERIES AND 100 MIDI SONGS IN DATABASE USING LINEAR INTERPOLATION ALIGNMENT.

F0 Normalization	Method	Rank		
		Top-1	*Top-5*	*Top-10*
N1	Proposed technique	54	75	82
	Median Filtering	58	77	80
N2	Proposed technique	55	73	82
	Median Filtering	60	77	79
N3	Proposed technique	53	62	72
	Median Filtering	45	55	60
N4	Proposed technique	50	71	82
	Median Filtering	57	74	79
N5	Proposed technique	34	65	77
	Median Filtering	40	69	78
N6	Proposed technique	47	73	81
	Median Filtering	51	72	80
N7	Proposed technique	47	72	82
	Median Filtering	52	74	79
N8	Proposed technique	56	73	83
	Median Filtering	59	77	79

TABLE V
TEST RESULT OF EXPERIMENT WITH 100 TEST QUERIES AND 300 MIDI SONGS IN DATABASE USING LINEAR INTERPOLATION ALIGNMENT.

F0 Normalization	Method	Rank		
		Top-1	*Top-5*	*Top-10*
N1	Proposed technique	37	63	73
	Median Filtering	40	69	74
N2	Proposed technique	43	64	72
	Median Filtering	42	71	75
N3	Proposed technique	40	59	63
	Median Filtering	32	49	54
N4	Proposed technique	40	64	72
	Median Filtering	41	70	72
N5	Proposed technique	23	46	58
	Median Filtering	26	54	67
N6	Proposed technique	36	60	70
	Median Filtering	43	62	74
N7	Proposed technique	35	61	70
	Median Filtering	41	62	71
N8	Proposed technique	43	65	73
	Median Filtering	42	71	75

TABLE VI
TEST RESULT OF EXPERIMENT WITH 100 TEST QUERIES AND 500 MIDI SONGS IN DATABASE USING LINEAR INTERPOLATION ALIGNMENT.

F0 Normalization	Method	Rank		
		Top-1	*Top-5*	*Top-10*
N1	Proposed technique	36	56	69
	Median Filtering	38	66	74
N2	Proposed technique	40	59	68
	Median Filtering	42	68	72
N3	Proposed technique	38	58	61
	Median Filtering	29	46	52
N4	Proposed technique	36	60	69
	Median Filtering	40	68	71
N5	Proposed technique	20	42	53
	Median Filtering	24	47	61
N6	Proposed technique	29	55	66
	Median Filtering	38	56	66
N7	Proposed technique	28	59	66
	Median Filtering	39	56	65
N8	Proposed technique	40	61	68
	Median Filtering	41	70	73

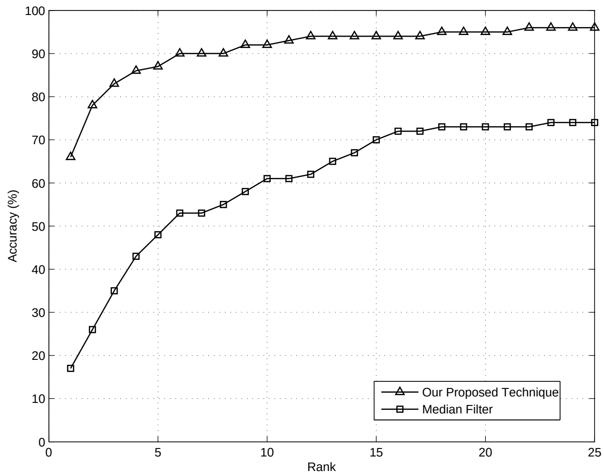

Figure 3. A graph is shown the performance of accuracy rate of our proposed technique and median filter method using N1 Normalization.

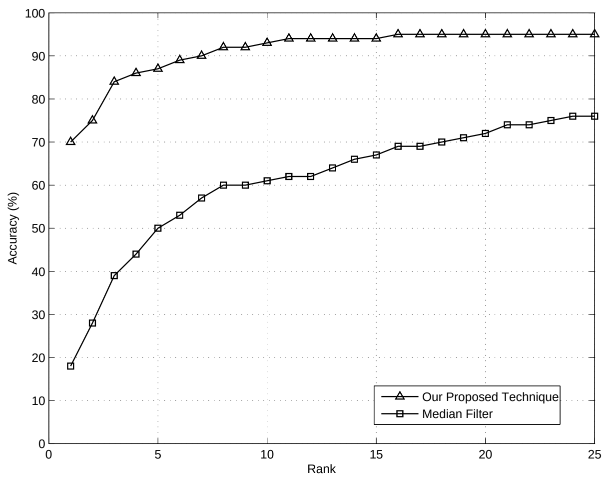

Figure 4. A graph is shown the performance of accuracy rate of our proposed technique and median filter method using N2 Normalization.

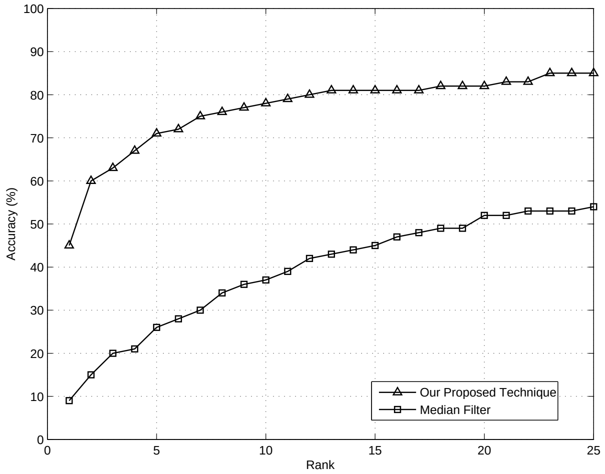

Figure 5. A graph is shown the performance of accuracy rate of our proposed technique and median filter method using N3 Normalization.

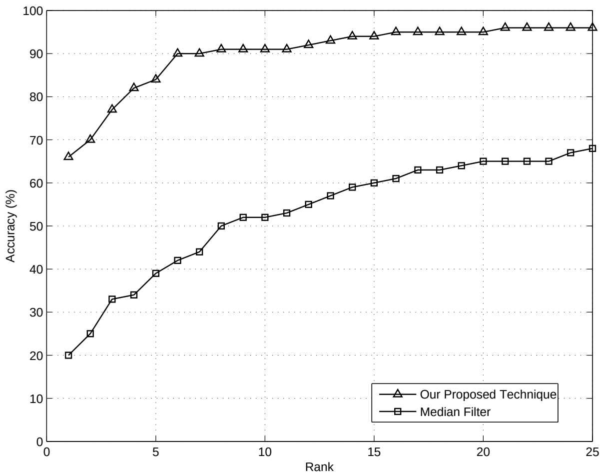

Figure 6. A graph is shown the performance of accuracy rate of our proposed technique and median filter method using N4 Normalization.

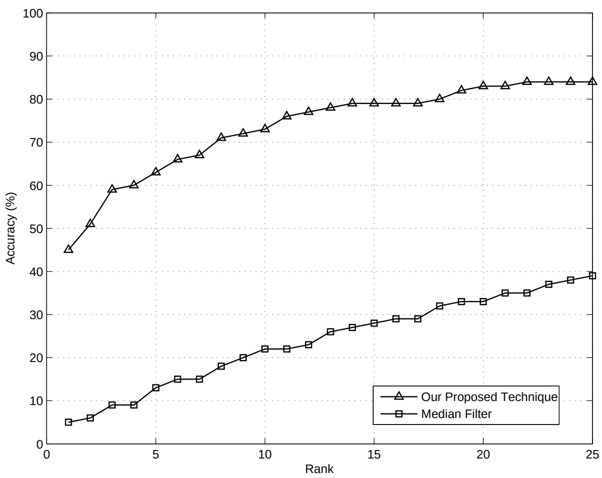

Figure 7. A graph is shown the performance of accuracy rate of our proposed technique and median filter method using N5 Normalization.

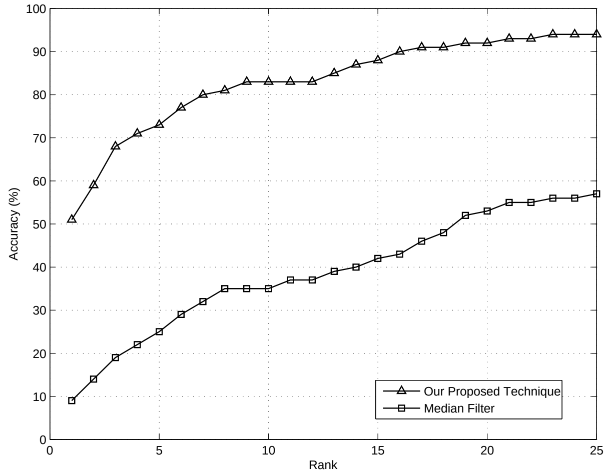

Figure 8. A graph is shown the performance of accuracy rate of our proposed technique and median filter method using N6 Normalization.

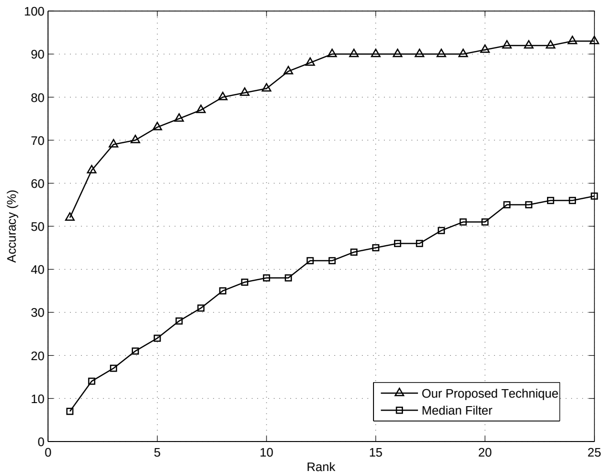

Figure 9. A graph is shown the performance of accuracy rate of our proposed technique and median filter method using N7 Normalization.

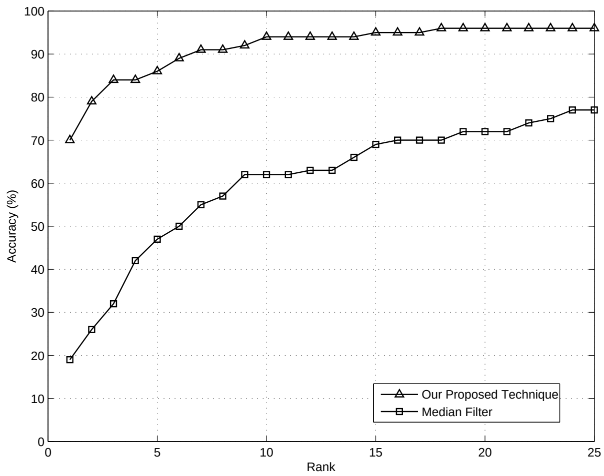

Figure 10. A graph is shown the performance of accuracy rate of our proposed technique and median filter method using N8 Normalization.

TABLE VII
TEST RESULT OF EXPERIMENT WITH 100 TEST QUERIES AND 100 MIDI SONGS IN DATABASE USING PIECEWISE CUBIC HERMITE INTERPOLATION POLYNOMIAL ALIGNMENT.

F0 Normalization	Method	Rank		
		Top-1	*Top-5*	*Top-10*
N1	Proposed technique	54	72	81
	Median Filtering	56	75	78
N2	Proposed technique	53	69	80
	Median Filtering	56	74	78
N3	Proposed technique	51	67	73
	Median Filtering	43	54	62
N4	Proposed technique	52	71	78
	Median Filtering	51	73	78
N5	Proposed technique	33	65	77
	Median Filtering	40	68	77
N6	Proposed technique	44	70	80
	Median Filtering	52	72	77
N7	Proposed technique	47	71	81
	Median Filtering	53	72	77
N8	Proposed technique	54	70	79
	Median Filtering	58	74	78

TABLE VIII
TEST RESULT OF EXPERIMENT WITH 100 TEST QUERIES AND 300 MIDI SONGS IN DATABASE USING PIECEWISE CUBIC HERMITE INTERPOLATION POLYNOMIAL ALIGNMENT.

F0 Normalization	Method	Rank		
		Top-1	*Top-5*	*Top-10*
N1	Proposed technique	39	61	71
	Median Filtering	40	66	74
N2	Proposed technique	39	63	71
	Median Filtering	35	69	73
N3	Proposed technique	40	61	64
	Median Filtering	33	48	53
N4	Proposed technique	39	63	71
	Median Filtering	36	68	72
N5	Proposed technique	23	41	58
	Median Filtering	25	53	65
N6	Proposed technique	33	57	68
	Median Filtering	41	58	71
N7	Proposed technique	34	59	66
	Median Filtering	41	57	72
N8	Proposed technique	39	64	71
	Median Filtering	39	70	74

TABLE IX
TEST RESULT OF EXPERIMENT WITH 100 TEST QUERIES AND 500 MIDI SONGS IN DATABASE USING PIECEWISE CUBIC HERMITE INTERPOLATION POLYNOMIAL ALIGNMENT.

F0 Normalization	Method	Rank		
		Top-1	*Top-5*	*Top-10*
N1	Proposed technique	37	55	66
	Median Filtering	38	66	74
N2	Proposed technique	39	57	66
	Median Filtering	35	66	71
N3	Proposed technique	38	57	61
	Median Filtering	29	46	50
N4	Proposed technique	35	59	65
	Median Filtering	35	65	71
N5	Proposed technique	19	41	49
	Median Filtering	24	48	61
N6	Proposed technique	28	53	66
	Median Filtering	35	56	67
N7	Proposed technique	28	55	65
	Median Filtering	36	55	65
N8	Proposed technique	39	59	66
	Median Filtering	39	67	71

TABLE X
TEST RESULT OF EXPERIMENT WITH 100 TEST QUERIES AND 100 MIDI SONGS IN DATABASE USING CUBIC SPLINE INTERPOLATION ALIGNMENT.

F0 Normalization	Method	Rank		
		Top-1	*Top-5*	*Top-10*
N1	Proposed technique	53	70	78
	Median Filtering	51	73	77
N2	Proposed technique	51	71	77
	Median Filtering	51	73	77
N3	Proposed technique	51	66	72
	Median Filtering	43	53	63
N4	Proposed technique	48	71	75
	Median Filtering	48	71	76
N5	Proposed technique	34	65	75
	Median Filtering	40	67	75
N6	Proposed technique	43	70	77
	Median Filtering	52	70	76
N7	Proposed technique	50	69	78
	Median Filtering	51	70	77
N8	Proposed technique	51	71	77
	Median Filtering	52	73	77

TABLE XI
TEST RESULT OF EXPERIMENT WITH 100 TEST QUERIES AND 300 MIDI SONGS IN DATABASE USING CUBIC SPLINE INTERPOLATION ALIGNMENT.

F0 Normalization	Method	Rank		
		Top-1	*Top-5*	*Top-10*
N1	Proposed technique	37	62	70
	Median Filtering	38	64	73
N2	Proposed technique	39	61	69
	Median Filtering	35	67	72
N3	Proposed technique	38	62	84
	Median Filtering	33	48	52
N4	Proposed technique	33	62	71
	Median Filtering	32	64	69
N5	Proposed technique	17	43	55
	Median Filtering	22	48	63
N6	Proposed technique	31	55	68
	Median Filtering	37	59	71
N7	Proposed technique	34	60	67
	Median Filtering	36	59	70
N8	Proposed technique	38	63	69
	Median Filtering	34	66	73

TABLE XII
TEST RESULT OF EXPERIMENT WITH 100 TEST QUERIES AND 500 MIDI SONGS IN DATABASE USING CUBIC SPLINE INTERPOLATION ALIGNMENT.

F0 Normalization	Method	Rank		
		Top-1	*Top-5*	*Top-10*
N1	Proposed technique	35	56	66
	Median Filtering	36	64	72
N2	Proposed technique	37	54	66
	Median Filtering	34	64	70
N3	Proposed technique	34	58	61
	Median Filtering	29	47	49
N4	Proposed technique	30	56	65
	Median Filtering	31	60	68
N5	Proposed technique	15	40	46
	Median Filtering	20	47	57
N6	Proposed technique	25	52	64
	Median Filtering	31	57	65
N7	Proposed technique	27	56	63
	Median Filtering	33	55	64
N8	Proposed technique	35	55	66
	Median Filtering	34	63	70