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 sufficient 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 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
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 spectral based 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 \( f_{\text{max}} \) be the fundamental frequency and the maximum frequency of \( A(f) \), respectively. Then the sum of harmonic amplitude is defined as

\[
SH = \sum_{n=1}^{N} A(nf_0),
\]

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

\[
SS = \sum_{n=1}^{N} ((n - 1/2) f_0).
\]

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

\[
SH = \sum_{n=1}^{N} LOGA(\log(n) + \log(f_0)).
\]

\[
SS = \sum_{n=1}^{N} LOGA(\log(n - 1/2) + \log(f_0)).
\]

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

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

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

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

Next, a difference function defines as

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

In searching for the maximum value, the position of the global maximum is located and denoted as \(\log (f_i)\). Then, starting from this point, the position of the next local maximum denoted as \(\log (f_j)\) is selected in the range of \([\log (1.9375f_i); \log (2.0625f_i)]\). 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, \(f_j\) is selected and the final pitch value is \(2^{f_j}\). Otherwise, \(f_i\) is selected and the pitch is \(2^{f_i}\). 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 defined 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:

$$D(i, j) = d(i, j) + \min \left\{ D(i - 2, j - 1), D(i - 1, j - 1), D(i - 1, j - 2) \right\},$$ (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

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

The best path is the one with the least global distance, which is the sum of cells along 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 \(r\) and \(t\). The \(k\)th element of \(W\) is defined as \(w_k = (i; f_k)\) so we have:

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

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

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 \(a-d \leq 1\) and \(b-b' \leq 1\). 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 \(a-b \geq 0\) and \(b-b' \geq 0\). 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.

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 ($T$).

![Graphs](image)

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 $\frac{1}{2} \log p(t)$ 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.

\[
p(t) = \frac{\log p(t) - \log \sigma_{p(t)}}{\log \sigma_{p(t)}}
\]

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

\[
p(t) = \frac{\log p(t)}{\log p(t)}
\]

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

\[
p(t) = \frac{\log p(t) - \min \log p(t)}{\max \log p(t) - \min \log p(t)}
\]

N4. Pitch normalization by pitch mean of each sequence.

\[
p(t) = \frac{p(t)}{\bar{p}(t)}
\]

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

\[
p(t) = \frac{\max p(t) - \min p(t)}{\max p(t) - \min p(t)}
\]

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

\[
p(t) = \frac{p(t) - \bar{p}(t)}{\sigma_{p(t)}}
\]

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

\[
p(t) = \frac{\log p(t) - \log \sigma_{p(t)}}{\log \sigma_{p(t)}}
\]

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

\[
p(t) = \frac{\log p(t)}{\log p(t)}
\]
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 University, Suan Dusit Rajabhat University Foundation and we would like to thank students of Suan Dusit Rajabhat 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


### TABLE I

<table>
<thead>
<tr>
<th>Normalization</th>
<th>Method</th>
<th>Top-1</th>
<th>Top-5</th>
<th>Top-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>Proposed technique</td>
<td>76</td>
<td>95</td>
<td>96</td>
</tr>
<tr>
<td>N2</td>
<td>Proposed technique</td>
<td>78</td>
<td>95</td>
<td>96</td>
</tr>
<tr>
<td>N3</td>
<td>Proposed technique</td>
<td>61</td>
<td>84</td>
<td>90</td>
</tr>
<tr>
<td>N4</td>
<td>Proposed technique</td>
<td>71</td>
<td>94</td>
<td>95</td>
</tr>
<tr>
<td>N5</td>
<td>Proposed technique</td>
<td>61</td>
<td>85</td>
<td>92</td>
</tr>
<tr>
<td>N6</td>
<td>Proposed technique</td>
<td>77</td>
<td>94</td>
<td>95</td>
</tr>
<tr>
<td>N7</td>
<td>Proposed technique</td>
<td>70</td>
<td>93</td>
<td>97</td>
</tr>
<tr>
<td>N8</td>
<td>Proposed technique</td>
<td>80</td>
<td>96</td>
<td>96</td>
</tr>
</tbody>
</table>

### TABLE II

<table>
<thead>
<tr>
<th>Normalization</th>
<th>Method</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>Proposed technique</td>
<td>69</td>
</tr>
<tr>
<td>N2</td>
<td>Proposed technique</td>
<td>21</td>
</tr>
<tr>
<td>N3</td>
<td>Proposed technique</td>
<td>54</td>
</tr>
<tr>
<td>N4</td>
<td>Proposed technique</td>
<td>66</td>
</tr>
<tr>
<td>N5</td>
<td>Proposed technique</td>
<td>54</td>
</tr>
<tr>
<td>N6</td>
<td>Proposed technique</td>
<td>56</td>
</tr>
<tr>
<td>N7</td>
<td>Proposed technique</td>
<td>59</td>
</tr>
<tr>
<td>N8</td>
<td>Proposed technique</td>
<td>71</td>
</tr>
</tbody>
</table>
### TABLE III
**Test result of Experiment with 100 test queries and 500 MIDI songs in database using DTW Alignment.**

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

### TABLE IV
**Test result of Experiment with 100 test queries and 100 MIDI songs in database using Linear Interpolation Alignment.**

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

### TABLE V
**Test result of Experiment with 100 test queries and 300 MIDI songs in database using Linear Interpolation Alignment.**

<table>
<thead>
<tr>
<th>F0 Normalization</th>
<th>Method</th>
<th>Rank</th>
<th>Top-1</th>
<th>Top-5</th>
<th>Top-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>Proposed technique</td>
<td></td>
<td>37</td>
<td>63</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td></td>
<td>40</td>
<td>69</td>
<td>74</td>
</tr>
<tr>
<td>N2</td>
<td>Proposed technique</td>
<td></td>
<td>43</td>
<td>64</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td></td>
<td>42</td>
<td>71</td>
<td>75</td>
</tr>
<tr>
<td>N3</td>
<td>Proposed technique</td>
<td></td>
<td>40</td>
<td>59</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td></td>
<td>32</td>
<td>49</td>
<td>54</td>
</tr>
<tr>
<td>N4</td>
<td>Proposed technique</td>
<td></td>
<td>40</td>
<td>64</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td></td>
<td>41</td>
<td>70</td>
<td>72</td>
</tr>
<tr>
<td>N5</td>
<td>Proposed technique</td>
<td></td>
<td>23</td>
<td>46</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td></td>
<td>26</td>
<td>54</td>
<td>61</td>
</tr>
<tr>
<td>N6</td>
<td>Proposed technique</td>
<td></td>
<td>36</td>
<td>60</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td></td>
<td>43</td>
<td>62</td>
<td>74</td>
</tr>
<tr>
<td>N7</td>
<td>Proposed technique</td>
<td></td>
<td>35</td>
<td>61</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td></td>
<td>41</td>
<td>62</td>
<td>71</td>
</tr>
<tr>
<td>N8</td>
<td>Proposed technique</td>
<td></td>
<td>43</td>
<td>65</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td></td>
<td>42</td>
<td>71</td>
<td>75</td>
</tr>
</tbody>
</table>

---

**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.**

<table>
<thead>
<tr>
<th>F0 Normalization</th>
<th>Method</th>
<th>Top-1</th>
<th>Top-5</th>
<th>Top-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>Proposed technique</td>
<td>34</td>
<td>72</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>36</td>
<td>75</td>
<td>78</td>
</tr>
<tr>
<td>N2</td>
<td>Proposed technique</td>
<td>53</td>
<td>69</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>56</td>
<td>74</td>
<td>78</td>
</tr>
<tr>
<td>N3</td>
<td>Proposed technique</td>
<td>51</td>
<td>67</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>43</td>
<td>54</td>
<td>62</td>
</tr>
<tr>
<td>N4</td>
<td>Proposed technique</td>
<td>52</td>
<td>71</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>51</td>
<td>73</td>
<td>78</td>
</tr>
<tr>
<td>N5</td>
<td>Proposed technique</td>
<td>33</td>
<td>65</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>40</td>
<td>68</td>
<td>77</td>
</tr>
<tr>
<td>N6</td>
<td>Proposed technique</td>
<td>44</td>
<td>70</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>52</td>
<td>72</td>
<td>77</td>
</tr>
<tr>
<td>N7</td>
<td>Proposed technique</td>
<td>47</td>
<td>71</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>53</td>
<td>72</td>
<td>77</td>
</tr>
<tr>
<td>N8</td>
<td>Proposed technique</td>
<td>54</td>
<td>70</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>58</td>
<td>74</td>
<td>78</td>
</tr>
</tbody>
</table>

### TABLE VIII

**Test result of Experiment with 100 test queries and 300 MIDI songs in database using Piecewise Cubic Hermite Interpolation Polynomial Alignment.**

<table>
<thead>
<tr>
<th>F0 Normalization</th>
<th>Method</th>
<th>Top-1</th>
<th>Top-5</th>
<th>Top-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>Proposed technique</td>
<td>39</td>
<td>61</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>40</td>
<td>66</td>
<td>74</td>
</tr>
<tr>
<td>N2</td>
<td>Proposed technique</td>
<td>39</td>
<td>63</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>35</td>
<td>69</td>
<td>75</td>
</tr>
<tr>
<td>N3</td>
<td>Proposed technique</td>
<td>40</td>
<td>61</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>33</td>
<td>48</td>
<td>53</td>
</tr>
<tr>
<td>N4</td>
<td>Proposed technique</td>
<td>39</td>
<td>63</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>36</td>
<td>68</td>
<td>72</td>
</tr>
<tr>
<td>N5</td>
<td>Proposed technique</td>
<td>23</td>
<td>41</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>25</td>
<td>53</td>
<td>65</td>
</tr>
<tr>
<td>N6</td>
<td>Proposed technique</td>
<td>33</td>
<td>57</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>41</td>
<td>58</td>
<td>71</td>
</tr>
<tr>
<td>N7</td>
<td>Proposed technique</td>
<td>34</td>
<td>59</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>41</td>
<td>57</td>
<td>72</td>
</tr>
<tr>
<td>N8</td>
<td>Proposed technique</td>
<td>39</td>
<td>64</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>39</td>
<td>70</td>
<td>74</td>
</tr>
</tbody>
</table>

### TABLE IX

**Test result of Experiment with 100 test queries and 500 MIDI songs in database using Piecewise Cubic Hermite Interpolation Polynomial Alignment.**

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

### TABLE X

**Test result of Experiment with 100 test queries and 100 MIDI songs in database using Cubic spline interpolation Alignment.**

<table>
<thead>
<tr>
<th>F0 Normalization</th>
<th>Method</th>
<th>Top-1</th>
<th>Top-5</th>
<th>Top-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>Proposed technique</td>
<td>33</td>
<td>61</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>38</td>
<td>64</td>
<td>75</td>
</tr>
<tr>
<td>N2</td>
<td>Proposed technique</td>
<td>39</td>
<td>61</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>35</td>
<td>67</td>
<td>72</td>
</tr>
<tr>
<td>N3</td>
<td>Proposed technique</td>
<td>38</td>
<td>62</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>33</td>
<td>62</td>
<td>71</td>
</tr>
<tr>
<td>N4</td>
<td>Proposed technique</td>
<td>32</td>
<td>64</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>17</td>
<td>43</td>
<td>55</td>
</tr>
<tr>
<td>N5</td>
<td>Proposed technique</td>
<td>22</td>
<td>48</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>31</td>
<td>53</td>
<td>66</td>
</tr>
<tr>
<td>N6</td>
<td>Proposed technique</td>
<td>34</td>
<td>60</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>36</td>
<td>59</td>
<td>70</td>
</tr>
<tr>
<td>N7</td>
<td>Proposed technique</td>
<td>38</td>
<td>63</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>34</td>
<td>66</td>
<td>75</td>
</tr>
</tbody>
</table>

### TABLE XI

**Test result of Experiment with 100 test queries and 300 MIDI songs in database using Cubic spline interpolation Alignment.**

<table>
<thead>
<tr>
<th>F0 Normalization</th>
<th>Method</th>
<th>Top-1</th>
<th>Top-5</th>
<th>Top-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>Proposed technique</td>
<td>39</td>
<td>61</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>38</td>
<td>64</td>
<td>75</td>
</tr>
<tr>
<td>N2</td>
<td>Proposed technique</td>
<td>39</td>
<td>61</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>35</td>
<td>67</td>
<td>72</td>
</tr>
<tr>
<td>N3</td>
<td>Proposed technique</td>
<td>38</td>
<td>62</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>33</td>
<td>62</td>
<td>71</td>
</tr>
<tr>
<td>N4</td>
<td>Proposed technique</td>
<td>32</td>
<td>64</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>17</td>
<td>43</td>
<td>55</td>
</tr>
<tr>
<td>N5</td>
<td>Proposed technique</td>
<td>22</td>
<td>48</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>31</td>
<td>53</td>
<td>66</td>
</tr>
<tr>
<td>N6</td>
<td>Proposed technique</td>
<td>34</td>
<td>60</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>36</td>
<td>59</td>
<td>70</td>
</tr>
<tr>
<td>N7</td>
<td>Proposed technique</td>
<td>38</td>
<td>63</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>34</td>
<td>66</td>
<td>75</td>
</tr>
</tbody>
</table>

### TABLE XII

**Test result of Experiment with 100 test queries and 500 MIDI songs in database using Cubic spline interpolation Alignment.**

<table>
<thead>
<tr>
<th>F0 Normalization</th>
<th>Method</th>
<th>Top-1</th>
<th>Top-5</th>
<th>Top-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>Proposed technique</td>
<td>35</td>
<td>56</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>36</td>
<td>64</td>
<td>72</td>
</tr>
<tr>
<td>N2</td>
<td>Proposed technique</td>
<td>37</td>
<td>54</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>34</td>
<td>64</td>
<td>70</td>
</tr>
<tr>
<td>N3</td>
<td>Proposed technique</td>
<td>34</td>
<td>58</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>29</td>
<td>47</td>
<td>49</td>
</tr>
<tr>
<td>N4</td>
<td>Proposed technique</td>
<td>30</td>
<td>56</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>31</td>
<td>60</td>
<td>68</td>
</tr>
<tr>
<td>N5</td>
<td>Proposed technique</td>
<td>15</td>
<td>40</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>20</td>
<td>44</td>
<td>57</td>
</tr>
<tr>
<td>N6</td>
<td>Proposed technique</td>
<td>23</td>
<td>52</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>31</td>
<td>57</td>
<td>65</td>
</tr>
<tr>
<td>N7</td>
<td>Proposed technique</td>
<td>27</td>
<td>56</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>33</td>
<td>55</td>
<td>64</td>
</tr>
<tr>
<td>N8</td>
<td>Proposed technique</td>
<td>35</td>
<td>55</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Median Filtering</td>
<td>34</td>
<td>63</td>
<td>70</td>
</tr>
</tbody>
</table>