An Improved Feature Extraction and Combination of Multiple Classifiers for Query-by-Humming

Nattha Phiwma[1] and Parinya Sanguansat[2]
[1]Department of Computer Science, Suan Dusit Rajabhat University, Thailand
[2]Faculty of Engineering and Technology, Panyapiwat Institute of Management, Thailand

Abstract: *In this paper, we propose new methods for feature extraction and soft majority voting to adjust efficiency and accuracy of music retrieval. For our work, the input is humming sound which is sound wave and Musical Instrument Digital Interface (MIDI) is used as the reference song in database. A critical issue of humming sound are variation such as duration, sound, tempo, key, and noise interference from both environment and acquisition instruments. Besides all the problems of humming sound we have mentioned earlier, whether humming sound and MIDI in different domain which will make the difficulty for two domains to compare each other. However, to make these two in the same domain, we convert them into the frequency domain. Our approach starts from pre-processing by using features for note segmentation by humming sound. The process consists of four steps as follows: Firstly, the MIDI is already a sequence of pitch while the pitch in humming sound is needed to extract by Subharmonic-to-Harmonic (SHR). Subsequently, the extracted pitch can be used to calculate all above attributes and then multiple classifiers are applied to classify the multiple subsets of these features. Afterwards, the subset contain the multiple attributes, Multi-Dimensional Dynamic Time Warping (MD-DTW) is used for similarity measurement. Finally, Nearest Neighbours (NN) and soft majority voting are used to obtain the retrieval results in case of equal scores. From the experiments, to achieve 100% accuracy rate at the early top-n rank in retrieving, the appropriate feature set should consist of five classifiers.*

Keywords: *Query-by-Humming, feature extraction, majority voting, multiple classifiers, MD-DTW, SHR.*

Received February 8, 2012; accepted May 22, 2012; published online January 29, 2013

1. Introduction

At present, the music becomes part of our lives both listening and singing to entertain and relax ourselves. The prevalent of problem, most users forget the name of the song, but they want to find a song for listening and singing. However, traditional approaches for retrieving music data were based on the textual information such as titles, composers, file names or singers. Because of their incompleteness, there are many difficulties in satisfying particular requirements of applications.

Therefore, many researchers have proposed techniques to query a song base on humming which is called Query-By-Humming (QBH) system [2, 13, 14, 20, 21, 26]. QBH system allows the user to retrieve an intended song based on humming some part of the song. The general framework of QBH system contains three main components, which are query processing module, melody database and matching engine [2]. Firstly, the system handles the Musical Instrument Digital Interface (MIDI) in database. Subsequently, the system process the users input humming signal then extract signal fundamental frequency, humming query is converted into melody representation. Finally, when a search is initiated, melody representation is used to match against the melody in the feature database, according to their similarities and return a rank list of songs.

Normally, natural sounds are a composition of a fundamental frequency with a set of harmonics. The frequency that the human ear interprets as the pitch of a sound is this fundamental frequency, even if it is absent in the sound. The pitch of natural sounds is important in many contexts. Pitch is the perception of how high or low a musical note sounds, which can be considered as a frequency which corresponds closely to the fundamental frequency or main repetition rate in the signal [15]. It is one of the most important parameters in the voice signal analysis and can be determined by the fundamental frequency of the unit frame [6]. For QBH system, pitch is the key feature of melody. As the humming sound consists of noise, pitch needs to be extracted and in order to get the most significant information.

2. Related Works

For early work, pitch is used in QBH as a feature [2, 8, 14]. There are many techniques to analyse and

extract pitch contour, pitch interval and duration from voice humming query [2]. In general, traditional methods for detecting pitches have been proposed in the past, it can be divided roughly into two domains to identify the pitch: time-domain based, frequency domain based [11, 15].

Pitch and fundamental frequency are important features, 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 describes the amplitude ratio between subharmonics and harmonics [23, 24]. For our system, we have implemented pitch tracking using SHR.

The Mel Frequency Cepstral Coefficients (MFCC) was adopted in many speech analysis applications. This type of feature extraction is being widely used in robust speech recognition systems inspired by human auditory perception and focusing on effective signal processing in the ear using cochlear filterbanks [1]. MFCCs were also used as features [7, 12]. From these experiments it shows that using MFCC with the dimension 13 and audio recognition will give better results than other dimensions. MFCC is used in our pre-processing.

Generally, to gather all attributes to use all at once might not give good result. Some features are appropriate but some are not. However, we need to find many classifiers to help with the result. To improve the performance of the system, there are many researches used a lot of information, such as pitch, duration, rhythm, inter-onset interval, start and end time to be mutually considered and make feature in [10, 14]. The most often used classifiers combination approaches in Multiple Classifiers System (MCS) include classifier selection, the majority voting, the weighted combination (weighted averaging), the probabilistic schemes, various rank-ordered rules and etc., [4]. Besides, MCS and majority vote is applied for off-line Arabic handwriting recognition, the accuracy is higher than individual classifier [9]. Therefore, MCS will be used to find the result and the easiest way to do is majority voting.

However, the feature still has variable length in the form of melody contour, hence the traditional Dynamic Time Warping (DTW) cannot be used for this feature. Multi-Dimensional Dynamic Time Warping time series (MD-DTW) algorithm was proposed for DTW on multi-dimensional time series, which the algorithm utilises all dimensions to find the best synchronization [3]. Multi-dimensional (time) series are series in which multiple measurements are made simultaneously. MD-DTW is applied with image texture [19], gesture recognition [3], time series [25], thus we have an idea to apply this to the QBH.

The segment of a note in the humming waveform is model by a Hidden Markov Model (HMM) while the pitch of the note is model by a pitch model using a Gaussian mixture model. The frame based analysis is performed on a note segment which usually has several frames. Multiple frames of a segmented note are used for pitch model analysis. After applying autocorrelation to those frames, pitch features are extracted. The first stage of the proposed algorithm is note segmentation, where the process of segmenting notes of a humming piece is conducted. First, a feature set which can characterize a note is chosen. Next, the HMM definition is chosen before training. During the training phase, notes' phone level HMMs are trained using the selected feature set. The trained note models are then used by the note decoder for note segmentation. Finally, the duration of a segmented note is label according to its relative duration change [5, 18, 19, 22].

3. Materials and Methods

3.1. Melody Contour Extraction Algorithm

The following algorithm describes how to extract pitch from humming sound to obtain the melody contour. Melody Contour Extraction, we have proposed in [16]. Let m represents melody contour and let p be the pitch. The variables of algorithm are described as follows: s is the size of the window for filtering, g is the gap of pitch difference, T is the threshold of standard deviation, and v is the variance of pitch interval. The Algorithm proceeds as follows:

Require: p, g, T, s
Ensure: m
Step 1: Smoothing p by median filter.
Step 2: Initial $m_1 \leftarrow p_1$
Step 3: $N \leftarrow$ *length of p*
Step 4: $j \leftarrow 1$
Step 5: While $t \leq N$ *do*
$d=|p_t - p_{t-1}|$
$Y \leftarrow \{p_{t-v}, p_{t-v+1}, \ldots, p_{t+v-1}, pt_{+v}\}$
$S_Y \leftarrow$ *Standard deviation of Y*
If $d>g$ *and* $S_Y<T$ *then*
$m_j \leftarrow p_t$
End if
$t \leftarrow t+s$
$j \leftarrow j+1$
End while
Step 6: Return m

The first step of this technique is to take a pitch to pass through the noise filtering process which uses the median filter in order to make the signal go smoothly. Then, find the different value of p by comparing with the defined g value by selecting only the exceed value. The value of s is determined in order to apply to find the range of signal that changes a little at that period of time. In other words, it discards the signal that changes rapidly in a short time comparing with this interval. There is the spread around the signal and it only needs the group of significant signals. Hence, it finds the range of signal which has a small value of the spread when comparing with the threshold of

standard deviation (T) as shown in Figure 1. The output of the algorithm melody contour contains 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.

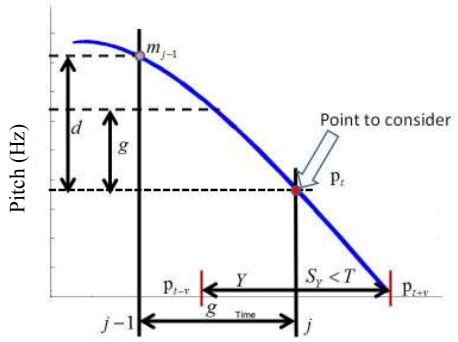

Figure 1. Example of pitch extraction by melody contour extraction.

3.2. Note Segmentation by Humming Sound

For this paper, we have proposed the method of note segmentation by humming sound to differentiate the sounds part from the silence parts in order to choose the most important part, which is the sound part, to use in the next process [17]. From the sound wave in Figure 2, the silence interval is removed manually as pre-processing before being fed to the HMM.

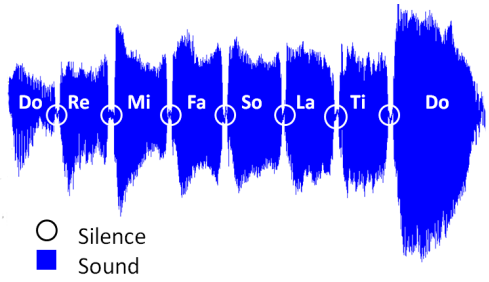

Figure 2. Sound wave from humming standard note in C major scale (do, re, me,...,do).

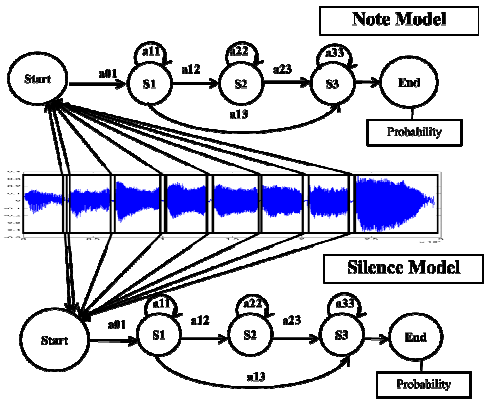

Figure 3. Note model and silence model.

As shown in Figure 3, the HMM contain 3 states with left-to-right topology using 2 Gaussian mixture distributions. Both the note and the silence are used to train these HMMs.

3.3. Multi-Dimensional Dynamic Time Warping

Multi-dimensional series consist of a number of measurements made at each instance. The number of measurements is the dimensionality of the series, the number of time instances its length. Note that multi-dimensional series need not be time signals, any situation in which several measurements are made simultaneously depending on one variable that gives a multidimensional series. They assume that measurements are stored in a matrix, in which columns are features and rows are time instances.

MD-DTW was proposed [3] as an approach to calculate the DTW by synchronizing multi-dimensional series, which is basically an extension of the original DTW, where the matrix D is created by computing the distance between k-dimensional points (where, differently from the original approach, k can be larger than This approach pre-processes the multi-dimensional series, which must have the same number of dimensions. The last step of this algorithm is the execution of the traditional DTW. However, in many cases, all dimensions will contain information needed for synchronisation therefore proposes MD-DTW for synchronising such series. The MD-DTW algorithm runs in 4 steps:

Let A, B be two series of dimension K and length M, N, respectively.
Step 1: Normalize each dimension of A and B separately to a zero mean and unit variance.
Step 2: If desired, smooth each dimension with a Gaussian filter.
Step 3: Fill the M by N distance matrix D according to:

$$D(i,j) = \sum_{k=1}^{K} \left\| A_i(k) - B_j(k) \right\|$$

Step 4: Use this distance matrix to find the best synchronization with the regular DTW algorithm.

Take two series A and B. DTW involves the creation of a matrix in which the distance between every possible combination of time instances $A(i) \leftrightarrow B(j)$ is stored. This distance is calculated in terms of the feature values of the points. Various norms are possible. In 1D-DTW, the distance is usually calculated by taking the absolute or the squared distance between the feature values of each combination of points.

For MD-DTW, a distance measure for two K-dimensional points must be calculated. This distance can be any p-norm. They use the 1-norm, i.e., the sum of the absolute differences in all dimensions. To combine different dimensions in this way, it is necessary to normalize each dimension to a zero mean and unit variance. For this, the dimensions must be comparable. If for instance one dimension contains real valued measurements and one is binary, comparing them directly is not possible and a more sophisticated distance measure must be found [3, 19].

4. Our Approach

In our approach, it consists of two steps which are pre-processing and processing. Pre-processing is process of note segmentation by humming sound. Processing, it consists of feature extraction and soft majority voting.

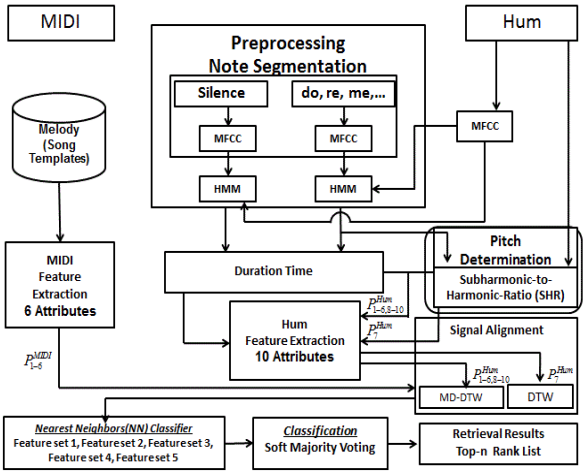

Figure 4. Block diagram of our approach.

We propose two methods in our framework as shown in Figure 4 which are feature extraction sections 4.1, 4.2 and 4.3 and majority voting extraction section 4.4. Our framework starts from pre-processing by using a feature to facilitate note segmentation by a humming sound. The process consists of four steps as follows: Firstly, the MIDI is already a sequence of pitch while the pitch in humming sound is needed to extract by SHR [23, 24]. Consequently, the pitch is extracted by our new feature extraction method and then multiple classifiers are applied to classify the multiple subsets of these features. Afterwards, MD-DTW is used for similarity measurement. Finally, Nearest Neighbors (NN) and soft majority voting are used to obtain the retrieval results in case of equal scores.

4.1. Feature Extraction

In this process, the principle function used for making feature extraction of input (humming sound) and reference songs in database MIDI.

- F 1. Normalized pitch:

$$N_1(p_t) = \frac{\log p_t - \log \overline{p_t}}{\log \sigma_{p_t}} \quad (1)$$

- F 2. Normalized duration of time:

$$N_{time}(T_t) = \frac{T_t}{\sum_t T_t} \quad (2)$$

where t represents note durations in seconds and T is the summation of duration time.

- F 3. Melody contour extraction (Melslope):

$$M(p_t) = melslope(p_t) \quad (3)$$

- F 4. String numeric relative (UDR):

$$udr(p_t) = \begin{cases} 0, |p_t - p_{t-1}| < \varepsilon \\ 1, |p_t - p_{t-1}| > \varepsilon \\ 2, |p_t - p_{t-1}| < -\varepsilon \end{cases} \quad (4)$$

4.2. Feature Extraction of MIDI

Conducting feature extraction of MIDI has four approaches as following:

- Normalized pitch:

$$P_1^{MIDI}(p_t) = N_1(p_t) \quad (5)$$

- Normalized duration of time:

$$P_2^{MIDI}(T_t) = N_{time}(T_t) \quad (6)$$

- Normalized duration of pitch:

$$P_3^{MIDI}(p_t) = N_{time}(p_t) \quad (7)$$

- String numeric relative (UDR):

$$P_4^{MIDI}(p_t) = udr(p_t) \quad (8)$$

4.3. Feature Extraction of Input

Conducting feature extraction of hum has six approaches as following:

- Normalized pitch:

$$P_1^{Hum}(p_t) = N_1(p_t) \quad (9)$$

- Normalized duration of time:

$$P_2^{Hum}(p_t) = N_{time}(T_t) \quad (10)$$

- Normalized duration of pitch:

$$P_3^{Hum}(p_t) = N_{time}(p_t) \quad (11)$$

- String numeric relative (UDR):

$$P_4^{Hum}(p_t) = udr(p_t) \quad (12)$$

- Melody contour extraction (Melslope):

$$P_7^{Hum}(p_t) = M(p_t) \quad (13)$$

- Melslope of pitch pass note segmentation:

$$P_8^{Hum}(pseg_t) = M(pseg_t) \quad (14)$$

where *pseg* represents pitch passed note segmentation. The MIDI is already a sequence of pitch while the pitch in humming sound is needed to extract by SHR [23, 24].

Due to the difference characteristic of MIDI and humming sound, the feature extraction P_1 to P_6 are performed to both of them as P_7 to P_{10} are only used for humming sound. While P_1^{Hum}, P_8^{Hum}, P_9^{Hum}, P_{10}^{Hum}

are compared with P_1^{MIDI}, P_1^{MIDI}, P_5^{MIDI}, P_6^{MIDI}, respectively.

Afterwards, the extracted pitch can be used to calculate all above attributes and then multiple classifiers are applied to classify the multiple subsets of these features. In case of the subset contain the multiple attributes, MD-DTW is used instead of DTW for similarity measurement. Finally, NN and soft majority voting are used to obtain the retrieval results.

4.4. Soft Majority Voting

Majority voting method is widely used in many tasks classification. Voting is a method for a group to make a decision. By the principle of voting, in general, the final decision is based on highest score. Nevertheless, in terms of equal vote, there are many ways of making decision, depending on particular situation. Thus, we propose to make important decisions if the vote is equal, based on the principle of minimum distance, which it is called soft majority voting.

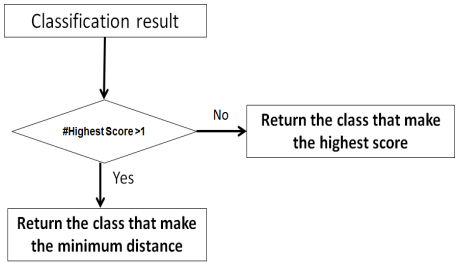

Figure 5. Soft majority voting method.

From Figure 5, if there is the only one highest score, it will return the class that has the highest score as the result. But if there are multiple highest scores, all members will be reconsidered minimum distance by soft majority voting method. The principle of soft majority voting makes all members reconsidered through distance.

5. Results

We have conducted extensive experiments to measure retrieval performance in terms of accuracy. 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, feature extraction, top-n rank and combination of feature.

5.1. Dataset

In this dataset, there are 500 MIDI format songs and they are divided into three subsets which are 100, 300 and 500. We used 100 tests humming sound to query songs in database. The test query is a humming sound which consists of hummed tunes with Da Da Da. We used 100 humming sounds from different people to test our system. The recording was done at 8kHz sampling rate, mono and time duration 10seconds, start at the beginning of song. The results are showed that when the number of MIDI in database is smaller, the accuracy rate is higher.

5.2. Variation of Feature Sets

In this paper, some single attribute are used for creating multiple attributes, such as *P1*, *P2*, *P4* and *P7*, as described in Tables 1 and 2. Experiments have shown the effectiveness of the system and according to the various conditions such as the variations of number of songs in database, feature extraction, top-n rank and combination of feature.

Table 1. List of single attribute.

Name	Single Attribute Description
P1	Normalized Pitch
P2	Normalized Duration Time
P3	Normalized Duration of Pitch
P4	String Numeric Relative (UDR)
P7	Melody Contour Extraction (Melslope)
P8	Melslope of Pitch Pass Note Segmentation

Table 2. List of multiple attributes.

Name	Multiple Attributes Description
P5	P1, P2
P6	P1, P4
P9	P2, P7
P10	P4, P7

Table 3. List of feature.

Feature	Attributes	Feature	Attributes	Feature	Attributes
1	P1	14	P2 P3 P8	27	P1 P3 P6 P8 P9
2	P2	15	P2 P4 P9	28	P1 P4 P6 P8 P10
3	P3	16	P2 P4 P6	29	P1 P6 P8 P9 P10
4	P4	17	P2 P6 P9	30	P2 P3 P4 P6 P9
5	P5	18	P2 P8 P9	31	P2 P5 P6 P8 P10
6	P6	19	P3 P6 P8	32	P2 P4 P8 P9 P10
7	P7	20	P4 P6 P8	33	P3 P4 P5 P6 P10
8	P8	21	P1 P2 P4 P6	34	P4 P5 P6 P8 P10
9	P9	22	P1 P2 P3 P8	35	P3 P4 P8 P9 P10
10	P10	23	P1 P4 P6 P8	36	P2 P4 P5 P6 P8 P10
11	P2 P4	24	P2 P4 P8 P9	37	P3 P4 P5 P6 P9 P10
12	P2 P6	25	P2 P7 P8 P9	38	P1 P2 P3 P4 P5 P6 P8
13	P1 P6 P10	26	P3 P5 P8 P10	39	P2 P3 P4 P6 P8 P9 P10

In this experiment, the number of classifiers was varied from two to ten. The feature sets for each classifier are defined in Table 4. It is fixed as *P7* in every classifiers. Since, our experiment, we have found that combining attribute *P7* with other features can achieve 100% accuracy rate, which faster than using random method. The detail of classifier is used in experimental as shown in Tables 3 and 4. Table 3

shows the attribute of each feature. Table 4 contains a feature set that is used as each classifier.

Table 4. List of classifiers.

# Classifier	Features Set
2	7 20
3	7 25 27
4	7 21 24 37
5	7 14 15 16 28
6	7 17 18 19 29 33
7	7 11 12 13 14 23 33
8	7 11 12 13 14 22 23 33
9	7 11 12 13 14 22 23 32 33
10	7 6 13 26 30 31 35 36 38 39

The performance evaluations vary top-n from top-1 to top-60. The experiments are shown in Tables 5, 6 and 7 for each dataset.

Table 5. Test results of experiment with 100 MIDI songs with variations of feature sets.

Top-n Rate(%)	Classifier								
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	56	67	74	73	71	72	72	71	71
5	91	90	94	96	91	87	91	91	94
10	97	96	97	100	94	96	96	96	97
15	97	97	99	100	97	97	97	98	97
20	97	99	100	100	99	99	99	99	99
25	97	100	100	100	100	100	100	100	99
30	97	100	100	100	100	100	100	100	100
35	98	100	100	100	100	100	100	100	100
40	98	100	100	100	100	100	100	100	100
45	99	100	100	100	100	100	100	100	100
50	99	100	100	100	100	100	100	100	100
55	99	100	100	100	100	100	100	100	100
60	100	100	100	100	100	100	100	100	100

Table 6. Test results of experiment with 300 MIDI songs with variations of feature sets.

Top-n Rate(%)	Classifier								
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	50	63	70	71	67	66	67	68	61
5	81	84	90	85	85	85	84	84	85
10	95	93	94	98	93	90	90	91	94
15	97	94	95	98	93	93	95	95	96
20	97	97	97	100	95	97	96	96	97
25	97	98	98	100	98	98	96	97	97
30	97	99	99	100	99	98	99	98	97
35	97	99	99	100	100	100	100	100	97
40	97	99	100	100	100	100	100	100	98
45	97	99	100	100	100	100	100	100	99
50	97	100	100	100	100	100	100	100	100
55	97	100	100	100	100	100	100	100	100
60	97	100	100	100	100	100	100	100	100

Table 7. Test results of experiment with 500 MIDI songs with variations of feature sets.

Top-n Rate(%)	Classifier								
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	42	58	65	64	63	62	65	61	56
5	78	82	86	84	80	78	76	78	81
10	86	91	91	89	87	86	87	87	88
15	93	94	92	94	93	90	91	92	92
20	95	95	93	97	93	93	93	95	94
25	96	95	93	98	93	95	96	95	94
30	96	97	93	99	96	97	97	97	94
35	96	97	95	100	98	97	97	98	94
40	97	97	97	100	98	97	98	98	95
45	97	97	98	100	100	99	100	99	96
50	97	97	99	100	100	100	100	100	97
55	97	97	100	100	100	100	100	100	99
60	97	98	100	100	100	100	100	100	100

The results of using five classifiers give the best performance in all datasets. That is, it can achieve 100% of top-n which are top-10 in case of 100 MIDI songs, top-20 in case of 300 MIDI songs, and top-35 in case of 500 MIDI songs. The feature set of five classifiers consists of features, 7, 14, 15, 16, 28, as shown in Table 4. That is the sets of attributes {P7}, {P2, P3, P8}, {P2, P4, P9}, {P2, P4, P6} and {P1, P4, P6, P8, P10}, as shown in Table 3. From this result, we found that P5 is not included in this set. While P8, which note segmentation, was processed, is employed. Feature sets that use two classifiers can achieve 100% accuracy rate at top 60 or feature set of ten classifiers can achieve 100% accuracy rate at top-30 for MIDI 100 songs in database, as shown in Table 5. Meanwhile, if MIDI songs in database increase feature set of 2 classifiers can only achieve 97% accuracy rate while feature set of 10 classifiers can achieve 100% accuracy rate at top-50 and top-60 for MIDI 300 and 500 songs in database, as shown in Tables 6 and 7 and Figures 6, 7 and 8.

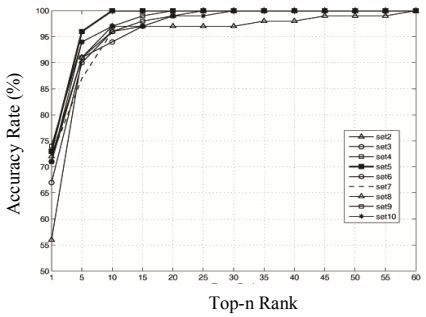

Figure 6. The performance of feature sets with 100 MIDI songs.

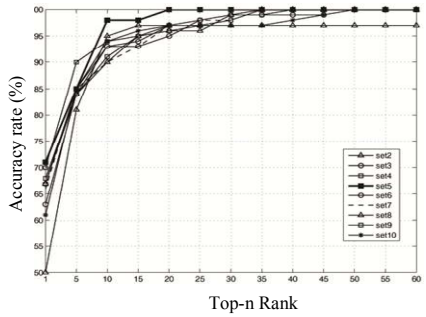

Figure 7. The performance of feature sets with 300 MIDI songs.

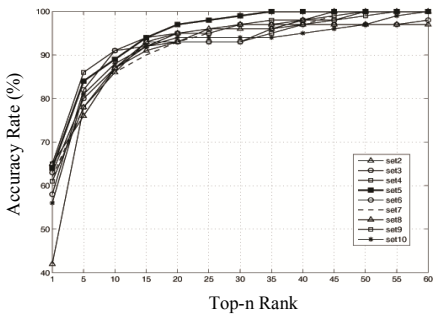

Figure 8. The performance of feature sets with 500 MIDI songs.

In addition, query time is used to measure the complexity of our proposed technique, as shown in

Table 8. We performed all the tests on a notebook with a CPU of Intel® Core™2 Duo processor 2.26GHz, 2GB of RAM. Normally, MCS with more classifiers take more query times.

Table 8. Test results of query time.

# Classifier	Query Time (Second)		
	100 MIDI Songs	300 MIDI Songs	500 MIDI Songs
2	1.39	5.84	10.00
3	2.63	7.99	13.73
4	3.33	10.14	17.50
5	4.10	12.24	21.14
6	4.73	14.46	24.90
7	5.42	16.69	28.76
8	6.15	18.86	32.58
9	6.77	20.89	36.16
10	7.50	22.97	39.84

6. Conclusions

In this paper, we propose new method for feature extraction and soft majority voting to make important decision if the vote is equal in application of QBH. Our approach consists of two processes which make humming sound go through note segmentation and then extract the feature to create many feature sets by using six approaches for input and four approaches for MIDI. The main feature we use in each set, it obtains from melody contour extraction algorithm, which we have proposed earlier. Next, soft majority voting will be used for making a decision to choose the best result.

The advantage of our approach is to increase efficiency and accuracy for retrieving data. From the use of multiple classifiers system by using soft majority voting as we have proposed, if the score is equal, all the members will get to reconsider by finding minimum distance, which we can look at this as an advantage. Moreover, using more than one feature can achieve better accuracy rate than one feature because of including more information and obtaining multiple aspects of that. From the experiments, using feature set which consists of 5 classifiers will get 100% accuracy at the early top-n rank in retrieving. Nevertheless, using a greater number of classifiers makes the system higher complexity and longer query time.

Acknowledgements

This work was assisted by Suan Dusit Rajabhat University through support with a scholarship and Rangsit University by providing the laboratory room for data processing. We would like to thanks 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] Behroozmand R. and Almasganj F., "Comparison of Neural Networks and Support Vector Machines Applied to Otimized Features Extracted from Patients' Speech Signal or Classification of Vocal Fold Inflammation," *in Proceedings of the 5th IEEE International Symposium on Signal Processing and Information Technology*, Athens, pp. 844-849, 2005.

[2] Ghias A., Logan J., Chamberlin D., and Smith B., "Query by Humming: Musical Information Retrieval in an Audio Database," *in Proceedings of the 3rd ACM International Conference on Multimedia*, New York, pp. 231-236, 1995.

[3] Holt A., Reinders T., and Hendriks A., "Multi-Dimensional Dynamic Time Warping for Gesture Recognition," *in Proceedings of the 13th Annual Conference of the Advanced School for Computing and Imaging*, Holland, pp. 23-32, 2007.

[4] Jiangtao H. and Minghui W., "Dynamic Combination of Multiple Classifiers Based on Normalizing Decision Space," *in Proceedings of WASE International Conference on Information Engineering*, Beidaihe, vol. 1, pp. 149-153, 2010.

[5] Jing Q., Wang X., Zhou M., and Liu X., "A Novel MIR Approach Based on Dynamic Thresholds Segmentation and Weighted Synthesis Matching," *in Proceedings of IET Conference on Wireless*, *Mobile and Sensor Networks*, Shanghai, pp. 1017-1020, 2007.

[6] Jun B., Rho S., and Hwang E., "An Efficient Voice Transcription Scheme for Music Retrieval," *in Proceedings of International Conference on Multimedia and Ubiquitous Engineering*, Seoul, pp. 366-371, 2007.

[7] Kim H. and Sikora T., "Audio Spectrum Projection Based on Several Basis Decomposition Algorithms Applied to General Sound Recognition and Audio Segmentation," *in Proceedings of the 13th European Signal Processing Conference*, Austria, pp. 1047-1050, 2004.

[8] Kosugi N., Nishihara Y., Sakata T., Yamamuro M., and Kushima K., "A Practical Query-By-Humming System for a Large Music Database," *in Proceedings of the 8th ACM International Conference on Multimedia*, New York, pp. 333-342, 2000.

[9] Leila C., Maamai K., and Salim C., "Combining Neural Networks for Arabic Handwriting Recognition," *International Arab Journal of Information Technology*, vol. 9, no. 6, pp. 588-595, 2011.

[10] Lemstrom K., Laine P., and Perttu S., "Using Relative Interval Slope in Music Information Retrieval," *in Proceedings of the International Computer Music Association*, China, pp. 317-320, 1999.

[11] Li P., Zhou M., Wang X., and Li N., "A Novel MIR System Based on Improved Melody

Contour Definition," *in Proceedings of International Conference on Multimedia and Information Technology*, China, pp. 409-412, 2008.

[12] Liu Y., Xu J., Wei L., and Tian Y., "The Study of the Classification of Chinese Folk Songs by Regional Style," *in Proceedings of International Conference on Semantic Computing*, Irvine, pp. 657-662, 2007.

[13] Lu L., You H., and Zhang H., "A New Approach to Query by Humming in Music Retrieval," *in Proceedings of International Conference on Multimedia and Expo*, Tokyo, pp. 595-598, 2001.

[14] McNab R., Smith L., Witten I., Henderson C., and Cunningham S., "Towards the Digital Music Library: Tune Retrieval from Acoustic Input," *in Proceedings of the 1st ACM International Conference on Digital Libraries*, New York, pp. 11-18, 1996.

[15] Mcnab R., Smith L., and Witten I., "Signal Processing for Melody Transcription," *in Proceedings of the 19th Australasian Computer Science Conference*, Australia, pp. 301-307, 1996.

[16] Phiwma N. and Sanguansat P., "A Music Information System Based on Improved Melody Contour Extraction," *in Proceedings of International Conference on Signal Acquisition and Processing*, Bangalore, pp. 85-89, 2010.

[17] Phiwma N. and Sanguansat P., "An Improved Note Segmentation and Normalization for Query-by-Humming," *Rangsit Journal of Arts and Sciences*, vol. 1, no. 2, pp. 139-148, 2011.

[18] Raphael C., "Automatic Segmentation of Acoustic Musical Signals using Hidden Markov Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 4, pp. 360-370, 1998.

[19] De-Mello R. and Gondra I., "Multi-Dimensional Dynamic Time Warping for Image Texture Similarity," *in Proceedings of the 19th Brazilian Symposium on Artificial Intelligence Savador*, Brazil, vol. 5249, pp. 23-32, 2008.

[20] Ryyndnen M. and Klapuri A., "Query by Humming of MIDI and Audio using Locality Sensitive Hashing," *in Proceedings of International Conference on Acoustics*, *Speech and Signal Processing*, Las Vegas, pp. 2249-2252, 2008.

[21] Shih H., Narayanan S., and Kuo C., "A Statistical Multidimensional Humming Transcription using Phone Level Hidden Markov Models for Query by Humming Systems," *in Proceedings of International Conference on Multimedia and Expo*, USA, vol. 1, pp. 61-64, 2003.

[22] Shih H., Narayanan S., and Kuo C., "Multidimensional Humming Transcription using A Statistical Approach for Query by Humming Systems," *in Proceedings of International Conference on Acoustics*, *Speech*, *and Signal Processing*, China, vol. 5, pp. 541-544, 2003.

[23] Sun X., "A Pitch Determination Algorithm Based on Subharmonicto-Harmonic Ratio," *in Proceedings of the 6th International Conference of Spoken Language Processing*, USA, pp. 676-679, 2000.

[24] Sun X., "Pitch Determination and Voice Quality Analysis using Subharmonic-to-Harmonic Ratio," *in Proceedings of International Conference on Acoustics*, *Speech*, *and Signal*, Orlondo, vol. 1, pp. 333-336, 2002.

[25] Vlachos M., Hadjieleftheriou M., Gunopulos D., and Keogh E., "Indexing Multi-Dimensional Time-Series with Support for Multiple Distance Measures," *in Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Springer-Verlag, pp. 216-225, 2003.

[26] Zhu Y., Xu C., and Kankanhalli M., "Melody Curve Processing for Music Retrieval," *in Proceedings of International Conference on Multimedia and Expo*, Tokyo, pp. 285-288, 2001.

Nattha Phiwma received her PhD degree in information technology from Rangsit University, Thailand in 2011. She is an assistant professor in the Department of Computer Science at Suan Dusit Rajabhat University, Thailand. Her research areas are music information retrieval and digital signal processing.

Parinya Sanguansat received his B.Eng, M.Eng. and PhD degrees in electrical engineering from the Chulalongkorn University, Thailand. He is an assistant professor in the Faculty of Engineering and Technology, Panyapiwat Institute of Management, Thailand in 2001, 2004 and 2007 respectively. His research areas are digital signal processing in pattern recognition including on-line handwritten recognition, face and automatic target recognition.