J Inf Process Syst, Vol.13, No.5, pp.1213~1228, October 2017
https://doi.org/10.3745/JIPS.02.0073

ISSN 1976-913X (Print)
ISSN 2092-805X (Electronic)

JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Thai Classical Music Matching Using t-Distribution on Instantaneous Robust Algorithm for Pitch Tracking Framework

Pheerasut Boonmatham*, Sunee Pongpinigpinyo*, and Tasanawan Soonklang*

Abstract

The pitch tracking of music has been researched for several decades. Several possible improvements are available for creating a good t-distribution, using the instantaneous robust algorithm for pitch tracking framework to perfectly detect pitch. This article shows how to detect the pitch of music utilizing an improved detection method which applies a statistical method; this approach uses a pitch track, or a sequence of frequency bin numbers. This sequence is used to create an index that offers useful features for comparing similar songs. The pitch frequency spectrum is extracted using a modified instantaneous robust algorithm for pitch tracking (IRAPT) as a base combined with the statistical method. The pitch detection algorithm was implemented, and the percentage of performance matching in Thai classical music was assessed in order to test the accuracy of the algorithm. We used the longest common subsequence to compare the similarities in pitch sequence alignments in the music. The experimental results of this research show that the accuracy of retrieval of Thai classical music using the t-distribution of instantaneous robust algorithm for pitch tracking (t-IRAPT) is 99.01%, and is in the top five ranking, with the shortest query sample being five seconds long.

Keywords

Pitch Tracking Algorithm, Instantaneous Robust Algorithm for Pitch Tracking, T-Distribution, Shortest Query Sample

1. Introduction

Content-based music systems are grown-up based on the features of melody. The song can be characterized as an arrangement of melodic notes in progression, and is additionally a grouping of pitches and lengths. The primary elements of melody are pitch, length, timbre and dynamics, where the pitch is the perceived fundamental frequency and the length is a particular time interval. In addition, the melody is progressively characterized as the non-abrasiveness or tumult of a sound, while timbre alludes to the tone color and is the nature of a voice.

Ghias et al. [1] published a study of identification song by humming. That study used the autocorrelation method to calculate the fundamental frequency (F0), while the vector of pitch was separated into notes. To help resolving the issue which arose that study used the autocorrelation method to make the fundamental frequency for a difference of music note, the obtained note was

Manuscript received March 22, 2017; first revision May 8, 2017; accepted June 17, 2017.
Corresponding Author: Sunee Pongpinigpinyo (pongpinigpinyo_s@su.ac.th)
* Dept. of Computing, Faculty of Science, Silpakorn University, Nakorn Pathom, Thailand (pheerasut.b@gmail.com, soonklang_t@su.ac.th)

transformed into three symbols: U (up), D (down), and S (same). Foote [2] proposed the audio retrieval with acoustic similarity. The model is generated using Mel-scaled coefficients for every sound document in the music archive, and the same strategy for the client's inquiry.

Chouet al. [3] used chords to establish the indexing of a song archive. A PAT-tree was used as an indexing frame to archive a music database. Kim and Whitman [4] proposed a methodology using voice coding features for the process of singer identification. They used acoustic features derived from linear predictive coding coefficients and frequencies for the construction of their model, which is used for singer identification. They implemented the model using a Gaussian mixture model and support vector machine before comparing their results. This research related to the pitch of vocals and searched music using dynamic time warping [5]. In summary, it can be seen that the pitch of the fundamental frequency is important in comparing music.

Previous research on music information retrieval has used text-based methods while novel studies revolve about content-based techniques. Music information retrieval systems have used various approaches such as finding certain features of sound to create a search index in the hash table. Prior research has conducted experiments using a comparison of the audio features of Thai classical music instruments to determine certain features of the music related to the pitch [6]. The Thai classical music scale is different from the western music scale [7], and previous research results have shown that pre-processing is required for the framing process. This requirement is the major problem of processing because the frames of the music that are compared to each other have a very little chance to determine the position of the frame with regard to the same position. The result of the calculation of features is directly affected by the framing, which applies to the comparison of temporal data. Thus, it affects searching in any part of the music. This means that the ability to estimate the pitch to deal with deviations in framing, in order to allow the comparison of music, gives the highest accuracy. A process for pitch estimation is therefore presented in this paper, in order to solve this problem of framing. The method of comparison used in these experiments provides search results with a percentage accuracy higher than 90%, using a t-distribution over pitch tracking. The longest common subsequence (LCS) is applied in a comparison of the similarity of the music in this paper [8]. This research presents a scheme for pitch tracking estimation using a t-distribution, in which the detection of pitch melody acts to restrict the recognition of pitch track candidates.

Thai classical music was created in its current form in the Thai royal court, based on a traditional Thai sound-scale of seven tempered notes. Generally, Thai classical music was used for the aim of providing a musical background for dramatics, rituals, and events. While the music does exist, they maintain only as a middle melodic outline around which instrumentalists extemporize. Thai classical instruments can be divided into four categories: plucked instruments, bowed string instruments, percussion instruments, and woodwind instruments. Thai classical musical bands are divided into three categories, as follows.

A piphat (Thai flute band) mainly comprises percussion instruments such as gongs, drums and the Thai flute (pi), which produces the melody, and also includes tempo instruments. A piphat band can be divided into subgroups consisting of between five and 14 musical instruments.

A khrueang sai (bowed string instrument band) consists of bowed string instruments such as the "*so-duang*", "*so-u*" and "*cha-khe*" as major instruments and also contains woodwind and the percussion instruments as constituents of the band. The khrueang sai band can be divided into subgroups consisting of between five and seven musical pieces.

A mahori (bowed string instrument band combined with a piphat band) consists of all kinds of instruments. The mahori band can be divided into subgroups consisting of between six and 23 musical pieces. The main difference in the mahori band is that the saw sam sai accompanies the vocalist, who takes a more prominent role in this band than in any other Thai classical music band.

A major challenge for this research is that the Thai musical scale is different from the western music scale due to the differences in musical tones and instruments between Thai and western Music, in terms of tonality, instrument shapes, the technology used in constructing instruments, musical composition techniques, and so on. The objective of this research is to create a Thai musical scale frequency filter, and to create the features of Thai classical music retrieval with the highest accuracy and the shortest query.

This paper is organized as follows. Section 2 provides a system overview, while Section 3 discusses pitch melody extraction. Section 3 also includes a discussion of pitch estimation using the IRAPT framework, and sampling distributions in IRAPT using the student's t-distribution. Section 4 examines the t-IRAPT sequence alignment comparison, while Section 5 presents an experiment, including experimental results of the percentage comparison of the t-IRAPT sequence alignment. Section 6 provides a discussion and finally presents the conclusions of this study.

2. System Overview

The music comparison system was designed to first import certain parts of music that serve as the query. Then, the indexing phase is used to extract the features to be used in indexing for a search. The indexed database of music performs the same function as the index of a query. The processes are divided into two phases, consisting of indexing and searching, as illustrated in Fig. 1 below.

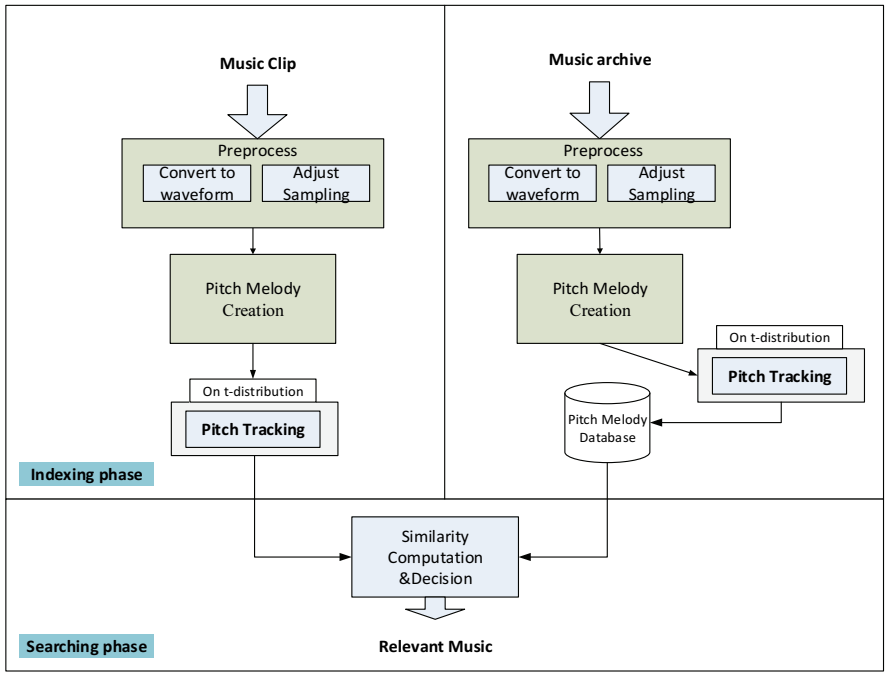


Fig. 1. System overview.

2.1 Indexing Phase

The aim of the indexing phase is to create a searchable list for each of the songs in the music archive. Pitch melody extraction is the focus of the emblematic description of the pitch melody, being related to the pitch sequence for each song in the music archive. It is not necessary to reduce the noise in the music in the main melody extraction method, to extract the pitch in the process of identifying the sequence of the fundamental frequency; IRAPT is a robust algorithm for generating a pitch track. The student's t-distribution was used to create the index. The index alignment was estimated from the likelihood of the pitch for each frame, which was calculated using IRAPT and the t-distribution.

2.2 Searching Phase

In this phase, the search uses a procedure that assigns the song with the greatest similarity, depend on the music query. It can be expected that every music query is a total expression onset or the underlying piece of that expression onset. Hence, the assignment is to discover the music archive that best matches the inquiry. A music wave record frame is changed over into a grouping of a pitch track utilizing pitch identification, in the same as that utilized as a part of the ordering stage. Then, the strategy consists in measuring the likeness between the pitch track of the inquiry and the pitch track of every music file. This research employs the string matching method to measure the similarities in melody tracking.

3. Pitch Melody Extraction

The requirement for the fundamental frequency of music is a subject that has attracted as much interest as any other topics in sound and music analysis. The study in [9] provides a very abbreviated survey of pitch tracking techniques, and then focuses on a complete description of a robust algorithm for pitch tracking, RAPT, that has proven effective in the context of basic research and synthesis engineering.

The term ‘pitch’ should be reserved for the auditory percept of tone. This can be measured, for example, by asking a listener to adjust the frequency of the sinusoid so that seems to share the same tone with the complex stimulus; the sinusoid frequency can be defined as the pitch of the complex signal. Although some computational auditory models are successful at predicting a perceived pitch containing complex signals, the pitch is not directly measurable from the signal and is a nonlinear function of the signal’s spectral and temporal energy distribution.

Fundamental frequency (F0) is the amount that is evaluated by essentially all pitch trackers. F0 is a characteristic property of periodic signals and tends to correspond well with realizing pitch. F0 estimators are required to cope with mixed excitation. For some applications, they must determine the presence or absence of glottis-induced periodicity, and thus the determination is referred to as the voicing classification. Given the nature of the speech or sound signal, it should now be clear that a whole range of excitation types is possible, from pure sound to pure non-sound. F0 is difficult to estimate due to the following factors.

1) F0 changes with time, frequently with each glottal period;
2) Sub-harmonics of F0 regularly give the ideas that are sub-multiples of the genuine F0;

3) As a rule, when solid sub-harmonics are displayed, the most sensible target gauge of F0 is obviously inconsistent with the sound-related percept;
4) Some F0 actually move up or down by an octave;
5) Voicing is often very irregular at voice onset and offset, leading to minimal wave-shape similarity in the adjacent period;
6) Panels of expert humans do not completely agree on the locations of music onset and offset;
7) The narrow band filtering of unvoiced excitation by certain vocal track configurations can lead to signals with significant apparent periodicity;
8) The amplitude of music sound has a wide dynamic range, from low (in silence) to high (in playing music).

For the various reasons mentioned above, the state of music will probably need to be determined in terms of pitch level, to create the pitch track. Pitch tracking is important for sound signal processing algorithms. Pitch tracking is created using a pitch detection algorithm, which is a method designed to estimate the fundamental frequency (F0) or "pitch" of a periodic signal, and is regularly a computerized recording of discourse or a melodic note or tone. Additionally, pitch detection can be carried out on the frequency domain, the time domain or both of them. Pitch tracking uses signal recognition processing for tone recognition, and uses disambiguation of homophones. The pitch is also valuable for prosodic variation in spoken language systems such as text-to-speech systems.

The level of pitch class is an arrangement of entire pitches that are a whole number of octaves separated. The pitch class C comprises the Cs in all octaves; it remains for all conceivable Cs, in whatever octave position [10]. Thus, using scientific pitch notation, the pitch class is the set shown in Eq. (1).

$$\{\mathrm{C_n}\} = \{\ldots, \mathrm{C_{-2}}, \mathrm{C_{-1}}, \mathrm{C_0}, \mathrm{C_1}, \mathrm{C_2}, \ldots\} \tag{1}$$

The pitch classes utilizing the whole number that start from zero, with each separately bigger number declaring to a pitch class higher than the previous class. Since octave-related pitches have a place with a similar class, when an octave is rising, the numbers start again at zero. This whole number notation alludes to the standard instance of chromatic twelve-tone scales, where the twelfth part is same to the first. This can define the fundamental frequency (F0) of the pitch as a real number U using Equation 2.

$$\mathrm{U} = 69 + 12\log_2(\frac{\mathrm{F}}{440}) \tag{2}$$

In addition, a pitch detection algorithm (PDA) is an algorithm designed to estimate the fundamental frequency (F0) of digital recording of speech or a musical audio. The pitch detection algorithm can be done in the other way such as IRAPT.

3.1 Pitch Estimation Using the IRAPT Framework

Most of the F0 estimation schemes consist of three main parts: 1) pre-processing or pretreatment of the signal; 2) creation of F0 candidate estimates for the actual audio signal; and 3) a post-processing stage that selects the best candidates and refinement for an estimated F0. The following to the original

robust algorithm for pitch tracking (RAPT), the algorithm gives better results for a wide range of signal sampling rates in the range 6 to 44 kHz [11]. This algorithm contains the following steps:

1) To reduce the cost of computing, 6 kHz is the lowest sampling rate selected. This method of down-sampling does not cause a loss of frequency resolution;
2) The estimation is generated of the instantaneous harmonic parameters associated with the processing operations analysis during the period. The method used in the calculation is improved from the analysis of a discrete Fourier transform (DFT)-modulated filter bank;
3) The harmonic frequency received will be checked repeatedly to eliminate the possibility of frequency duplication that may occur due to overlapping frequencies;
4) The function used for instantaneous period candidate generation is an approximate calculation. An approximation is used instead of the direct evaluation of the instantaneous normalized cross-correlation function (NCCF) in order to reduce the cost of computing. The approximation used here is the inverse fast Fourier transform (IFFT);
5) The selection of the best candidate pitch is carried out using a dynamic programming technique;
6) In this process, an assessment is made at the outset of the pitch, which is represented by "Pitch Estimate 1" as it is shown in Fig. 2. The accuracy of the pitch is restricted by the size of the inverse fast Fourier transform (IFFT) that is used in the prior step and which is degraded more by a high frequency transition. The pitch contour is used to estimate the time-warped signal obtained from the down-sampled source using the all-pass sinc filters with the estimated F0;
7) Instantaneous harmonic parameters were used to estimate the signal. In Step 2, a DFT-modulated filter bank is used, although there is no overlap of channels and each channel of the filter bank processes only one correspondent harmonic;
8) The new pitch values (denoted as "Pitch Estimate 2" in Fig. 2) is calculated.

The above steps are shown in Fig. 2.

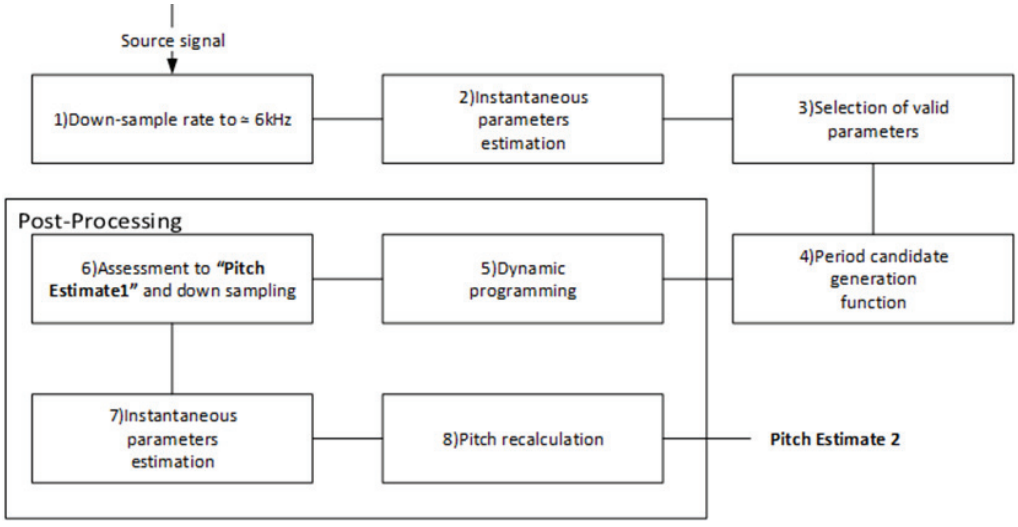
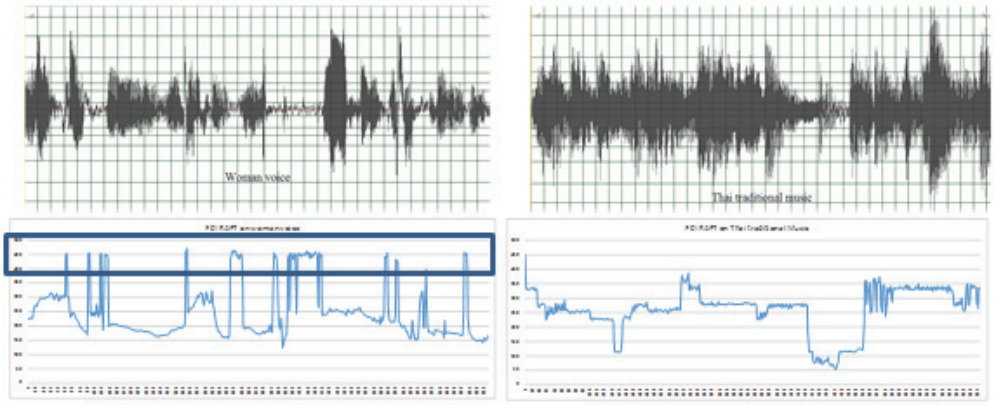


Fig. 2. IRAPT process.

IRAPT is the process used in estimating the pitch, which is demonstrated in the experiments [12]. Pitch extraction is feasible, and word boundaries can be segmented using an unvoiced position to separate these from each other. This method is applied to speech recognition [13], which is characterized by discontinuities as shown in Fig. 3(a). In this figure, the frequency in the box shows the frequencies employed to segment the words; however, the current research uses music involving a contiguous pitch. A portion of the audio with many voices is used to express the characteristics of combined frequency bands. The characteristics of Thai classical music are different from western music. Such as western musical instruments can generate sounds for more tones than the existing Thai musical instruments. Western musical instruments can make sounds that are lighter, easier, or incorporate tremolo, and can control the length of the sound. For example, when piano keys are pressed, multiple notes are sounded simultaneously; while the key is pressed, a long note is sounded, and when it is lifted, the note stops immediately. The Thai gamelan cannot make a sound stop immediately, because it does not use vibrating equipment or a pedal. The music signal is the whole pitch but it cannot identify the actual sound that shown in Fig. 3(b). Thus, the improved methods of IRAPT produce the melody line of the music with the statistical method.



(a) (b)

Fig. 3. IRAPT Pitch extraction between speech (voice) and music.

3.2 Sampling Distributions on IRAPT Using the Student's T-Distribution

Typically, the distribution of the population would be distributed over a wide distribution of the sample. This means that the distribution of the population $\boldsymbol{s}$ is larger than the estimated sample. $\overline{\boldsymbol{x}}$ is the average of the sample. $\boldsymbol{s}_{\overline{\boldsymbol{x}}}$ is the standard deviation of the distribution of the average of the sample, $\boldsymbol{s}$ is the standard deviation of the distribution of the population and $\boldsymbol{n}$ is the number of samples. Thus, when the number of samples increases, the standard deviation is reduced.

The t-test method for statistical hypothesis testing is an approach based on the student's t-distribution [14]. It is used to determine whether two sets of data are significantly different from each other. The method used in this case involves a small amount of data ($n < 30$). The article in [15] showed that a small number of samples does not allow the distribution to match the standard normal distribution. The formula of t-score is shown in Eq. (3):

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (3)$$

where S is the standard deviation of the distribution of the average of the sample, n_x is the number of samples, $\bar{x}_1$is the average of sample 1, $\bar{x}_1$is the average of sample 2, S_1^2 is the standard deviation of sample 1, and S_2^2 is the standard deviation of sample 2.

The current research uses this property in the application of IRAPT to produce an index that is used to compare the music with the music samples. The process of creating an index for an instantaneous robust algorithm for the pitch tracking framework with t-distribution uses the following steps:

1) The process begins with an example of the process for estimating pitch using the IRAPT framework for a sample of Thai classical music, as shown in Fig. 4.
2) When the process has approximated the pitch using IRAPT, F0 is used to create the index, as shown in Fig. 5.

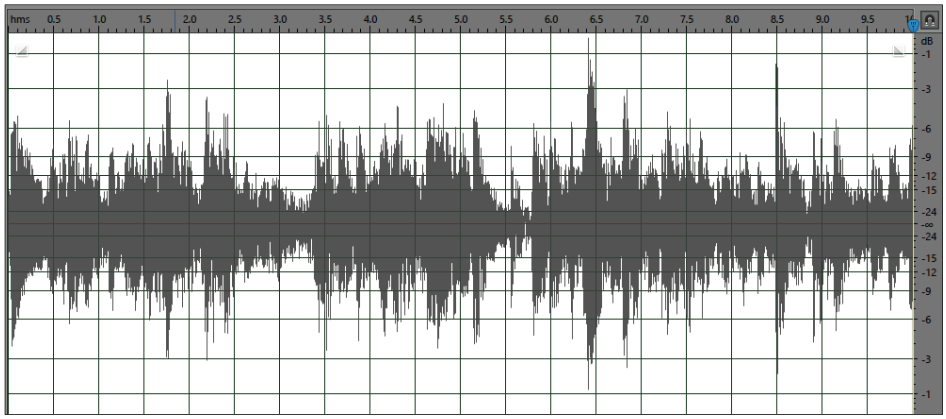


Fig. 4. Thai classical music sound waves.

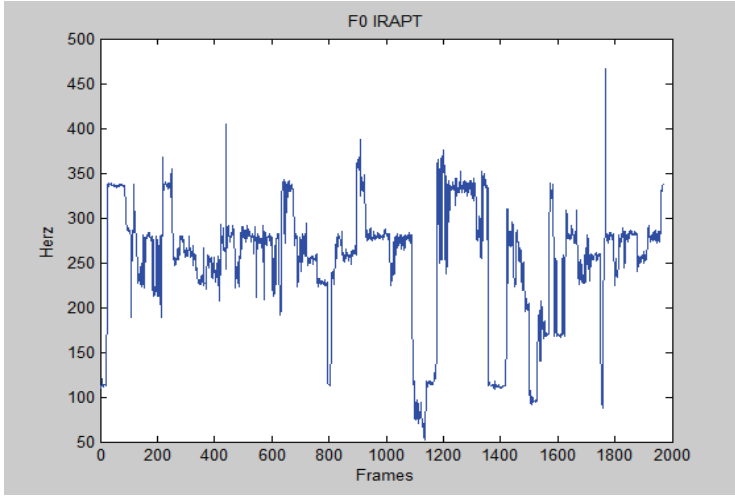


Fig. 5. Fundamental frequency (F0) from IRAPT framework.

3) This step determines the extent of the pitch. Using a default frame size of 20 ms, it starts with small frames which are then merged together to form a larger frame. This process uses frames of size 50 ms [16], which is suitable as it is close to the frame size of the pitch of Thai classical music. The level of the frequency occurs respectively are assessed by the F0 apply in the search system [17]. It uses the fundamental frequency range which determines the frequency bin, as shown in Table 1. The frequency range shown below is derived from the frequencies used by a mahori band in Thailand, matched with a range of frequencies from the diatonic scale; the frequency is then adjusted to the fundamental frequency (F0). Finally, the frequency of Thai classical music is filtered by range to the frequency bin corresponding to the fundamental frequency as shown in Fig. 6.

Table 1. Frequency of notes

No. of frequency bin	Thai note	Frequency of F0 range (Hz)	Western note
1	DO	220.36-247.35	B^b
2	RE	247.36-277.65	C
3	ME	277.66-302.39	D
4	FA	302.40-330.18	E^b
5	SOL	330.19-370.61	F
6	LA	370.62-403.65	G
7	TI	403.66-437.73	A^b

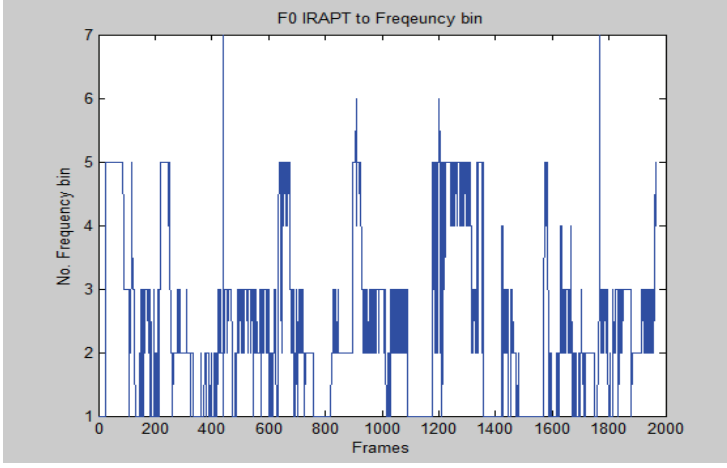


Fig. 6. F0 to frequency bin.

4) Following, the group of frame sequences of F0 is set up using sets of 15 frames, as shown in Fig. 7. This is a comparison of the differences between the two groups. The results of hypothesis testing of the change in F0 are used to determine changes in the level of the pitch of the sound, and this uses the following steps:
   1. The number of F0 frames is determined;
   2. F0 is divided into sets of 15 frames (in the experiment using changing number of frames, the values were 15, 20 and 25 respectively). The Thai musical notes have durations of

between 0.07 and 0.12 seconds, shown in Fig. 8 using circles which display the example of the boundary of one Thai musical note appears clearly. Thus, dividing the frame into the range of Thai musical notes gives 15 frames of length 0.07 seconds and 25 frames of length 0.13 seconds.

3. The statistical hypothesis is $H_0: \mu_1 = \mu_2$ and $H_1: \mu_1 \neq \mu_2$, where $\mu_1$is the F0 average of set N and $\mu_2$is the F0 average of set N+1;
4. The level of significance was 0.05, and the crisis use the t -score from the t-test table;
5. The t-test is used to test hypotheses using statistical calculation based on Equation (3);
6. Decisions are made using the t-value from the t-test table and the t-test from Equation (3); if the t-test is lower than the value in t-test table, this shows that the pitch of the test is the same pitch.

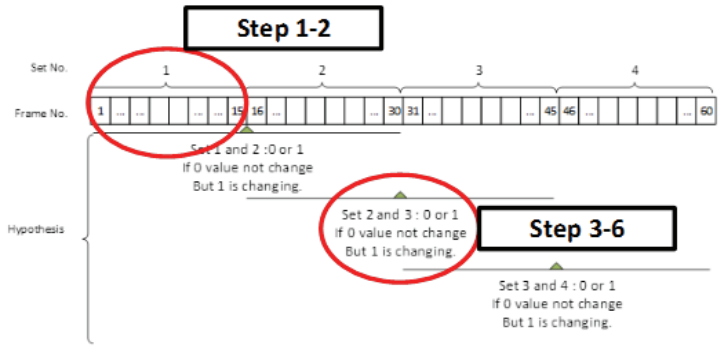


Fig. 7. Sequence of framesets in hypothesis testing.

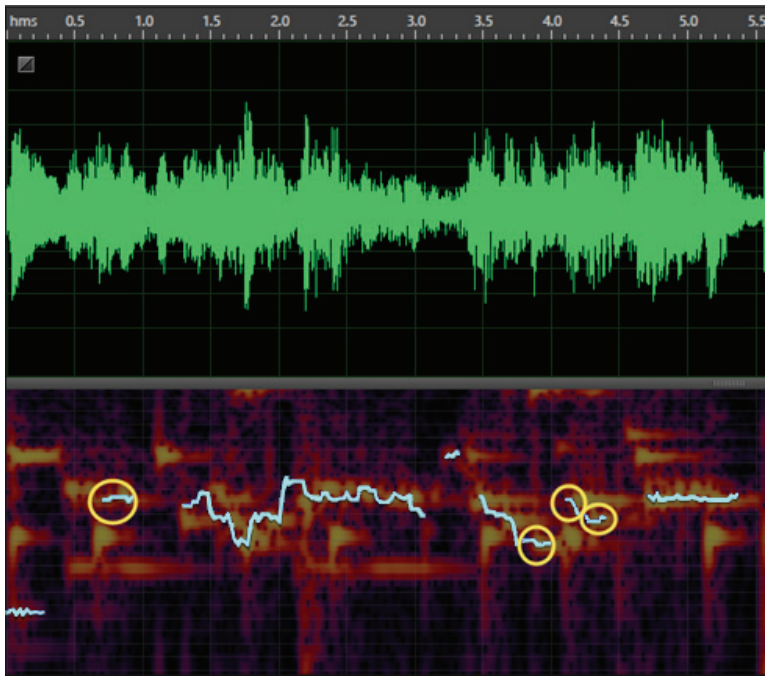


Fig. 8. Duration of a single Thai musical note.

Finally, the t-distribution of instantaneous robust algorithm for pitch tracking (t-IRAPT) sequence alignment is calculated using the t-distribution of the sequence of F0 frames, as shown in Fig. 9.

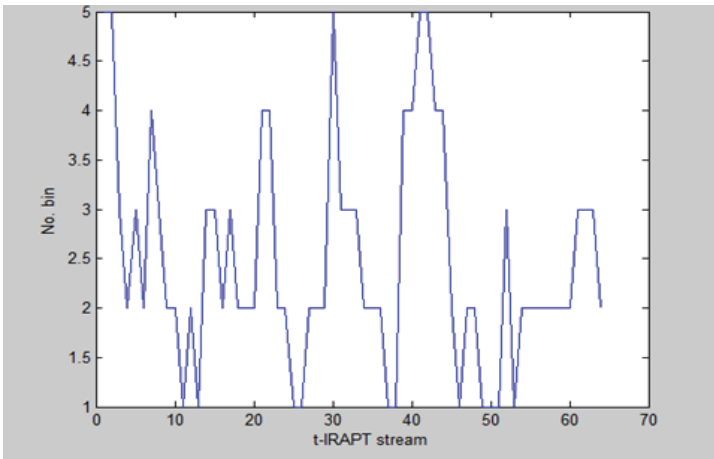


Fig. 9. t-IRAPT sequence alignment with T-test.

4. Sequence Alignment of Melody Similarity

The longest common subsequence is an algorithm for the comparison of strings [18]. The longest subsequence is a sequence that appears in the same order and is necessarily contiguous in both the strings. This is a recursive solution which decomposes the larger problem into responses from sub-problems. When there is an overlap of many sub-problems, however, then the dynamic programming method performs better, as shown in Fig. 10.

Base Cases: If any of the strings are null, then LCS will be 0.

Check whether the i^{th} character in string A is equal to the j^{th} character in string B

Case 1: Both characters are same:

LCS[i][j] = 1 + LCS[i-1][j-1] (add 1 to the result, remove the last character from both the strings and check th e result for the smaller string.)

Case 2: Both characters are not the same:

LCS[i][j] = 0

At the end, traverse the matrix and find the maximum element in it; this will be the length of the longest com mon subsequence.

Fig. 10. Concept of the longest common subsequence.

In this research, the sequence of frequency bin number is used as it is t-IRAPT sequence alignment of each music that was stored in the music database. Similarly, the t-IRAPT sequence alignment of the

piece of music was constructed as the query sequences of frequency bin numbers. Then we compare the similarity of any part of music that was stored in the music database with the part of music query.

A comparison of the t-IRAPT between the music (archive) and the query (music examples) using the LCS is shown in Fig. 11. This shows the beginning of a comparison of the difference between the music and the query. The bar chart represents the different tracks and compares the query at the frames. Fig. 12 compares the results used to find the maximum element; this will be the length of the longest common subsequence. It can be observed from the bars shown on the graph that frames 43 to 58 of the music are similar to the query.

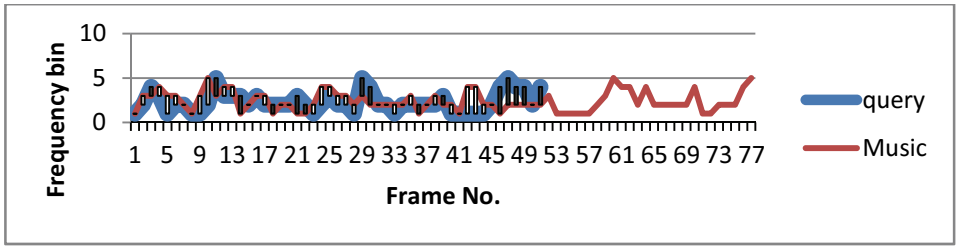


Fig. 11. Initial state of approximate matching.

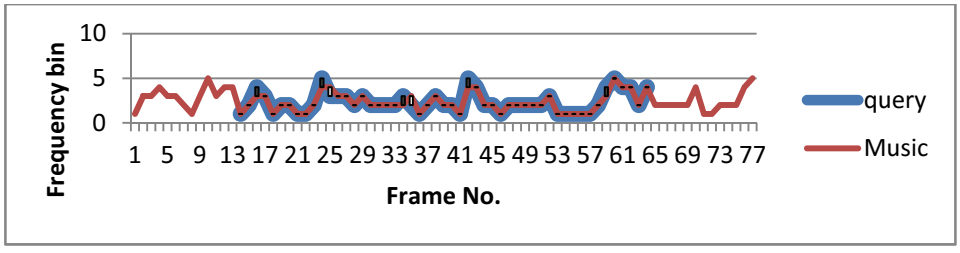


Fig. 12. Result of the least difference with LCS approximate matching.

5. Experiment Results

The music archive used in this research consisted of 102 songs extracted from Thai classical music. The extracted waveform signals use a sampling rate of 22.05 kHz, stereo channels with a bit depth of 16 bits. Each song in the music database is 30 seconds long. A total of 102 queries were collected from a random selection from the database, in which each query was associated with one of the songs in the database. The query files were also recorded at a sampling rate of 22.05 kHz. The length of each query range was 5, 10 and 15 seconds respectively. The performance of the retrieval was measured using the regulation of song accuracy as defined in Equation 6. All the samples can be heard at the following link: https://goo.gl/LxGrlO.

$$\text{accuracy}(\%) = \frac{\#\text{Queries on songs correct}}{\#\text{Queries}} \times 100\% \tag{6}$$

Table 2. Comparisons of accuracy of t-IRAPT and IRAPT

	TOP-N	Frame size (n) of t-IRAPT			Only IRAPT
		15	20	25	
Query length: 15 seconds					
Brute force	1	0	0	0	0
	3	0	0	0	0
	5	0	0	0	0
DTW	1	25 (24.50)	21 (20.59)	23 (22.55)	6 (5.88)
	3	27 (26.47)	-	-	-
	5	-	-	-	-
LCS	1	92 (90.19)	91 (89.22)	85 (83.33)	18 (17.65)
	3	93 (91.17)	-	87 (85.29)	-
	5	-	-	-	-
Query length: 10 seconds					
Brute force	1	0	0	0	0
	3	0	0	0	0
	5	0	0	0	0
DTW	1	19 (18.63)	19 (18.63)	17 (16.67)	6 (5.88)
	3	-	-	18 (17.65)	-
	5	-	-	-	-
LCS	1	95 (93.14)	95 (93.14)	97 (95.10)	28 (27.45)
	3	96 (94.12)	97 (95.10)	98 (96.08)	-
	5	-	-	-	-
Query length: 5 seconds					
Brute force	1	0	0	0	0
	3	0	0	0	0
	5	0	0	0	0
DTW	1	28 (27.45)	35 (34.31)	24 (23.53)	34 (33.33)
	3	29 (28.43)	36 (35.29)	-	-
	5	-	-	-	-
LCS	1	94 (92.16)	101 (99.01)	92 (90.20)	43 (42.16)
	3	98 (96.08)	-	95 (93.14)	-
	5	-	-	96 (94.12)	-

Values are presented as number (%).

We also observed the situation where the music information retrieval system returned the results to the list of top N ranked songs for each choice of system. In this case, the top N accuracy is computed and it is defined as the percentage of queries of the songs which were among the top N ranked.

Preliminary tests were conducted to find music using only IRAPT datasets with the same method and dataset. The results show that the query can retrieve the songs correctly with an accuracy of 42.16% with a five-second query by using the longest common subsequence (LCS) as shown in Table 2. Table 2 shows the comparisons of accuracy of experimental t-IRAPT and IRAPT data. Since the process of converting the signals of a sample is the important factor in calculating the estimated value of the pitch, it is unlikely to succeed because IRAPT frames of query and archive are the beginning overlay. So IRAPT needs to adjust the pitch of these contiguous framesets by statistical tests. The results are shown in three groups, where each group uses query sizes 5, 10 and 15 seconds respectively. The similar comparison of music samples uses three methods: brute force string matching, dynamic time warping

(DTW) with the optimal warping path [19] that is the best alignment of the sequences by the minimized value with the distance function, and the longest common subsequence (LCS). The results show that comparisons using brute force string matching cannot find the song correctly, since the music archive and query are vectors of t-IRAPT; the query cannot examine every element from the searched music to match this against the query. The DTW search method provides a maximum of 35.29% accuracy with a query length of five seconds. The search method with the highest percentage accuracy of 99.01% of search results was LCS, with a query length of five seconds. The data used in this experiment were generated using t-IRAPT frame sizes of 15, 20 and 25, respectively.

6. Conclusions

This experiment requires the ability to compare the estimated pitch of the query and the data in the database, as it is necessary to search for a similar t-IRAPT vector. The t-IRAPT, derived from the instantaneous robust algorithm for pitch tracking, is a framework that is highly effective for estimating the pitch of Thai classical music. The Thai musical scale is different from the western musical scale, due to the differences in musical tones and instruments between Thai and western music; there must therefore be a difference in filtering frequency in the index creation process of western music. The experimental use of LCS and DTW is approximately compared to find similar patterns in the two vectors. For the longest common subsequence, the results show that the index created with the proposed algorithm (t-IRAPT) improves the instantaneous robust algorithm for pitch tracking using a t-test to analyze the dynamics of a pitch that is changing rapidly; the IRAPT is not efficient enough to be used as an index for searching. It can be seen that the newly created index is effective enough to be used as representative to search any part of the music. The prior index is a sequence of number of frequency bin that transforms to vector. This new index is estimated based on numerical analysis using the t-distribution. This procedure is made up of t-IRAPT vector elements that inclined to closer some integer. This change in the value of number of frequency bin is represented in each element of the vectors. Brute-force or exhaustive searching is a simple method of pattern matching which cannot match the corresponding sequence of t-IRAPT to the correct answer. The longest common subsequence (LCS) is the way of finding the longest subsequence common to all successions in an arrangement of groupings; this is an improved method and is more flexible. Since only a portion of the query sequence matches the sequence of the data being searched in, the longest sequence is the answer to the matching.

Additionally, the comparison of music using dynamic time warping (DTW) was also restricted to the issue of determining the starting point of the comparison and had a relatively limited scope for comparison. Dynamic time warping is a way for computers to find the correct matching of two sequences under certain restrictions. In order to determine the distribution of a fixed unit of time, the result is a distance and the best warping path (alignment). The distance and the alignment of an optimal match indicate the possibility that some of the competitors provide better results. However, when compared to the performance in the identical pairing of the t-IRAPT vector, the LCS proves to be more accurate.

Finally, the songs related to the music queries were retrieved by matching the best sequence. The

pitch was presented as the answer to the music search. The pitch sequence alignment comparison applied the longest common subsequence to fit the characteristics of the index used to retrieve it. The experimental results of this research show that the accuracy of Thai classical music retrieval using the t-distribution of instantaneous robust algorithm for pitch tracking (t-IRAPT) was 99.01% in the top five ranking, with a five-second shortest query sample.

References

[1] A. Ghias, J. Logan, D. Chamberlin, and B. C. Smith, "Query by humming: musical information retrieval in an audio database," in *Proceedings of the 3rd ACM International Conference on Multimedia*, San Francisco, CA, 1995, pp. 231-236.

[2] J. T. Foote, "Content-based retrieval of music and audio," in *Proceedings of the Multimedia Storage and Archiving Systems II*, Dallas, TX, 1997, pp. 138-147.

[3] C. C. Liu, J. L. Hsu, and A. L. P. Chen, "An approximate string matching algorithm for content-based music data retrieval," in *Proceedings of the IEEE International Conference on Multimedia Computing and Systems*, Florence, Italy, 1999, pp. 451-456.

[4] Y. E. Kim and B. Whitman, "Singer identification in popular music recordings using voice coding features," in *Proceedings of the 3rd International Conference on Music Information Retrieval*, Paris, France, 2002, pp. 164-169.

[5] H. M. Yu, W. H. Tsai, and H. M. Wang, "A query-by-singing system for retrieving karaoke music," *IEEE Transactions on Multimedia*, vol. 10, no. 8, pp. 1626-1637, 2008.

[6] P. Boonmatham, S. Pongpinigpinyo, and T. Soonklang, "A comparison of audio features of Thai Classical Music Instrument," in *Proceedings of the 7th International Conference on Computing and Convergence Technology (ICCCT)*, Seoul, Korea, 2012, pp. 213-218.

[7] P. Boonmatham, S. Pongpinigpinyo, and T. Soonklang, "Musical-scale characteristics for traditional Thai music genre classification," in *Proceedings of the International Computer Science and Engineering Conference (ICSEC)*, Nakorn Pathom, Thailand, 2013, pp. 227-232.

[8] D. Gusfield, *Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology*. Cambridge, UK: Cambridge University Press, 1997.

[9] W. J. Hess, "Pitch and voicing determination of speech with an extension toward music signals," in *Springer Handbook of Speech Processing*, Berlin, Germany: Springer, 2008, pp. 181-212.

[10] A. Whittall, *Serialism*. New York, NY: Cambridge University Press, 2008.

[11] D. Talkin, "A robust algorithm for pitch tracking (RAPT)," in *Speech Coding and Synthesis*, Amsterdam, Netherlands: Elsevier, 1995, pp. 495-518.

[12] E. Azarov, M. Vashkevich, and A. Petrovsky, "Instantaneous pitch estimation based on RAPT framework," in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, Bucharest, Romania, 2012, pp. 2787-2791.

[13] K. Hotta and K. Funaki, "On a robust F0 estimation of speech based on IRAPT using robust TV-CAR analysis," in *Proceedings of the Asia-Pacific, Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Siem Reap, Cambodia, 2014, pp. 1-4.

[14] L. Jiaxi, "The application and research of T-test in medicine," in *Proceedings of the 1st International Conference on Networking and Distributed Computing (ICNDC)*, Hangzhou, China, 2010, pp. 321-323.

[15] W. G. Cochran, "ES Pearson, John Wishart, "Student's" Collected Papers," *The Annals of Mathematical Statistics*, vol. 15, no. 4, pp. 435-438, 1944.

[16] F. Sha and L. K. Saul, "Real-time pitch determination of one or more voices by nonnegative matrix factorization," in *Proceedings of the Advances in Neural Information Processing Systems 17*, Vancouver, Canada, 2004, pp. 1233-1240.

[17] N. H. Adams, M. A. Bartsch, and G. H. Wakefield, "Note segmentation and quantization for music information retrieval," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 131-141, 2006.

[18] M. Zhao, Z. Li, Y. Wang, and Q. Xu, "Longest common sub-sequence computation and retrieve for encrypted character strings," in *Proceedings of the 19th International Conference on Network-Based Information Systems (NBiS)*, Ostrava, Czech Republic, 2016, pp. 496-499.

[19] P. Senin, "Dynamic time warping algorithm review," Collaborative Software Development Laboratory, *Technical Report CSDL-08-04,* 2008.

Pheerasut Boonmatham

He received M.S. degrees in Faculty of Information Technology from King Mongkut's Institute of Technology North Bangkok, Thailand in 2006. He is with the Department of Computing, Faculty of Science, Silpakorn University, Thailand as a PhD candidate. His research area includes Multimedia, Information Technology, and Music Retrieval.

Sunee Pongpinigpinyo

She received the Ph.D. degree in computer science and engineering from Chulalongkorn University, Thailand, the M.S. degree in computer science from University of Tasmania, Hobart, Australia, and the B.S. degree in Statistics. Currently, she is a faculty member in the Department of Computing, Faculty of Science, Silpakorn University, Thailand. Her research interests include Machine Learning, Data Mining, Expert Systems, and Distributed Databases.

Tasanawan Soonklang

She received Ph.D. in School of Electronics and Computer Science from University of Southampton in 2008. Currently, she is a lecturer of department of computing, Faculty of Science, Silpakorn University, Thailand. Her research interests include Natural Language Processing and Machine Learning.