# Music Scale Modeling for Melody Matching

Yongwei Zhu
Institute for Infocomm Research
21 Heng Mui Keng Terrace
Singapore 119613
ywzhu@i2r.a-star.edu.sg

Mohan Kankanhalli
School of Computing
National University of Singapore
3 Science Drive 2, Singapore 117543
mohan@comp.nus.edu.sg

## ABSTRACT
Several time series matching techniques have been proposed for content-based music retrieval. These techniques represent a melody by a time series of pitch values and use time warping distance measures for melody matching. These methods have shown to be robust and effective for music retrieval by acoustic inputs, such as query-by-humming. However, due to the key transposition issue, all the current methods need to search a large space for the proper key in melody matching. This computation can be prohibitive for a practical music retrieval system with a large database.

In this paper, we present a music scale modeling technique for melody matching. The root note of music scale (Major or Minor) of a melody is estimated by fitting the notes to a music scale model. The estimated root note can then be used as the key in melody matching. To the best of our knowledge, this is the first approach that utilizes music scale knowledge for retrieval. In our experiments, 96% of the songs in the database (3000 melodies) can fit into the music scale model. Promising results for query-by-humming retrieval have been obtained by using this novel approach.

## Keywords
Content-based music retrieval, query-by-humming

## 1. INTRODUCTION
The spread of the Internet has made a large amount of music data easily available. This has created a need for people to search for and to identify musical clips using computers. Traditional text retrieval methods are inadequate since text information such as artist's name or song title is often unavailable. Searching for music by content, therefore, becomes necessary.

Several melody matching techniques have been proposed for content-based music retrieval [2-7]. In the earlier approaches [2,3], a melody is represented by a sequence of pitch differences between contiguous music notes, called the melody contour. Melody contour matching is done by approximate string matching algorithms. However these methods require the notes in the melody to be accurately identified, which is very difficult for acoustic inputs such as humming.

A few time series matching methods have recently been proposed for melody matching [4-7]. The melody is represented by a time series of pitch values, where accurate note segmentation is not necessary. [4,5] used dynamic time warping to do the time series matching. [6] proposed a melody skeleton technique for melody alignment and matching. A time series indexing scheme is proposed in [7] for query-by-humming. The time series approach has shown to be robust and effective for music retrieval using acoustic input, such as for query-by-humming. However, due to the key transposition issue in melody matching, all these techniques need to search a large space for the correct key when doing melody matching. The computation cost can be enormous for a practical music retrieval system with a large database.

All the previous methods represent melody using a sequence of local melody features, such as the relative pitch difference or the absolute pitch value. The global information of the melody has not been considered for melody matching. However, from the music theory point of view, the pitches/notes in a melody do have a global structure, which is the *scale*. A music scale is a particular set of notes that are defined by musicians as being appropriate for a song [1]. Most music pieces are created (composed) using particular music scales, and the notes only or mainly from the scale are used in the composition. The Major scale and Minor scale are the widely used scales in music.

We have found that most music pieces are consistent in the scale and usually only a small portion of a melody can reveal its music scale and the root note. This has motivated us to use music scale information to help solve the key transposition problem in melody matching.

In this paper, we propose a scale modeling technique to estimate the scale root of a melody. The scale root of the melody is then used for proper key transposition in melody matching using the time series method. The computation for key transposition in melody matching is drastically reduced (compared with previous approaches), which makes the time series matching approach more practical for a music retrieval system. This paper is organized as follows: section 2 presents background knowledge on music scale and our proposed scale model for Major scale and Minor scale; section 3 presents how to estimate the scale root of a melody in either symbolic form or acoustic form; section 4 presents a melody matching method by using scale root; section 5 presents the experimentation; and section 6 concludes the paper.



## 2. MUSIC SCALE MODELING
### 2.1 Music Scales
Equal-tempered tuning is the de facto standard for music note tuning in modern and western music [1]. In equal-tempered tuning system, each octave has 12 notes, and the pitch difference between contiguous notes is uniformly 1 semitone. The 12 notes are

conventionally named as: A, A#, B, C, C#, D, D#, E, F, F#, G, G#. The pitch difference between any 2 notes, measured in semitones, is called the pitch interval.

A music scale is a series of notes defined by musicians as appropriate for a music piece. A particular scale is specified by the pitch intervals between the series of notes in 1 octave (and repeated in all octaves). The Major scale and the Minor scale are the most widely used music scales. Both of these 2 scales have 7 notes in 1 octave. The pitch intervals for a Major scale are: 2 2 1 2 2 2 1. And the pitch intervals for a Minor scale are: 2 1 2 2 1 2 2. A scale is usually referenced to a root note (e.g. C). In equal-tempered tuning any note can be the root note of a scale, thus key transposition is possible.

For example, Major C scale (Major scale using C as the root note) consists note C D E F G A B, Major D scale consists note D E F# G A B C#, and Minor A scale consists note A B C D E F G.

Figure 1 illustrates the Major scale and the Minor scale. It can be seen that a Major scale coincides with a Minor scale with its 6th note as the root note. And a Minor scale coincides with a Major scale with its 3rd note as the root note.

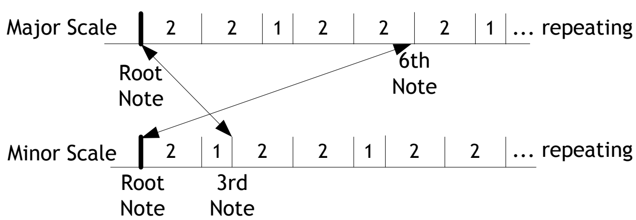


**Figure 1: Pitch intervals in Major scale and Minor scale.**

Typically a musician composes a music piece with the scale determined *a priori*. The notes of the music are mostly from the notes in the scale, although notes outside of the scale may occasionally be used. A performer will also need to have the scale to perform the music. This is even true for a hummer, who is not a musician or a trained singer. Although the hummer may not know whether it's a Major C or a Minor G, the scale (the root and the series of notes can be used) is decided (maybe subconsciously) when the tune is being hummed.

## 2.2 Music Scale Model

The music scale information is usually not encoded in the music data file, either in MIDI or in the acoustic waveform formats. But with the knowledge of the scale type and the assumption that notes of a melody are mostly from the scale, it is possible to determine the scale and its root note from the melody.

We propose a music scale model for major and minor scales for music scale estimation. As illustrated in figure 2, the music scale model is a circle of 12 equally spaced notes. The highlighted notes labeled by (0, 2, 4, 5, 7, 9, 11) correspond to the note in the music scale (Major or Minor), and we call them a *scale note*. The notes labeled by (1, 3, 6, 8, 10) correspond to the notes outside of the scale, and we call each of them a *non-scale note*. A major scale is represented by the clockwise enumeration of the scale notes starting with note 0. And a minor scale is the list starting with note 9. In other words, this model represents both major scale and minor scale.

If a melody can fit into the model (all notes can match with the scale notes in the model), the note in the melody corresponding to note 0 of the model is then the root note of a major scale. Moreover, the note in the melody corresponding to note 9 of the model is also the root note of a minor scale.

Fortunately the ambiguity between major and minor scales for this model will not affect the melody matching task. We would only need to estimate the root note of one scale (say the major scale), and use this root note to do key transposition in melody matching.

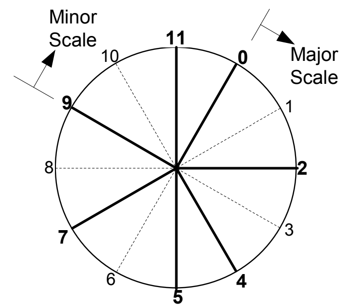


**Figure 2: A music scale model for Major and Minor scale**

Estimating the root note of Major scale for a melody implies finding which note in the melody corresponds to note 0 in the scale model. This can be done by fitting the notes of the melody into the model and computing a fitting error. A small fitting error will indicate that the quality of the scale estimate is good. The details of the major scale estimation procedure are presented in the next section.

# 3. MUSIC SCALE ESTIMATION

This section presents how to estimate the major scale root of a melody using the scale model. There are 2 melody formats in music retrieval: symbolic and acoustic. Scale estimation for these 2 formats is presented in section 3.1 and 3.2 respectively.

## 3.1 Symbolic Melody

MIDI is a popular music encoding format, in which the music notes are represented by numbers (from 0 to 127). A melody is just a sequence of such numbers. The pitch range of a melody is usually within 2 octaves (24 semitones).

Estimating the scale of a symbolic melody consists of 2 steps: (a) construction of a note histogram for the melody; (b) comparing of the note histogram with the scale model and locating the root note.

(a) Since music scale is cyclic in the octaves, the notes of a melody are mapped into a single octave by constructing a note histogram with 12 bins numbered from 0 to 11. The notes in the melody are assigned to the bins by using the following equation:

$$B = MOD(P,12), \tag{1}$$

where P is the note number (from 0 to 127), B is the bin number (from 0 to 11), and MOD is the modulus operation.

(b) The note histogram is compared with the scale model by doing a clockwise alignment of the histogram bins with the notes of the scale model. There are 12 possible alignments when the note 0 of the model is matched to each different bin of the histogram. For each alignment, a model fitting error is computed by summing the histogram bins matched with non-scale notes in the model. If the fitting error is under a threshold for an alignment, good scale estimation is claimed.

For example, the melody of "Auld Lang Syne" is composed by the notes: 58, 63, 62, 63, 67, 65, 63, 65, 67, 65, 63, 67, … And the note histogram is [23] [0] [3] [50] [0] [37] [0] [30] [0] [0] [27] [0]. The scale model fitting results is shown in Table 1. It can be seen that the fitting error is 0 when bin 2 is matched to note 0. This means that the note in bin 2 is estimated to be the root note for major scale.

**Table 1: Scale model fitting for "Auld Lang Syne"**

| Start Bin Number | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fitting Error | 77 | 93 | 0 | 167 | 3 | 114 | 56 | 50 | 120 | 0 | 137 | 33 |

It can be seen that the fitting errors are also small for bin 4 and bin 9. So the notes in bin 4 and 9 can also be claimed to be the root of major scale. Multiple root notes are claimed for this melody because not all 7 Major scale notes are used in the melody. In fact, this melody is mainly composed with 5 scale notes (refer to the note histogram). The issue of ambiguity of root note will be discussed in section 4 on melody matching.

## 3.2 Acoustic Melody

Humming is an important way of providing melody input, which can be used by laypersons without any formal music knowledge or skills. The melody from acoustic input can be transcribed to a time series of pitch values by pitch tracking [4], where each pitch value corresponds to a frame (about 50ms) of the acoustic input. The pitch values are continuous and are measured in semitones and cents. One cent is 1/100 semitone, and 1 octave pitch range has 1200 cents. The lowest pitch considered is 55Hz, which is corresponds to pitch value 0.00. And the highest pitch considered is 880Hz, which has the pitch value 48.00 (4 octaves higher than 0.00).

Scale root estimation for acoustic melody is conducted at the precision of a cent. The estimation also involves 2 steps: (a) construction of a pitch value histogram for the pitch time series; (b) comparison the pitch histogram with the scale model in order to locate the root pitch.

(a) The pitch histogram has 1200 bins numbered from 0 to 1199. The pitch values are assigned to the histogram using the following equation:

$$B = MOD(P \times 100, 1200) \quad (2)$$

where P is the pitch value (from 0.00 to 48.00, 4 octaves), B is the bin number (from 0 to 1199), and MOD is the modulus operation. Frames with no pitch are not assigned to the pitch histogram.

(b) The pitch histogram is compared with the scale model by aligning the histogram bins with the notes of the scale models. Only 7 bins can be exactly aligned with a note of the scale model, since there are 1200 bins but only 12 scale model notes. Some bins can be 100 cents away from the scale model note. Starting by the scale model note 0, there are 1200 possible alignments. For each alignment, a model fitting error is computed by summing the distance of each bin to the model by using the following equation:

$$D = \sum_{i=0}^{1199} N_i \times d_i \quad (3)$$

where $N_i$ is the number of counts in the $i^{th}$ bin and $d_i$ is the distance of the $i^{th}$ bin to a closest scale model note for the alignment. $d_i$ is measured in cents and takes the value from 0 to 100.

If the model fitting error is under a threshold, a good scale estimate can be claimed.

Figure 3 shows a short portion of Auld Lang Syne produced by humming. The scale model fitting error is illustrated in figure 4. It can be clearly seen that the local minimums at 77 and 577 with fitting error of 25 correspond to the root of the major scale. The local minimum at 1074 of fitting error of 30 could also be regarded as a root of the major scale. The actual pitch of the root note in the melody can be easily derived from the pitch histogram bin number.

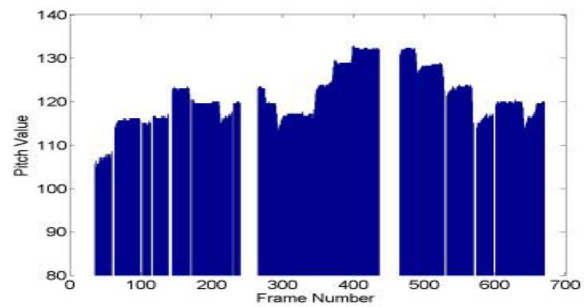


**Figure 3: A humming input of Auld Lang Syne**

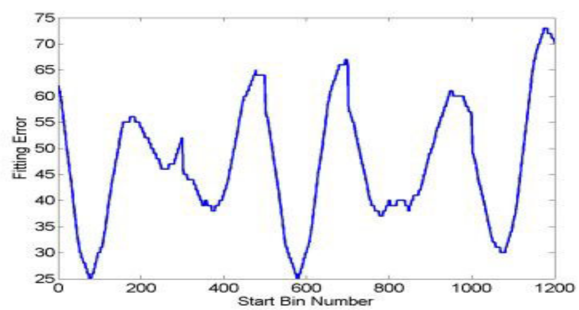


**Figure 4: Scale model fitting error for the hummed melody**

# 4. MELODY MATCHING

Music retrieval can be done by similarity matching of melodies. This section presents a melody matching method by utilizing the estimated scale root. The matching consists of 2 processes: first, transposing the query melody by matching the keys (root); and second, computing the similarity by using dynamic time warping distance.

## 4.1 Key Transposition

Usually a melody query is a short excerpt of the target melody. Since the query contains a subset of the notes of the original melody, the estimated scale root of the query is then a superset of the estimated scale root of the original melody. This is because, the less the notes used, the smaller are the constraints in scale model fitting, and there are more possible estimated scale roots. As a result, we only need to use 1 root per octave for the original melody in the database, and all the roots estimated for the query are used for melody transposition. It is guaranteed that no valid key transposition will be missed.

Key transposition for melody matching is straightforward with the scale roots available, which is demonstrated using the example shown in figure 5. The root note for the original melody is R1. R2 is the same root 1 octave higher than R1. The query melody has 3 roots estimated: Ra, Rb and Rc. Rd is the same root 1 octave higher than Ra. The key transposition can be done by matching any root in the original melody with any root in the query melody. So there could be multiple valid key transpositions. However, only (R1 to Ra) and (R2 to Rd) transposition are valid in this example. This is because the pitch range of the query has to be within the pitch range of the original melody after transposition. Furthermore, since the 2 valid transpositions are redundant, there is only one key transposition for this melody matching (R1 to Ra).

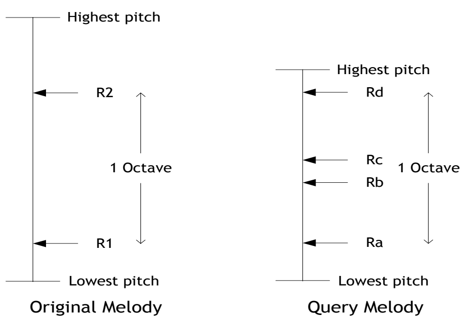


**Figure 5: Key transposition for melody matching**

Without the scale roots, any transposition that allows query pitch range to be within the original melody pitch range can be a potential valid transposition. It can thus be clearly seen that key transposition using scale root has drastically reduced the computation in key transposition for melody matching. For example, if the pitch range margin is 6 semitones, and pitch precision is in 10 cents, then the computation for transposition is reduced to 1/60 by utilizing scale roots.

## 4.2 Melody Similarity Measure

With the melodies properly transposed, a time series similarity measure can be used for melody matching. Since the tempo or the query might vary from the original tempo, time warping distance can be a good similarity measurement [5]. The distance can be computed by accumulating the local pitch distance between the 2 sequences using a dynamic programming algorithm. Boundary conditions can be imposed to accelerate the time warping distance computation, such as to skip a matching when local distance is larger than 2 semitones.

# 5. EXPERIMENTS

To investigate the scale modeling approach, we applied the scale root estimation to 3000 MIDI (Karaoke) melodies including western pop songs, Chinese pop songs and Japanese pop songs. The results have shown that 96% of the melodies can fit to our scale model (Major/Minor), when at least 1 root note can be estimated (using the threshold: 5% of total number of notes of a melody). This result shows that most popular songs are written in Major/Minor scales. About 30% among the melodies fit to the scale model have more than 1 root estimated. These melodies are mostly short melodies with a small number of scale notes used.

A query-by-humming experiment was conducted to evaluate the key transposition and melody matching performance. We compare this method to a previous method by [5], in which the time warping distance is computed based on the first note in the query. 20 hummed queries are used in the experiment. The result is shown in Table 2. The average accuracy for rank 1 is 75% compared with 45% for the previous approach. The average search time is reduced to about one-third of the previous method.

**Table 2: Query-by-humming experiments**

| Top Rank List Size | 1 | 5 | 10 | 20 |
|---|---|---|---|---|
| Without Scale Estimation[5] | 45% | 50% | 60% | 60% |
| With Scale Estimation | 75% | 85% | 85% | 90% |

# 6. CONCLUSTION AND DISCUSSION

We presented a music scale modeling approach for melody matching. The scale root for major (and minor) scale can be estimated for key transposition in doing melody matching. The experiments have shown that both retrieval accuracy and efficiency are improved using this technique.

This work is one of the first efforts to utilize music scale information for music retrieval. Although major and minor scale is assumed in this work, other types of scale can be considered in our future work. We have seen that it might be possible to estimate the scale type using the scale modeling technique. We are also investigating this method in polyphonic music retrieval.

# 7. REFERENCES


[1] Harrison, M. Contemporary Music Theory: Level One. Hal Leonard. January 1999.

[2] A. Ghias, J. Logan, and D. Chamberlin. "Query By Humming". *Proceedings of ACM Multimedia 95,* November 1995, pages 231-236.

[3] A. Uitdenbogerd and J. Zobel. "Melodic Matching Techniques for Large Music Database". *Proceedings of ACM Multimedia 99,* November 1999, pages 57-66.

[4] J.S. R. Jang, H.R. Lee, "Hierarchical Filtering Method for Content-based Music Retrieval via Acoustic Input", *Proc. ACM Multimedia 2001*.

[5] T. Nishimura, H. Hashiguchi, J. Takita, J. Zhang, M. Goto, R. Oka. "Music Signal Spotting Retrieval by a Humming Query Using Start Frame Feature Dependent Continuous Dynamic Programming", *Proc. 3rd International Symposium on Music Information Retrieval*, Indiana, USA, October 15-17, 2001.

[6] YW. Zhu, M. Kankanhalli, "A Robust Music Retrieval Method for Query-by-Humming", International conference on Information Technology: Research and Technology, New Jersey, USA, Aug. 10-13, 2003.

[7] YY. Zhu, D. Shasha, "Query-by-humming: a Time Series Database Approach" ACM SIGMOD/PODS 2003, San Diego, June 9-12, 2003.