13 Microtonal Matching with MTRI

Alexandra L. Uitdenbogerd and **Iman S. H. Suyoto**
School of Computer Science and Information Technology
RMIT University
Melbourne, Australia

Abstract

Most systems designed for locating music based on melodic input have catered exclusively to music in common Western notation and its audio analogues. This implies music which, at the auditory level, uses an equal-tempered twelve-tone scale as its basis. Obviously, this standardized tuning system is not appropriate for large bodies of the world's music. In this paper, we explore one approach to the problem of content-based retrieval of microtonal music. We conclude that a microtonal music matching system is useful for specialized types of music queries such as retrieving pieces that use specific tuning systems, and precisely defined microtonal melodic queries. We also describe a prototype system developed for a preliminary study of the problem.

Music Analysis East and West (Computing in Musicology 14).
Published by CCARH and The MIT Press, 2005-06.

13.1 Microtonal Music

13.1.1 Contexts and Definitions

Microtonal music refers to music whose notes cannot be adequately represented within an equal-tempered twelve-tone system. Many kinds of microtonal music use a twelve-tone scale in which the intervals are of irregular size. Scales with greater and lesser numbers of tones per octave constitute another category of microtonal music. There is no real limit to the number of tones into which an octave may be divided except that imposed by the threshold of human hearing.

Our application has been built on a particular kind of music found among the Sundanese people, a mountain population living in Western Java. Sundanese music uses three distinct pentatonic scales, which are numbered from highest to lowest pitches. The most common scale of the Sundanese, and the one we investigate, is the *saléndro*. The tuning system is used for both instrumental and vocal music, in both folk and modern genres. It is also used in the accompaniment of puppetry and dance. In performance, pitch fluctuation can come about partly through the substitution of a two-string fiddle (*rebab*) for a bamboo flute.

In the equal-tempered twelve-tone scale notes are by definition 100 cents apart. The tones of the *saléndro* scale fall 240 cents apart, 40 cents greater than a whole tone in the diatonic scale. One approximate tuning, in relation to the equal-tempered C-Major scale, is C↓-D-E↑-G↓-A (-C↓). (Arrows represent the flattening and sharpening of pitches.) An octave spans 1200 cents. Theorists maintain that the intervals of *saléndro* are precisely equidistant (therefore constituting the *padantara* class of tunings), although empirical evidence generally notes small deviations from equidistance. The other two pentatonic scales in common use in Java have intervals which, it is generally agreed, are more irregular.

13.1.2 Representing Microtonal Music

The representation of unequal temperaments and scales of greater or fewer than twelve tones has not been as well explored as have systems which exclusively treat the equal-tempered twelve-tone scale.

MIDI files can use pitch-wheel events to control tuning, but these methods (and other proposed extensions) have not been widely adopted. EsAC (Essen Associative Code; Schaffrath 1997) can accommodate modes with arbitrary numbers of tones but does not indicate tuning and is not suited to polyphonic music. The *Humdrum Toolkit* (1985) allows microtonal representations in several ways. Its ***cents* representation accommodates all the features required here. *Csound* also provides a cents representation. Both can also represent dynamics and articulation. *Csound* alone can represent timbre. XML hypothetically can accommodate note representations below the level of a half step, but actual implementations have not yet appeared. All of these codes are searchable, and therefore any of them could be used to search sound files of microtonal music.

13.2 Microtonal Representations

In microtonal matching, a user may want to identify instances of a piece for which a melody fragment is known, or search for particular patterns in order to determine their prevalence. A user might wish to seek pieces that use a specific tuning system. Queries may be expressed in terms of the precise intervals occurring in these tuning systems if a cents representation is used. One would search for Sundanese pieces using the 240-cent interval.

We believe that music researchers and highly trained musicians could be potential users. Queries would be specified symbolically, via graphical or textual representations. A non-expert user, in contrast, would be more likely to use metadata or a sung query. We are not confident about sung queries of microtonal materials, however, because the precision of singers is often insufficient to distinguish tuning systems.

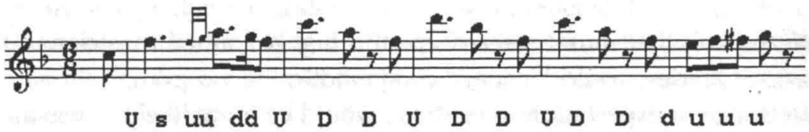
The nearest analogues come from projects related in their use of 3,000 Chinese melodies. These were encoded in EsAC (Schaffrath, 1992) to work with software designed for searches by mode, phrase structure, and other musical features. (The original data is still maintained at *www.esac-data.org/*.) The 2,241 Chinese songs (which constitute roughly 30 per cent of the original EsAC corpus) are overwhelmingly pentatonic.

Their incipits (opening phrases) have long been included in the *Themefinder* database at CCARH (*www.themefinder.org*; see under "folksongs"), but no special provision for pentatonic searching is included in the symbolic tools provided. For example, a "scale-degree" search must use diatonic, rather than pentatonic, terminology. The pentatonic scale(s) of China uses three intervals of two half steps and two of three half steps, so they are not equidistant. Modal varieties are possible, since (by Western logic) any available tone may serve as a primary, final, or principal one. Multiple methods of representation and calculation exist. In the absence of a dominant interpretive tradition, *Themefinder* groups the pentatonic tones as 2+3 and maps them to 1-2-4-5-6 of a diatonic scale, but a pentatonic search should ideally employ only the positions 1–5.

13.2.1 Microtonal Representation for Information Retrieval (MTRI)

We developed the Microtonal Representation for Information Retrieval (MTRI) for use with music which is both microtonal and polyphonic. It is designed to provide access to melodic, harmonic, and rhythmic information. Each melody is represented in both a pitch specification file (MTP), which contains tuning information, and a score file (MTS), which contains event information. In the first, the parameter N designates the number of pitches in the scale. To date 2,390 tunes have been encoded in the system. The MTP gives performance pitches in cents.

The MTS can operate on several levels of pitch resolution. Every piece in the music database undergoes *standardization*, which is the process of representing music as retrievable sequences of symbols (Uitdenbogerd, 2002). One implementation uses five symbols: U, u, r, d, D, representing "up a lot", "up a little", "repeated pitch", "down a little", and "down a lot" respectively. A typical representation might read "UUudsu" (up a lot, up a lot, up a little, down a little, same, up a little). Here we adopted the intervallic transition codes developed by Denis Parsons (1976). An example for Western music is shown below:



Duration also employs a relative description with the tokens R (the same), S (small), and L (large). These schemes correspond to the descriptions of *refined pitch contour* (prc) and *gross metric contour* (mgc) in Sapp et al. (2004).

When pitch contour and metric contour are combined, this method can give reasonable results. However, for much larger collections containing hundreds of thousands of pieces, relevant microtonal melodies are less likely to be found in response to queries, unless the queries are both longer and more precise. For example, in a collection of about ten thousand MIDI files, a three-symbol contour query required a query length of 80 to achieve the same precision average as a ten-note query using an interval-based representation (Uitdenbogerd, 2002). Sapp et al. (2004) showed in searches of a database of more than 100,000 melodies that the combination of *duration gross contour* (dgc) and *pitch gross contour* (pgc) was particularly effective in culling appropriate match candidates, since there is little overlap in the two representations. Yet combining dgc with a more refined representation of pitch or interval adds little benefit.

For our purposes, the strength of contour representation is that it is less likely to misrepresent a singer's intentions than a more exacting system. In other words, imperfect queries are more likely to find matches in more generalized representations. However, these matches require longer queries.

13.2.2 The MTRI Microtonal Matching System

Our prototype system allows users to find answers to precisely defined microtonal queries. The potential answers produced are ranked according to their degree of similarity to the query. Specifically, they are evaluated according to *mean reciprocal* *rank* (MRR, the probability that a relevant tune is returned within the first ten answers) and *mean highest false match* (MHFM).

In its current implementation, the MTRI system accepts microtonal queries that are specified in a text-based format. Any polyphonic music in the collection is first converted to a series of individual melody tracks, so that the query can be compared against each of these. The pieces are first converted into a standard form for matching. The system can also be used for tuning-system queries.

The current prototype uses a dynamic-programming technique for matching, but a system for release would use an index look-up approach based on melody fragments in order to speed up queries. The current test collection consists of MIDI files downloaded from the internet and a small number of *saléndro* pieces that have been manually transcribed. We hope to expand this with the aid of automatic transcription.

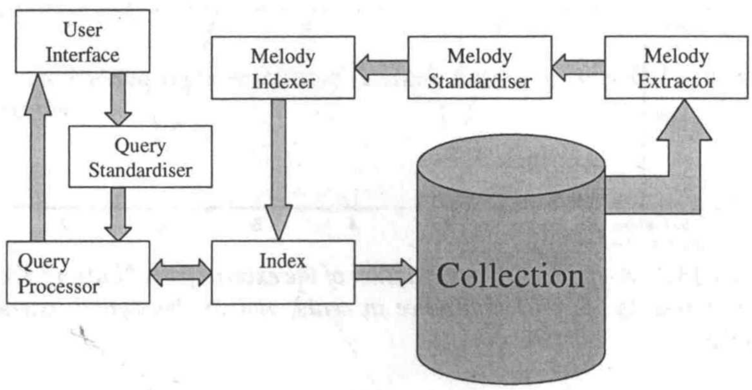


Figure 13.1. The architecture of the MTRI system. For the MTRI microtonal matching system the melody and query standardizers convert melodies to microtonal standard forms. Currently microtonal matching is against standard forms of all pieces in the collection, but a working system would use an index for increased speed.

For microtonal matching purposes, we developed the *exact microtonal interval* standardization (Suyoto and Uitdenbogerd, 2004). In this standardization, every note is represented by the difference between its pitch (measured in cents) and the previous note. The first note serves as the base and is not represented. This makes the standardization transposition-independent. A ten-note (nine-interval) example sequence from a live performance of "Lutung Kasarung" (Monash University, 14 September 2000, under the direction of Lili Suparli) using this standardization procedure is

424 70 −494 424 70 −494 424 70 0

The second note is 424 cents higher than the first, the third is 70 cents higher than the second, the fourth is 494 cents *lower* than the third, and so on. Figure 13.2 graphs the standardization sequence. ("Lutung Kasarung" is a dance-drama in the Sundanese tradition.)

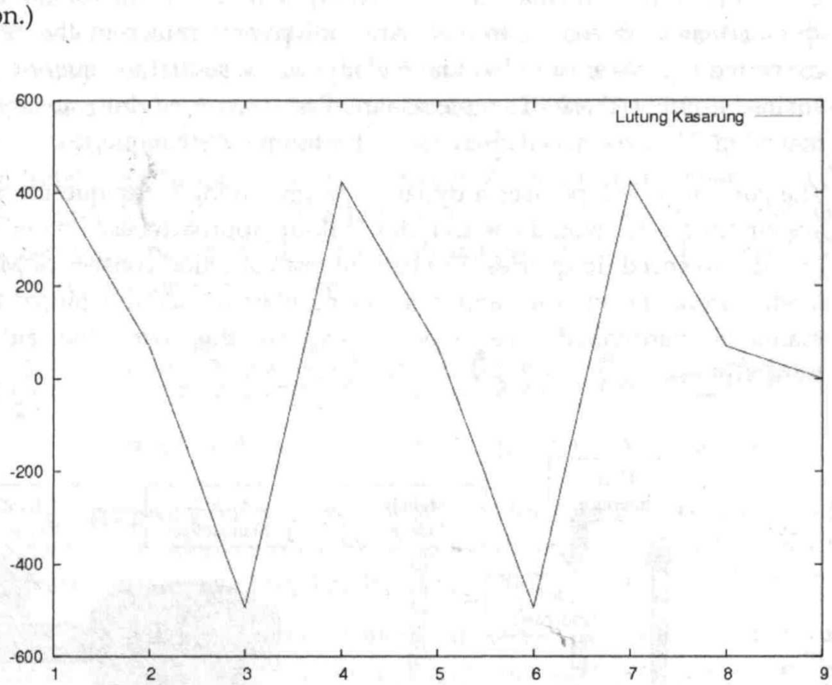


Figure 13.2. A graphical representation of the excerpt from "Lutung Kasarung." The vertical axis represents the pitch difference in cents, and the horizontal axis shows the note-event number.

13.2.3 Implementation

To answer a query, we compare it to standardized versions of the pieces in the database using a dynamic-programming approach. For example, if there were a query for "Lutung Kasarung" in the 12-tone system (Figures 13.3, 13.4),

400 100 -500 400 100 -500 400 100 0

it would be a better match than

470 130 -450 470 130 -450 470 130 0

As we can see from Figures 13.3 and 13.4, the separation between the curves in Figure 13.4 is wider than the one in Figure 13.3. Obviously, a piece such as the one shown in Figure 13.5 is likely to have a low rank in the list of matches.

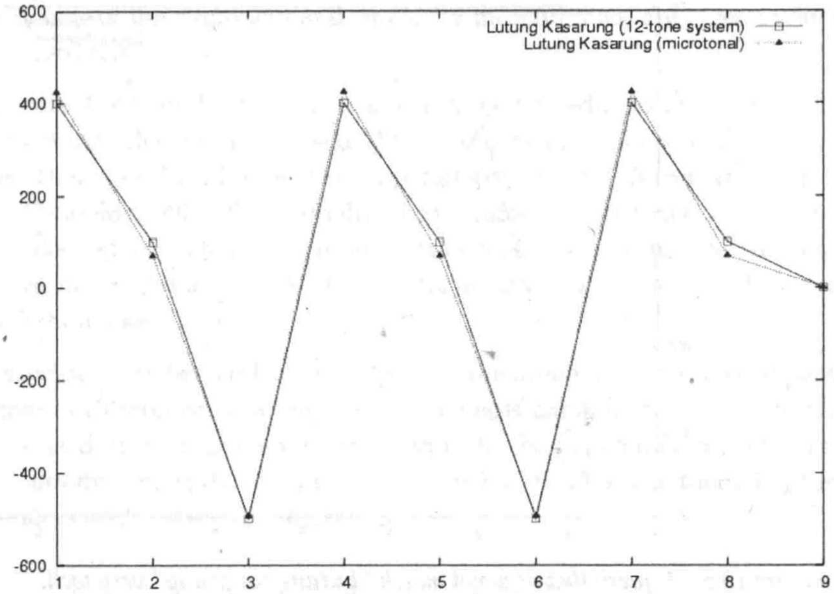


Figure 13.3. Twelve-tone representation of "Lutung Kasarung" matched against the original microtonal version.

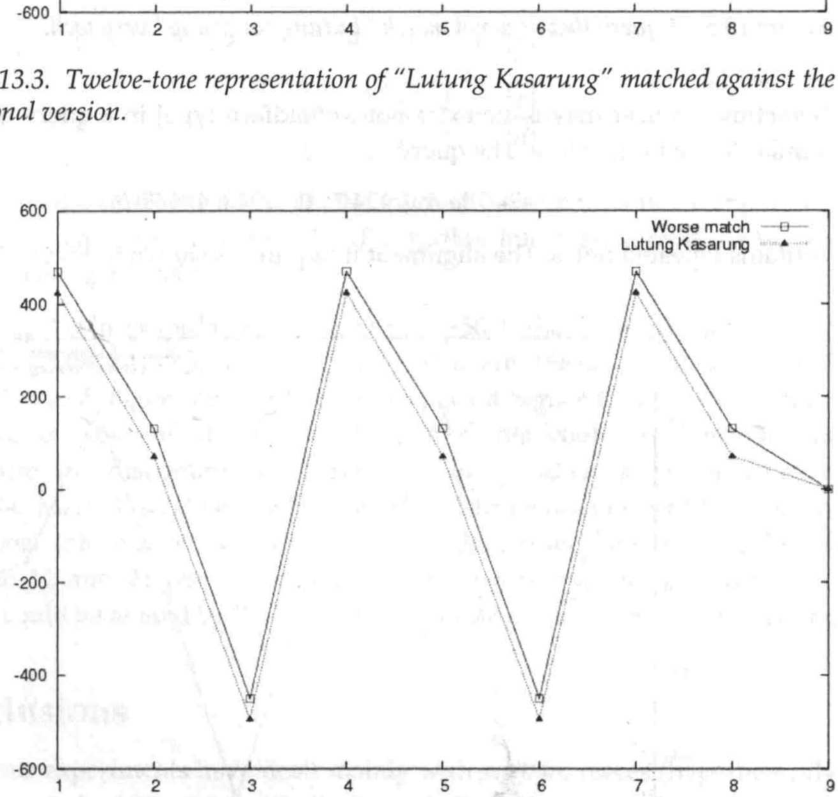


Figure 13.4. A somewhat worse match with respect to tuning.

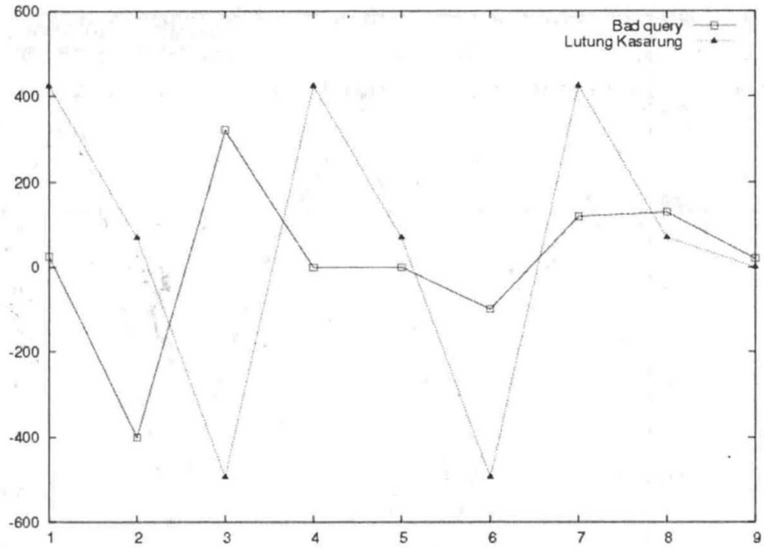


Figure 13.5. A query that does not match "Lutung Kasarung" very well.

Sometimes, a user may issue extra notes [boldface type] in a query that still sounds similar to the target piece. The query

424 70 −494 424 70 **0** −494 **0** 424 70 0

contains repeated notes. The alignment it requires is shown in Figure 13.6.

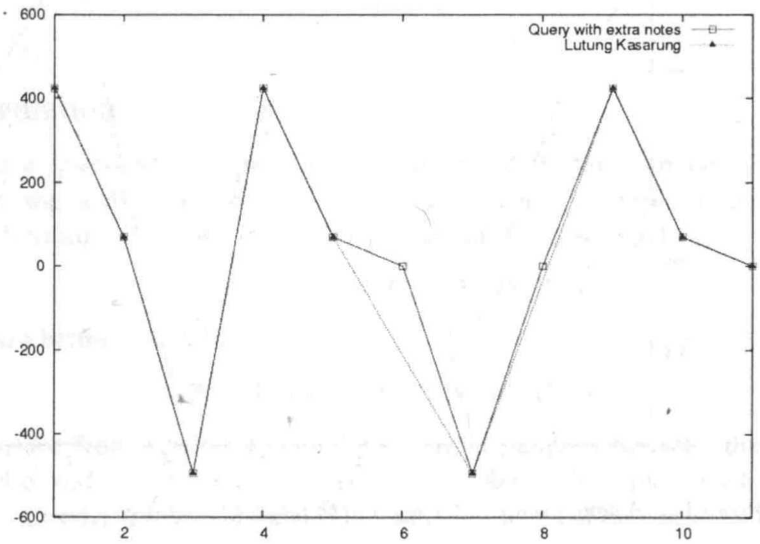


Figure 13.6. A query with repeated notes. The matching process ensures that alignment between the queries can take place despite repeated notes.

Further details of the alignment and similarity measurement are given in Suyoto and Uitdenbogerd (2004).

To determine how similar two tunes are, a variant of edit-distance (an approximate string-matching algorithm) is used. Every character difference is called an *edit operation*. There are three kinds of edit operations: *mismatch* (*substitute*), *insertion*, and *deletion* (Gusfield, 1997). Traditionally, a favorable edit-distance score is the minimum number of operations required to transform one string into the other. This scheme is not suitable for exact microtonal interval sequences, because many mismatches are likely to occur.

In our scheme, matches and mismatches are rewarded or penalized depending on the degree of difference between the two symbols compared. An insertion/deletion penalty is used in the alignment process, typically for matching a melody to a similar one with omitted or repeated notes, such as in Figure 13.6. The following formula is used to calculate the reward/penalty score:

$$w = \begin{cases} -25 & ; \left|\dfrac{i}{d}\right| \geq 10 \\ 25 - \left\lfloor \left(\dfrac{i}{d}\right)^2 \right\rfloor & ; \left|\dfrac{i}{d}\right| < 10 \end{cases}$$

where w is the similarity score, i is the interval (between the two compared symbols) in cents, and d is approximately 12. (For further information on d, please see Suyoto and Uitdenbogerd [2004]).

Applying the insertion/deletion penalty of −50.0, the similarity scores of the pairs shown in Figures 13.3, 13.4, 13.5, and 13.6 are 198.00, 85.00, 25.00, and 100.00 respectively. A higher similarity score implies a higher rank. It was noted that the presence or absence of insertion/deletion points could influence the optimum procedure for discrimination. Different scoring schemes can produce different MHFMs. MHFM is valuable when there is no difference in MRR. The contribution of durational information was slight. The original tests were run on 50 queries of between 12 and 21 notes. Six of the queries were microtonal. We recognize that results could be altered by the use of larger datasets and a greater number of queries.

13.3 Conclusions

While our experiments have dealt mainly with *saléndro* pieces (hypothetically within a larger standard Western collection), we believe that the techniques of microtonal matching can be extended to apply to other scales and tunings. Our MTRI system is an experimental prototype that allows different microtonal matching approaches to be tested. Preliminary experimental results published elsewhere (Suyoto and

Uitdenbogerd, 2004) showed that the concept was feasible. However, larger collections are needed to allow better testing of matching techniques.

The ability to retrieve microtonal pieces from a collection based on a variety of queries is likely to be needed by a very small number of users. In some cases special technology is probably not required, but for melody queries on large collections, or queries on underlying tuning systems, microtonal retrieval systems may prove to be useful.

References

Bainbridge, David (1997). *"Csound"* in *Beyond MIDI: The Handbook of Musical Codes*, ed. E. Selfridge-Field (Cambridge, MA: The MIT Press), pp. 111–142.

Gusfield, Daniel (1997). *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*. Cambridge University Press, Cambridge, UK.

Sapp, Craig Stuart, Yi-Wen Liu, and Eleanor Selfridge-Field (2004). "Search Effectiveness Measures for Symbolic Music Queries in Very Large Databases," *ISMIR 2004*.

Schaffrath, Helmut (1992). "The retrieval of monophonic melodies and their variants: Concepts and strategies for computer-aided analysis" in Alan Marsden and Anthony Pople (eds.), *Computer Representations and Models in Music*, (London: Academic Press), pp. 95–110.

Schaffrath, Helmut (1997). "The Essen Associative Code" in *Beyond MIDI: The Handbook of Musical Codes*, ed. Eleanor Selfridge-Field (Cambridge, MA: The MIT Press), pp. 343–361.

Supandi, A., U. Ngalagena, I. Djunaedi, D. Sain, and R. S. Riswara (1976). *Teori Dasar Karawitan*. 3rd edn. Pelita Masa, Bandung, Indonesia.

Suyoto, Iman S. H., and Alexandra L. Uitdenbogerd (2004). "Exploring Microtonal Matching" in C. L. Buyoli and R. Loureiro (eds.), *ISMIR: International Symposium on Music Information Retrieval*, 5: 224–231.

Uitdenbogerd, Alexandra L. (2002). *Music Information Retrieval Technology*. Ph.D. thesis, School of Computer Science and Information Technology, RMIT University, Melbourne, Australia.

Yampolsky, Philip, et al. "Indonesia" in *The New Grove Dictionary of Music and Musicians* (London: Macmillan), accessed in November 2004.