

I. **Title:** Evaluation of Key Finding Algorithms

II. **Problem Description:**

Determination of the key is a prerequisite for any analysis of tonal music. As a result, extensive work has been done in the area of automatic key detection. However, among this plethora of key finding algorithms, what seems to be lacking is a formal and extensive evaluation process. We propose this first step in the evaluation of key-finding algorithms at the 2005 MIREX.

There are significant contributions in the area of key finding for both audio and symbolic representation. This evaluation process will consider algorithms in each domain separately. Algorithms that determine the key from audio should be robust enough to handle frequency interferences and harmonic effects caused by the use of multiple instruments.

III. **Possible Participants:**

Symbolic Key-Finding

[likelihood of entering]

1. Tuomas Eerola (ptee@cc.jyu.fi) and  
Petri Toivainen (ptoivai@cc.jyu.fi) [high]
2. Arpi Mardirossian (mardiros@usc.edu) and  
Elaine Chew (echew@usc.edu) [high]
3. David Temperley (dtemp@theory.esm.rochester.edu) and  
Daniel Sleator (sleator@cs.cmu.edu) [high]
4. Olli Yli-Harja (yilharja@cs.tut.fi),  
Ilya Schmulevich (is@ieee.org), and  
Kjell Lemström (kjell.lemstrom@cs.helsinki.fi) [high]
5. Craig Sapp (craig@ccrma.stanford.edu) [moderate]

Audio Key-Finding

6. Ching-Hua Chuan (chinghuc@usc.edu) and  
Elaine Chew (echew@usc.edu) [high]
7. Emilia Gómez (egomez@iua.upf.es) and  
Perfecto Herrera (perfecto.herrera@iua.upf.es) [high]
8. Steffen Pauws (steffen.pauws@philips.com) [high]
9. Ozgur Izmirlı (oizm@conncoll.edu) [moderate]
10. Yongwei Zhu (ywzhu@i2r.a-start.edu.sg) and  
Mohan Kankanhalli (mohan@comp.nus.edu.sg) [low]

#### IV. Evaluation Procedures:

The following evaluation outline is a general guideline that will be compatible with both audio and symbolic key finding algorithms.

##### Test Set

The test set we propose to use will consist of 30 second excerpts of pieces for which the keys are known. For example, symphonies and concertos by well-known composers often have the keys stated in the title of the piece. The excerpts will typically be the beginnings of the pieces as this is one part of the piece for which establishing of the global and known key can be guaranteed.

##### Input/Output

The input to the system should be some musical excerpt (either audio or MIDI) and the output should be a key name, for example, C major or E flat minor.

##### System Calibration

It is reasonable to assume that each key finding algorithm will have its own set of parameters. The creators of the system should pre-determine the optimal settings for the parameters. The participants will be provided with training data that they may use in determining the optimal settings. The training data will be randomly selected and a representative subset of the actual test data that will be used in the evaluation process.

##### Submission

The participants will provide the evaluation committee a copy of their system as well as the output from their algorithm for the training data. This will serve as a way to test how the algorithm performs in the evaluation environment.

##### Evaluation

The error analysis will center on comparing the key identified by the algorithm to the actual key of the piece. The key of the piece is the one defined by the composer in the title of the piece.

We will then determine how 'close' each identified key is to the corresponding correct key. Keys will be considered as 'close' if they have one of the following relationships: distance of perfect fifth, relative major and minor, and parallel major and minor.

A correct key assignment will be given a full point, and incorrect assignments will be allocated fractions of a point according to the following table:

Relation to correct key	Points
Same	1
Perfect fifth	0.5
Relative major/minor	0.3
Parallel major/minor	0.2

The team with the highest total score will be the designated winner. Further details of how each system performed within each genre will also be provided.

### Comments

Many excellent suggestions were made in the review process. Some of the ideas included: using actual audio files from recordings for the audio portion of the contest, employing other metrics used in information retrieval literature, using test data from a wider variety of genres, and considering the detection of key modulations.

As this is a first attempt at evaluating key-finding across different systems employing a variety of algorithm combinations, we have opted to keep the evaluation procedure as simple and streamlined as possible. The results of this contest will lay the groundwork from which we can expand the techniques for key-finding evaluation.

## V. **Relevant Test Collections:**

### Symbolic Data

*MIDI Collections:* MIDI data are an event-based representation of music. It provides a numeric representation of the pitch and onset/offset time and velocity for every event in a musical piece. Classical Archive website (<http://www.classicalarchives.com>) provides more than thirty thousands full length classical music files by more than two thousands composers in MIDI format. All the files are presented with full name, and composer. Also, most of files state the key clearly. Music by different composers may be used to test the range of the algorithm.

### Audio Data

*Synthesized MIDI:* Audio data can be generated by synthesizing the MIDI data proposed above. By using the same data for both the symbolic and audio key-finding methods, we will be able to evaluate and compare both approaches. It should be noted that even though synthesized MIDI is a simple alternative to actual audio, it is an appropriate approach for an evaluation where we are consider both audio and symbolic algorithms. Also, this controlled method eliminates possible tuning issues that are sometimes present in recorded audio.

Audio-from-MIDI data can be synthesized using either software or hardware. The software synthesizers include freeware such as Winamp and commercial software such as Cakewalk. The hardware synthesizers, for instance, a Roland XV5080, can receive MIDI commands and use built-in synthesizers to produce more realistic sound.

### Test Data

The test data can be obtained from the Classical Archive website (<http://www.classicalarchives.com>). This site provides a large collection of classical music. Examples of pieces with labeled keys appropriate for the test data set include, but are not limited to, the following:

*Pieces from the Baroque period:*

- Bach (<http://www.classicalarchives.com/bach.html>) – Keyboard Works, Chamber Works, and Orchestral Works.
- Vivaldi (<http://www.classicalarchives.com/vivaldi.html>) – Concerti and Chamber Works.

*Pieces from the Classical period:*

- Handel (<http://www.classicalarchives.com/handel.html>) – Orchestral Works, Keyboard Works, and Chamber Works.
- Haydn (<http://www.classicalarchives.com/haydn.html>) – Keyboard Works, Chamber Works, and Orchestral Works.
- Mozart (<http://www.classicalarchives.com/mozart.html>) – Keyboard Works, Symphonies and Concertos, and Chamber Works.
- Early Beethoven (<http://www.classicalarchives.com/beethovn.html>) – Piano Works, Symphonies, Concertos, and Chamber Works.

*Pieces from the Romantic period:*

- Late Beethoven (<http://www.classicalarchives.com/beethovn.html>) – Piano Works, Symphonies, Concertos, and Chamber Works.
- Brahms (<http://www.classicalarchives.com/brahms.html>) – Keyboard Works, Chamber Works, Concertos and Orchestral Works.
- Chopin (<http://www.classicalarchives.com/chopin.html>) – Piano Works.