U N I V E R S I T Y O F T A R T U
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
Institute of Computer Science

Anna Aljanaki

Automatic musical key detection

Master's thesis (30 ECTS)

Supervisor: Konstantin Tretyakov, M.Sc.

Autor: ..... "....." mai 2011
Juhendaja: ..... "....." mai 2011

Lubada kaitsmisele
Professor: ..... "....." mai 2011

TARTU 2011

Contents

1 Introduction **7**

2 Music Theoretical Background **11**
2.1 Basic terms . . . 12
2.2 Key and mode . . . 14
2.3 Musical traditions of the world . . . 15
2.3.1 Indian raga . . . 17
2.3.2 Arab music . . . 17
2.3.3 Chinese music . . . 18

3 Acoustical background **19**
3.1 Basic terms . . . 19
3.2 Harmonic series . . . 20
3.3 Equal temperament and MIDI numbers . . . 21

4 Related work **23**
4.1 Pitch class profiling . . . 24
4.2 Tree model . . . 25
4.3 Spiral Array . . . 26
4.4 Key detection methods applied cross-culturally . . . 26

5 Key detection **29**
5.1 Feature representation . . . 29
5.1.1 Pitch class profiles . . . 29
5.1.2 Interval distribution . . . 30
5.2 Determining the key . . . 33
5.2.1 Computing template profiles . . . 33

6 Evaluation **35**
6.1 Data . . . 35
6.1.1 Modes . . . 35
6.1.2 Symbolic dataset . . . 37
6.1.3 Acoustical dataset . . . 37
6.2 Evaluating results . . . 39
6.3 Experiments . . . 39
6.3.1 Experiments on the symbolic dataset . . . 40
6.4 Acoustic dataset . . . 42
6.5 Comparison by mode . . . 43

Summary **45**

Resümee (eesti keeles) **47**

References **48**

Appendices **52**

Acknowledgments

First of all, I would like to thank my supervisor Konstantin Tretyakov, who has put a lot of time and effort into guiding my research work. I'm also grateful to Peeter Vähi for his advice and help on musical aspects. Finally, I thank Andreas Ehmann, who provided access to MIREX dataset, used in this thesis.

Chapter 1

Introduction

For a long time appreciation and analysis of music have been regarded as activities, only amenable to human beings. In recent times, the situation has started to change: computerized approach is needed where earlier manual solutions sufficed. The reason lies in the increasing amounts of music, that both music theorists and average users are exposed to today, in large due to web projects like last.fm and allmusic.com.

Music Information Retrieval (MIR) is a new rapidly developing interdisciplinary research field (for details see, for instance, website of International Society for Music Information Retrieval [Ism]). MIR attempts to apply computational methods to music research. Here, research progress is stipulated on the one hand by user demand, and on the other hand by requirements of the music industry. The needs of an average user include, for example, automatic arrangement of a personal collection by artist or style, music recommendations based on samples, playlist construction. Professional musicians, in turn, are waiting for solutions in the field of automatic accompaniment and score generation. The music industry needs methods for identifying audio samples in a large database (acoustic fingerprinting).

Automatic key detection is a step helpful for most of these tasks. It is particularly useful in classification by genre, detecting chords and their functions in tonality, automatic accompaniment and audio-to-score transcription. Key and mode belong to the main characteristics, that shape a piece of music. They are associated with certain emotions, stylistic belonging, epoch, nationality and sometimes even particular authors. For example, if we tell a musician that a piece of music written in the bebop scale is going to be played, he will anticipate a jazz piece.

For a trained musician determining the key and mode of musical piece, composed in a familiar musical tradition, is usually not a difficult task. For a computer, though, it is quite challenging. It involves automatic transcription and requires understanding of how human perception of tonality works. Although this problem has been a subject of research for a long time, the prevalent majority of the works concentrated only on major and minor tonalities, characteristic of Western classical and popular music. There are two reasons

Name of the song	**Actual key**	***Mixed in Key***	***Rapid Evolution***
Placebo, Black Eyed	Fis major	Fis major	Fis major
Cranberries, Animal Instinct	E minor	E minor	*C major*
Franz Ferdinand, Outsiders	D minor	D minor	*D major*
Scriabin, Piano concert	Fis minor	Fis minor	Fis minor
Big Joe Turner, Sun riser blues	G blues	*E minor*	*E minor*
Milt Jackson, Blues mood	Es blues	*Es minor*	*Gis major*
King Crimson, Fracture	Cis whole tone	*A major*	*A major*

Table 1.1: Comparison of two harmonic-mixing programs.

for this. Firstly, Western music has probably the strongest commercial influence internationally. Secondly, classical and popular music have quite strong and unambiguous tonal implications. Hence, many musical genres have been marginalized: blues music, based on pentatonic-like set of scales, jazz with its bebop and whole-tone scale, most of traditional folk music. The latter has been gaining importance in recent years and penetrating our culture through soundtracks and festivals such as, for example, WOMAD [WoMA] and Glasperlenspiel (Estonia) [Gla].

The spread of alternative styles of music (and hence, alternative musical modes) is largely disregarded by the MIR community. It is neglected by commercial software as well. Table 1.1 provides a comparison of results, obtained by two harmonic mixing[1] programs: the commercial *Mixed in key* and the open-source *Rapid Evolution*. Tonalities in italicized letters are misclassified. From this preliminary evaluation we can see that, although both programs perform very good on rock, pop and classical tracks, they can't handle blues and jazz music. There is virtually no support for these modes.

In this thesis we have developed and evaluated an algorithm for key detection that supports these modes. Our approach is based on the most basic musical properties that are present in all of the world's music: hierarchy establishment through pitch duration and interval proportion. The algorithm was tested both on symbolic and acoustic datasets.

[1] *Harmonic mixing* is a technique, used by DJs to create smooth transitions between songs. It involves automatic key detection.

Thesis organization. The thesis consists of 7 chapters. In the current chapter we have explained the goals of our research work. Chapter 2 (Music Theoretical Background) gives a brief description of the music theoretical questions that are necessary to understand the rest of this work. It also provides an overview of the world's musical traditions and their modes. Chapter 3 (Acoustical background) deals with the description of relevant acoustical concepts coming from physics and computer science. In Chapter 4 (Related work) we provide an overview of the existing approaches to automatic key detection. Chapter 5 (Key detection) explains the principles of the algorithm, developed in the current work. Chapter 6 (Evaluation) describes experiments performed on two datasets.

Chapter 2

Music Theoretical Background

> The diatonic scale wasn't
> invented, it was discovered.
>
> *Anton Webern*

Music is sometimes called a universal language of the world. Indeed, music is a form of art, present in virtually every culture on Earth. Musical language is not a symbolic one, it conveys a universally understandable emotional message, which appeals to every human being, regardless of race and nation. But when we come to a definition of music, we are bound to discover, that this is a surprisingly difficult question, and various nations would give a different response. Music is profoundly connected to its culture of origin, to spiritual practices, traditions, and, of course, musical instruments. Music can play ritual role, as in the case of a national anthem or a funeral march, and it can serve for entertainment. In the XXth century traditional music of the world expanded beyond its original realm of ethnomusicology. International musical cultures, along with their native instruments, their modes and manner of performance have spread around the globe. In 1980s the new term emerged to refer to non-Western folk music outside of its natural context — *world music*.

> *In a word, this* world music *is traditional music repackaged and marketed as popular music. This world music, too, owes its origins to the 1980s, when the executives of record companies and advertising specialists determined that popular music from outside the Anglo-American and European mainstreams needed a distinctive name. During the 1980s, the record industry toyed with a few other names* — worldbeat, world fusion, ethnopop, *even* tribal *and* new age — *but by the 1990s, it was* world music *that enjoyed by far the greatest currency.* [Boh02]

In this work we decided to explore the music of the world in the above understanding. Thus, our main target is folk music adapted to Western

ears; traditional music, which can be performed with conventional musical instruments, such as piano or guitar.

The reason for this limitation is simple: this is the way, how we encounter different musical traditions most often and this is the form, in which this music is mostly available and mostly confused with Western music.

2.1 Basic terms

Note — a musical sound, characterized by pitch (corresponding to sound wave frequency) and duration.

Pitch class — a set of all pitches that are a whole octave apart. For instance, pitch class C consists of all Cs that are one octave apart. Figure 2.1 shows all the notes of a piano keyboard, that belong to pitch class C.

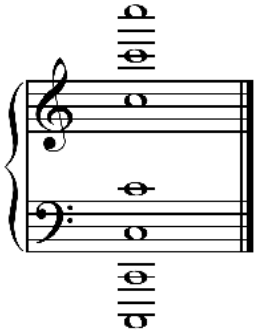


Figure 2.1: C pitch class on piano.

Octave — a musical interval between two notes having frequency ratio 2:1. Notes which are an octave apart are given the same name and are perceived as having the same tone. This phenomenon is called octave equivalence.

Enharmonic equivalence — equivalence of notes with different names. Notes such as G♭, F♯ and E𝄪 refer to the same pitch and are thus enharmonic equivalents. They have different diatonic functions and in a just temperament they would have different tuning.

Tonality (Key) — a musical system, which has a tonal center (named "tonic") and a hierarchy of strong and weak notes. Tonality and key can be regarded as synonyms. Examples of tonalities are *A minor* or *C♯ major*. In major and minor tonalities the most important steps are I, IV and V — *tonic*, *subdominant* and *dominant* respectively. Figure 2.2 illustrates the three main triads of *A minor*, which are built on these steps.

Modulation — change of tonality for a musical piece or a part of it. Very short modulation (less than a phrase) is called tonicization.

Transposition — the shifting of all notes in a compositions by a constant interval. As a result, the tonality of the composition also shifts by the same

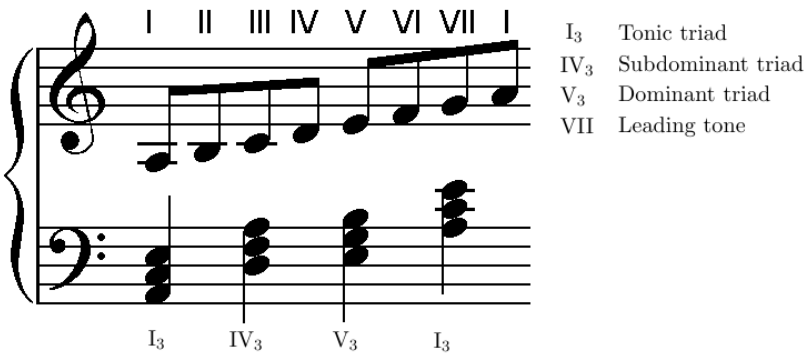


Figure 2.2: Minor key

interval.

Mode — a musical concept, which involves an ordered sequence of notes and melody patterns. Unlike key, mode doesn't define a starting note (tonic). Mode can be expressed via its constituent intervals.

For instance, the *major mode* can be expressed as the following sequence of intervals: tone, tone, semitone, tone, tone, tone, semitone.

Circle of fifths — a circular representation of relationships among the twelve pitch classes and the associated major and minor keys. The keys that are close on circle of fifths, share many common notes. They are perceived as close keys. It is easier to modulate to a close key, than to a distant one.

Relative key — two keys, usually major and minor, that have the same signatures. For instance, *A major* and *F♯ minor* are relative keys. Such keys consist of the same set of pitches, but have different hierarchy between them.

Interval — a combination of two notes with a particular frequency ratio. The most common intervals are called *second, third, fourth, fifth, sixth or seventh*. Prefixes are such as *perfect*, *diminished*, *augmented*, *major* or *minor* are used to indicate modified versions of some intervals. The size of an interval can be measured as a frequency ratio (for instance, a perfect fifth corresponds to the frequency ratio $3 : 2$) or in a system of *cents*:

$$n = 1200 \cdot \log_2 \frac{\text{frequency}_2}{\text{frequency}_1}\,.$$

100 cents correspond to one semitone (a minor second) in the equal temperament.

Melody type — a set of melodic patterns used in a composition. Melodic types are predecessors of scales and are used in non-Western folk music.

Phonic structure — organization of musical sounds. Can be *monophonic*, if music consists of a single line and is performed either in unison, or an octave apart, and *polyphonic*, if multiple lines of music are performed simultaneously. Western classical music is traditionally polyphonic. Some musical modes (mainly quarter-tonal) are built in such a way that chords

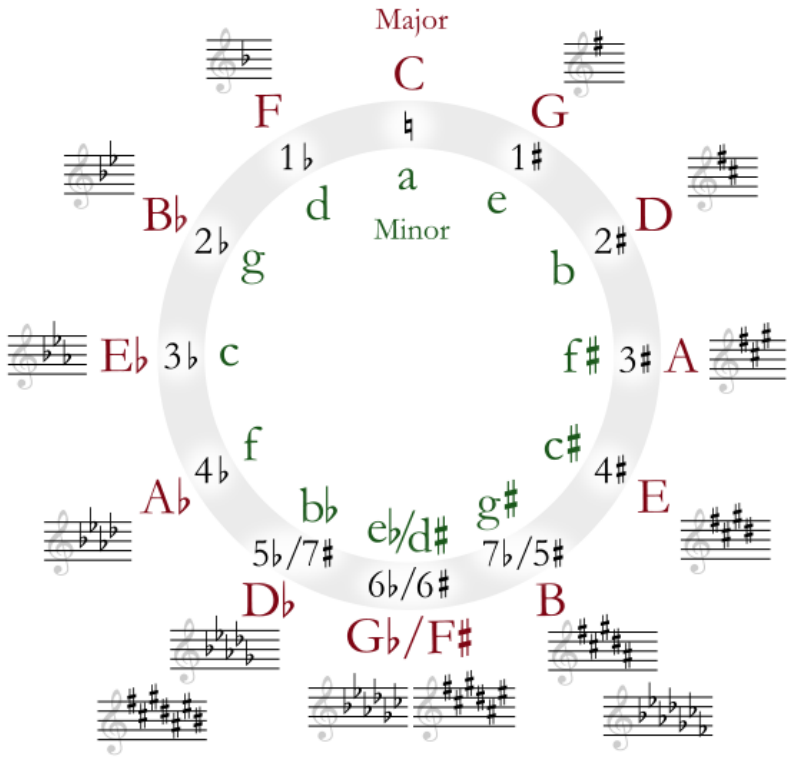


Figure 2.3: Circle of fifths with corresponding key signatures. [Wika]

are not pleasant-sounding. Such music is usually monophonic, or has very rudimentary polyphony. This is true for some modes used in this thesis.

2.2 Key and mode

The variety of existing musical modes is limited only by the constraints of human ear, which is capable of perceiving sound from 20 Hz to 20 000 Hz and distinguishing between two consecutive sounds differing by around 10 cents [Ben07]. In every musical culture there is a specific set of modes based on a historically developed *tuning system*.

> *The term* tuning system *denotes the entire collection of pitch frequencies commonly used in a given musical tradition. Tuning systems are culturally determined. Our ears become accustomed to the tuning system of the music we hear on a regular basis.* ⟨...⟩ *The basis for most tuning systems around the world is the* octave. *In the most commonly used European tuning system (equal-tempered tuning), the octave is divided into twelve equal parts. In the Thai classical music tradition, however, the same octave is divided into only seven equal parts. The tuning systems common to some traditions use more than thirty discrete pitches within a*

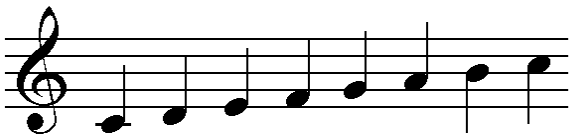


Figure 2.4: Diatonic scale.

single octave. [MS06]

In some musical traditions modes may consist only of as few as two or three pitches [MS06]. For instance, didjeridoo — a wind instrument, developed by indigenous Australian population — is capable of producing only one sound. Music of Australian aborigens consists usually of nature-imitating sounds, accompanied by various drums. The most common instrument, found in many indigenous cultures on earth is a mouth harp. Its playing range also consists of only one sound. [MS06]

In this work, however, we will be interested in less exotic modes. We shall consider scales ranging from five to seven pitches. Five-note scales are called pentatonic, six-note — hexatonic. The Western music is usually based on heptatonic (seven-note) scales. Diatonic scale is a particular class seven-note scales, which, in the layman's terms, correspond to the "white keys" of the piano. The diatonic scale is the most widespread in the world. In particular, traditional European music is based on it. As compared to other scales, the diatonic scale has very high number of consonant intervals, containing six major and minor triads.

It was not easy to select several modes from the diversity of the world music for the purpose of this thesis. The selection was dictated mostly by the availability and hence, degree of penetration of the studied music.

2.3 Musical traditions of the world

In this section we will shortly introduce the musical traditions of the world, that are not based on a diatonic scale. Table 2.1 presents a particular selection of those.

Of course, the presented list does not cover all of the diversity of the world's musical traditions and modes associated with them. Besides traditional modes, there are many experimental and newly developed ones. For instance, one branch of music developed in the XXth century is not based on any mode at all. Such music is denying pitch hierarchy and is called atonal (or 12-tone serialism). Arnold Schoenberg is the composer mostly associated with it.

As long as we confine our research with music that can be performed in equal temperament, we must automatically exclude some regions, such as Oceania and Sub-Saharan Africa and the gamelan music of Indonesia.

Region	Countries	Description
Oceania	Australia, Papua New Guinea, Polynesia	Music is mostly vocal, accompanied by drums, mouth harps. The most ancient musical traditions of the world are still flourishing in this region.
Africa	Ghana, Zimbabwe, Central Africa, South Africa	Rhythmically complex, vocally challenging, polyrhythmical.
India	India	Has developed a tonal system (*system of thāts*), similar to the Western, which consists of a variety of modes. Scale is usually heptatonic, a subset of 22 possible different pitches.
Indonesia	Java, Bali	Gamelan (ensemble, consisting primarily of idiophones).
Middle East	Egypt, Iran, Turkey, Israel and other Arab countries	Music is highly ritualized. Accepted only in form of Quran recitations. Developed system of "modes" — maqamat. Each has characteristic set of melodic patterns.
East Asia	China, Japan, Korea, Vietnam, Tibet, Thailand	Pentatonic scales.
Europe	European countries, America, Cuba	Heptatonic scales, major and minor mode. Sometimes European folk music, such as flamenco, features other scales, but they are mostly variations of the diatonic scale.

Table 2.1: Musical traditions of the world (based on [MS06]).

Overall, there are four major giants: Indian, Western, Arab and Sino-Japanese tradition.

2.3.1 Indian raga

Indian music and underlying modes can be compared to the Western ones in degree of complexity and theoretical base. Tuning system of India is usually said to be consisting of 22 pitches as opposed to 12 in the West. Indian melodic pattern is called "rāg", and modes are called "thāt".

> *The creation of raga is a highly controlled musical process, with established constructional boundaries — even if it allows for nearly unlimited individual variations within these boundaries. Raga is comprised of several elements, one being tonal material (what might be called a "scale"). These "scales" consist of hierarchy of strong and weak notes, a set of typical melodic figures, and a set of extra-musical associations with such things as moods, times of the day, and magical powers. Ragas are sometimes represented pictorially as individual human beings in miniature paintings called ragamalas.* [MS06]

South-Indian music is different and is called "carnatic". The Carnatic system is, at least on the surface, unusually extensive, because theoretically there are so many possible modes. If you allow for all possible arrangements of the seven pitches (some available in three forms), there are 72 possible mode forms. Practically, only a small number of those are commonly used.

Both North- and South-India music have developed quite complex and outstanding musical traditions. It would have been quite hard to cover them in a dataset, so these remained for future work.

2.3.2 Arab music

The music of the Arab world can be united in one huge group, and the quintessence of it is found in Egypt, which is actually territorially outside Middle East and is situated in North Africa. [MS06]

In Arab music, the closest equivalent of the Western mode would be the *maqam*. Like in Western music, *maqamat* are usually heptatonic and octave-repeating, though some may span two octaves. However, there are also many distinctions. Maqam is usually associated with a certain tonic — the starting note. They can be also transposed, but only to a certain set of tonics. For example, maqam Bayati usually starts from D, G and A. Also, each maqam is associated with mood and prescribes certain melodic development.

Two Arab modes were included in the dataset. One of them (phrygian) is quite frequent in the music of other nations. Being a rotation of the diatonic scale, it is also used in flamenco, in Persian music and in European folk music.

2.3.3 Chinese music

The music of China, Tibet, Korea and Japan has one important feature in common: it is essentially pentatonic. Although East-Asian music can be united in one large group, it demonstrates significant variety in timbre, manner of performance and modes.

Chinese music is constructed similarly to its Western equal-tempered sibling, so it doesn't usually create in Western listeners a sense of "out-of-tunness". It can contain modulations, so all seven pitches can be used in a piece of music, despite the pentatonic base.

Two Chinese modes and two Japanese ones were included in our dataset.

Chapter 3

Acoustical background

Acoustics is a subfield of physics dealing with sound. The purpose of this thesis — key detection — involves interpreting the audio signal and analysing such sound wave properties as frequency and amplitude in order to perform pitch-extraction. Identifying the underlying score from audio can range from a fairly easy task to an extremely difficult problem. It depends on many conditions, such as the quantity of simultaneously sounding notes, tempo, instrumental ensemble.

3.1 Basic terms

Fourier transform — a decomposition of signal into its constituting frequencies. It is based on the fact that any periodic function can be represented as an infinite sum of sines and cosines – the *Fourier series*.

$$S_N f(x) = \frac{a_0}{2} + \sum_{n=1}^{N} [a_n \cos(nx) + b_n \sin(nx)], \qquad N \geq 0$$

Frequency domain — a representation of signal, using frequency rather than time.

Discrete Fourier Transform — transforms the signal from the time domain to frequency domain. Abbreviated as DFT. For an audio signal x, the DFT transforms a sequence of its time-samples $(x_0, ..., x_{N-1})$ to a vector of complex numbers of the same length $(X_0, ...X_{N-1})$, representing amplitude and phase of the different sinusoidal components of the input, according to Formula 3.1

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn}, \qquad k = 0, ..., N-1 \tag{3.1}$$

Fast Fourier Transform (FFT) — an efficient algorithm to compute the DFT with $O(N \log N)$ complexity.

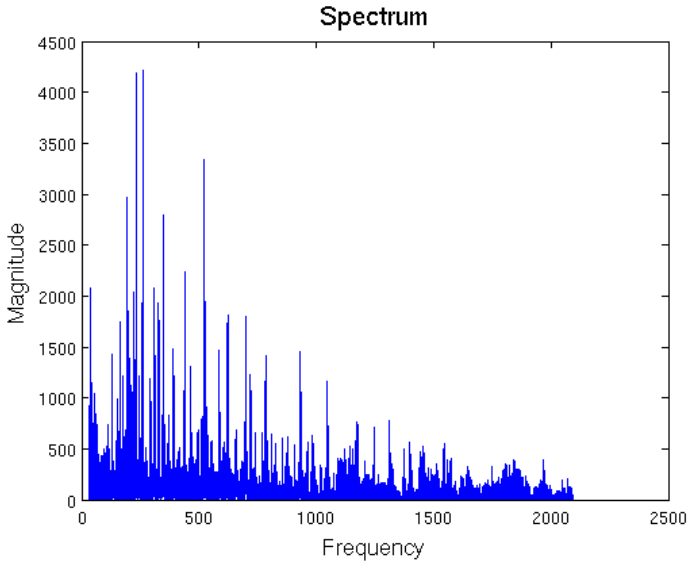


Figure 3.1: Spectrum

Spectrum — the distribution of energy as a function of frequency for a particular sound source (see Figure 3.1).

Chromagram — representation of frequencies that are mapped onto a set of 12 chroma values. Frequencies are assigned to bins according to ideal pitches of equal-temperament and octave-folded.

3.2 Harmonic series

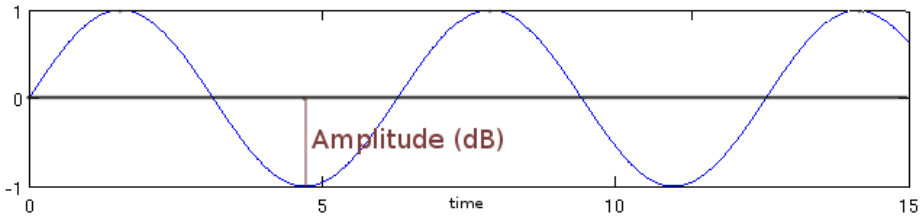


Figure 3.2: Sinusoidal wave

Sound is the vibration of solid, liquid or gaseous substance. Musical sounds are produced by a vibrating source (such as a string or human vocal cords) and transmitted through air. The most basic sound waves are sinusoidal. The loudness of a sinusoidal sound wave is determined by its amplitude. The pitch height is determined by the frequency (see Figure 3.2).

When a note on a string or wind instrument sounds at a certain pitch, let's say with a frequency f, the resulting soundwave is periodic with that frequency. The Fourier theory shows that such a wave can be decomposed into a sum of sinewaves with various phases and frequencies being integer

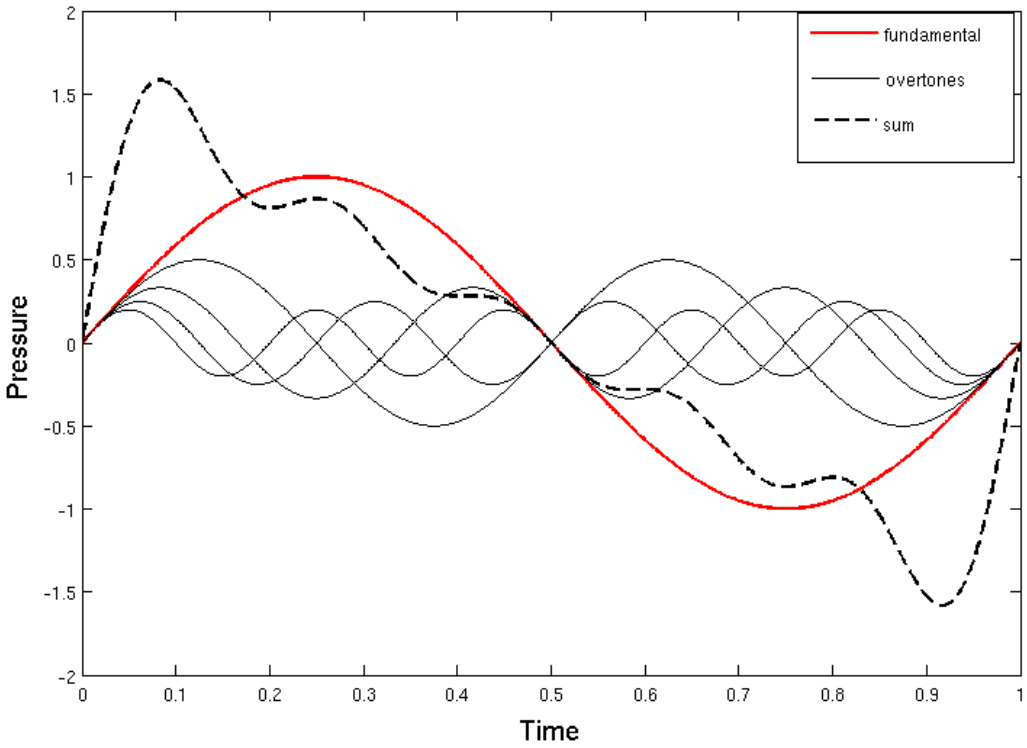


Figure 3.3: Fundamental and its four overtones.

multiples of the base frequency f. The component of the sound with frequency f is called the fundamental. The component with frequency $m \cdot f$ is called the m-th harmonic. [Ben07]

Figure 3.3 illustrates the fundamental wave and four of its harmonics. The dashed line shows the resulting function. An interesting fact to note is that the first five overtones of a sound form a major chord. This notion has been used to justify the "natural" foundation of major tonality.

In real world musical tones, all the possible harmonics need not be present in the sound wave. Some instruments (like clarinet) only produce odd (1, 3, 5 etc.) harmonics. Also, when we hear all the upper partials, we tend to hear the fundamental, even if it is not present. This phenomenon of missing fundamental is used in practice, for example, in production of deep organ tones [RD95].

3.3 Equal temperament and MIDI numbers

An alternative representation of musical sound can be provided by a sequence of discrete notes. The most popular file format supporting such a representation is MIDI (*Musical Instrument Digital Interface*). In MIDI files the notes are represented by their number rather than by their frequency. For

music in equal temperament, the concert pitch $a_4 = 440$ Hz can be used as a reference pitch. In equal temperament, the ratio of two adjacent pitches is always $\sqrt[12]{2}$. Thus, by knowing the MIDI number we can obtain the frequency of the n-th channel using the following formula:

$$\text{freq}_n = \text{freq}_r \cdot 2^{(n-a)/12} \tag{3.2}$$

where freq_n is the desired frequency, freq_r is the reference frequency (440 Hz), n is the pitch number in the MIDI system and a is the MIDI number corresponding to the reference frequency (57).

The reverse transformation (from frequency to a MIDI number) can be done using Equation 3.3.

$$n = (a + \log_2 \frac{\text{freq}_n}{\text{freq}_r} \cdot 12) \tag{3.3}$$

Chapter 4

Related work

The problem of automatic identification of tonality in a piece of music has been researched by musicologists, computer scientists and psychologists for several decades [HM71], [DR93], [S.04], [EP04], [KM09]. However, it is still not quite clear, how do we, humans, discover the tonal center in music. Both experienced musicians and musically non-trained people, provided that they are exposed to a familiar musical tradition, are able to discover tonal hierarchies in music [DE08].

Two basic ways of modelling human key perception have been outlined: the structural and functional approach [H.88]. According to this distinction, structural approaches imply that listeners derive the key from pitch-content material, judging by the prevalent pitches. However, it is quite easy to find examples where reordering of pitches results in a change of tonality perception. For example, on Figure 4.1 the excerpt *a* suggests *C major*, whilst excerpt *b* suggests *G major*. Functional approach adherents argue that tonal hierarchy is based in the first place on the musical context. Both approaches have received empirical support [H.88], [D.89], [Kru90]. These two approaches can also be regarded as complementary [DR93].

Despite the fact that works on key detection are numerous, most of the effort was dedicated to just two Western modes: major and minor (see review on cross-cultural studies on musical pitch and time [C.04]). Moreover, most studies are using Western classical music to evaluate their results, thus narrowing the problem even more. The most influential approaches in key-finding are presented below.

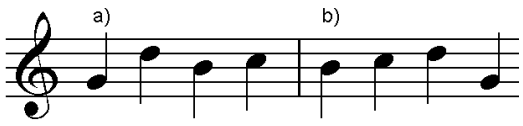


Figure 4.1: Excerpts suggesting different tonalities

4.1 Pitch class profiling

One of the most seminal approaches in automatic key detection is a key-profile method, introduced by Krumhansl and Kessler in [LJ82] and developed later in [Kru90]. This approach is based on an assumption, that a hierarchy of pitches in a key is established through pitch repetition and cumulative duration. According to this theory, the most often occurring note in music should be the tonic. Krumhansl and Kessler proposed a set of key-profiles, representing stability of each pitch-class relative to each key. Their theory is supported both by music theoretical knowledge and experimental results.

The Krumhansl-Kessler approach [LJ82] is often compared to earlier attempts. In 1971, a simple model was created by Longuet-Higgins and Steedman [HM71]. The algorithm proposed by them processed a melody note by note, eliminating all the keys that didn't contain the current note. David Temperley argues [Tem07] that this approach can be expressed in terms of key profiles, using a profile, consisting of 1 and 0 values. Figure 4.2 illustrates such a "flat" major profile.

$$[1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1]$$

Figure 4.2: "Flat" profile for major mode.

The Longuet-Higgins and Steedman approach has many problems. Firstly, the algorithm is likely to end with several possible keys. They proposed to choose the correct key judging by the first note of the piece (it should have been either the tonic or the dominant). Clearly, this is not always the case. Secondly, the chromatic notes prevent the algorithm from selecting the correct key.

The key-profiles of Krumhansl were based on experimental data. The participants of the experiment were presented a key-establishing context, followed by a probe tone. Each participant had to rate, how well the probe tone fitted into previously presented musical context. The results were averaged and key-profiles (vectors of 12 real values) for major and minor modes were created. These profiles did not distinguish between enharmonic notes. They were then rotated around the circle of fifths to obtain profiles for each of the 24 available keys. In the experiment, diatonic pitches received higher values than chromatic ones. Tonic had the highest rating of all, followed by dominant on the second place. These results reflect the basic principles of Western harmony. Figure 4.3 presents the original profiles from the paper [LJ82].

The algorithm, developed in [Kru90], judges the key of a piece by generating its profile. The profile contains 12 values according to cumulative duration of each pitch class in a piece. The key is found by calculating the correlation to each of the 24 original empirical profiles. The maximum

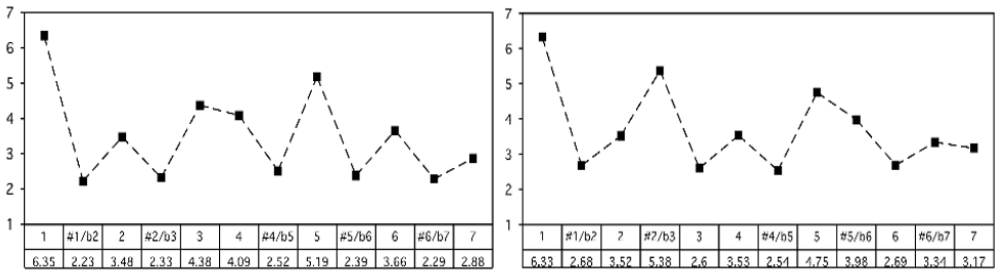


Figure 4.3: Original profiles obtained by Krumhansl and Kessler. [LJ82]

correlated one is selected as the key.

The approach was tested on 48 preludes of Bach's Well-Tempered Clavier and achieved the overall accuracy of 83 % . Secondly, it was tested on preludes by Shostakovich, achieving accuracy of 70%.

In 2001 David Temperley reconsidered and complemented the key-profile method [D.82]. He introduced profiles recognizing enharmonic distinctions, and implemented handling of modulations. To detect modulations, the profiles are calculated not for the whole piece, but only for a passage of music within a moving window. In order to prevent algorithm from modulating too often, a penalty is introduced. Temperley also introduced some other minor modifications, such as different handling of repeating notes.

The modified version of algorithm by Temperley was tested on 46 excerpts from the Kostka-Payne theory textbook "Tonal Harmony" by Stefan Kostka and Dorothy Payne. The version, which recognized enharmonic distinctions, attained 87.4% accuracy [D.99].

Though the Krumhansl-Kessler-Schmuckler approach has had numerous advocates, it also was subject to criticism, because key profiling ignores important aspects in the structure of music — melodic patterns, cadence, harmony, thus discarding many important music-theoretical concepts. In a study of D. Temperley and E. Marvin [DE08], the empirical evidence is reported, that it is not just pitch-distribution that engages in the listener's sense of tonality.

4.2 Tree model

In the work [DFJ03] D. Rizo et al. have proposed a tree model of monophonic symbolic music, which can be used for key finding. In [DIP06] they extended it to polyphonic music. In this model, each melody is represented by a tree, where each leaf is a note. Note duration is encoded in the leaf's distance from the root. Nodes are labelled by pitch height.

The key estimation algorithm estimates all the possible keys for every segment and chooses the most frequent one. Estimation of the key is per-

formed by calculating a rating for each key, applying the principles of tonic triad dominance. On Figure 4.4 the ratings, employed by Rizo, are shown.

Constant	Rate
FULL_TRIADS_I_V	16
FULL_TRIADS	15
2NOTES_TRIADS_I_V	9
2NOTES_TRIADS	8
NOTES_CHORDS_I_V	10
2NOTES_CHORDS	9
TONAL_DEGREES	4
MODAL_DEGREES	3
SCALE_NOTES	2

Figure 4.4: Rate values for set of pitches in the node. [DIP06]

This approach can be applied only to Western music, as far as it uses Western harmony principles.

4.3 Spiral Array

Other influential approach is the Spiral Array model. It was introduced by E. Chew in [E.02]. This model is essentially distributional, like Krumhansl's. In this model, pitches are located in a three-dimensional space, where every key holds its characteristic place. The key of a passage of music can be identified by finding the average position of all pitches in this continuum. The desired key should be the closest to this position. This model, as well as [DFJ03], works on assumption of tonic triad dominance.

4.4 Key detection methods applied cross-culturally

As we have already mentioned, the problem of key detection in non-Western, as well as Western popular music, has not been researched much.

In particular, one of most interesting questions might be, how pitch hierarchy is established in non-Western music.

This question was researched using the probe-tone technique (for atonal [Kru90], North Indian [ABK84] and Balinese [Kru90] music).

North Indian musical tradition is similar to its Western sibling in that it has a well-developed hierarchy of pitches. The most important tone is called Sa. It corresponds to Western tonic. The fifth scale tone, Pa, is considered the second most stable in the system. Experiment showed, that tone duration correlated with their importance.

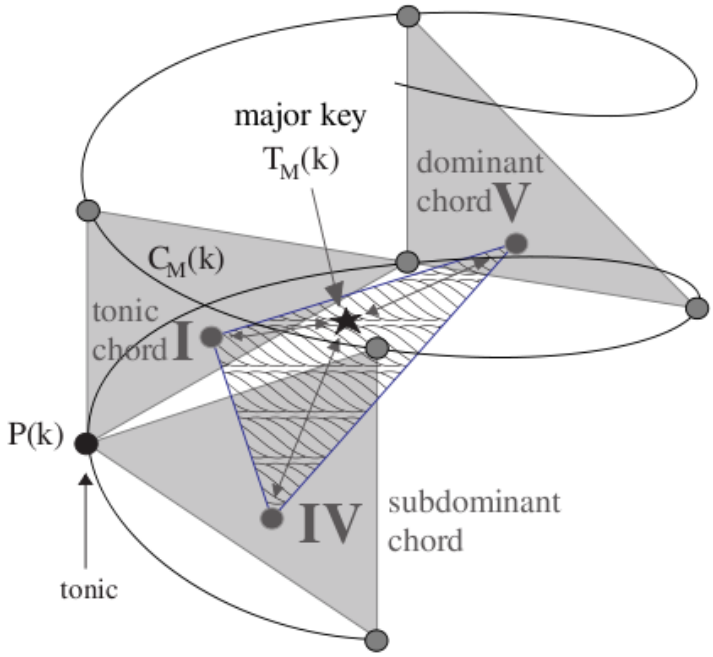


Figure 4.5: Spiral Array model [E.02]

Balinese gamelan music is quite different from Western and would sound very exotic, unusual and out of tune for a new listener. Gamelan ensemble employs two different tuning systems: pelog and slendro. The slendro tuning system is composed of five pitches per octave with intervals ranging from less than 200 cents to 300 cents. Pelog system consists of seven pitches per octave and intervals vary from 100 to 300 cents. The probe-tone experiment in [Kru90] included both Western and native listeners and showed, that both Western and native listeners were capable of deriving tone hierarchies from music, guided by pitch duration.

The same fact was proved for sami yoiks and Finnish spiritual hymns [C.94].

An interesting attempt was made by R.Lewis [RR03]. Lewis is trying to handle 35 different modes (actually less, because many of them are repeating under different names). The author generates a set of action rules, that permit to detect underlying mode and to manipulate music, modifying these rules. Although the goal (modifying music in accordance to user desire) is quite unusual to say the least, the approach is still interesting. Lewis assumes, that it would be possible to judge the key by processing monophonic melodic sequences and comparing them with query-answering-system's standards. His approach is very much resembling Longuet-Higgins and Steedman approach, but it is applied to many different modes. The problem of this approach is the same as its predecessor's. Flat key profiles are incapable

of discriminating between several rivalling tonalities and they can't handle chromatisms occurring, for instance, during tonicization.

Chapter 5

Key detection

> We must ask whether a cross-cultural musical universal is to be found in the music itself (either its structure or function) or the way in which music is made, heard, understood and even learned
>
> Dane Harwood

The question of underlying principles of world music is by no means trivial. Two basic characteristics — *rhythm* and *repetition* — distinguish musical type of sound from noise and speech. All the music of the world is featuring these elements. As far as we are trying to handle many different musical traditions, we should discard approaches, that rely only on Western harmonic principles [E.02, DIP06].

As was already shown in [Kru90] and [C.94], both European, Indian and Finnish music rely on pitch duration to lay emphasis on a certain note. We will exploit an assumption, that this is true for most of the world's music.

Overall, the algorithm for detecting the key of a given composition will consist of two steps. First, we represent a musical piece in terms of its *pitch class profile* and *interval distribution*. Once this is done, we compare the resulting features to the "template" profiles, computed on a pre-labeled training dataset and return the key of the best matching profile.

5.1 Feature representation

5.1.1 Pitch class profiles

Pitch class profiles (PCP), as described in 4 (Related work), are vectors of twelve real values, that express cumulative duration of each pitch class in a piece of music. In our case we were dealing with excerpts no longer than thirty seconds. The excerpts contained no modulations. Therefore, PCPs were calculated for the entire excerpt.

In case of symbolic (MIDI) data, the distribution of notes was weighted by note duration according to Parncutt's durational accent model [R.94].

Durational accent is the following function of event x (a note):

$$\text{duraccent} = (1 - \exp\frac{-d_x}{\tau})^i\,, \tag{5.1}$$

where d_x is duration of x, τ is saturation duration proportional to the duration of echoic note (chosen to be 0.1 seconds in our experiments) and i is an accent index, which is equal to the minimal discriminable note duration (chosen to be 0.3).

Figure 5.1 illustrates the dependence of durational accent on note duration in seconds, according to Equation 5.1.

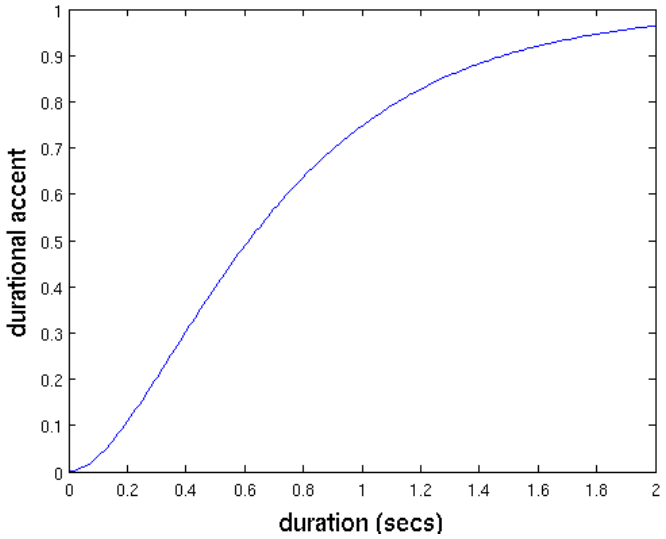


Figure 5.1: Durational accent

5.1.2 Interval distribution

A second useful set of features is provided by the interval distribution. Intervals belong to the basic music structures. The influence of interval distribution on music perception has been studied in [Coo59], [MPE04], [LA04]. Intervals can be measured in the number of semitones. One can distinguish between upward and downward intervals or discard this distinction. On Figure 5.2 the intervals, that can occur within octave in equal temperament, are presented, along with notation, used in this thesis.

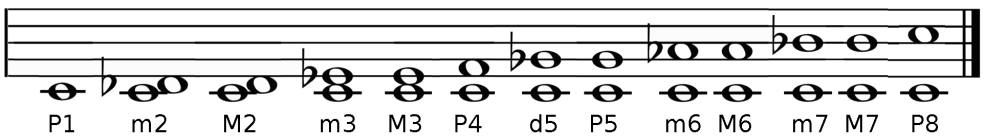


Figure 5.2: Intervals from prime to octave.[Wikb]

In case of symbolic data, we could calculate proportion of intervals present in the melody. For our purposes we define the "melodic" line as the highest note at a given time and extract all intervals between the successive notes.

The intervals are numbered from 0 (prime) to 11 (major seventh). The interval size is computed by the formula 5.2, where num_x and num_y are MIDI numbers of respectively pitch x and y (or would be their diatonic step, in case of audio). Firstly, note that, according to this formula, intervals are regarded module octave. That means that a *major thirteenth* and a *major sixth* are regarded as the same interval. Secondly, the "direction" of interval within the melody is also ignored. As all modes considered in this work are octave-repeating, the presence of primes and octaves is not counted to avoid artefacts of a particular piece.

$$\text{interval}(x, y) = |\text{num}_x - \text{num}_y| \mod 12 \tag{5.2}$$

When dealing with different modes, interval distribution can be an important factor.

Consider two melodies: first in the Pentatonic Minor (Figure 5.3) and second in the Blues Minor mode (Figure 5.4). The melodic intervals of these melodies are given under the score. The modes differ only by one note: the chromatic lowered fifth step in the blues mode. This chromatic step is quite important for the blues style. Thus, passages, including minor seconds, are characteristic of blues music.

In contrast, minor seconds can not occur in the Pentatonic Minor mode. Hence, this mode should contain none or a very small quantity of such intervals. Consequently, only judging by the pitch class distribution, it would have been possible to confuse blues with a pentatonic mode. However, by regarding the interval distribution it is possible to discriminate the two modes.

The pitch class profiles of the two melodies are given on Figure 5.5

Figure 5.6 shows interval distribution of the melodies 5.3 and 5.4. Note that the piece in Pentatonic Minor contains big seconds, small thirds and perfect fourths. There are no small seconds or tritones.

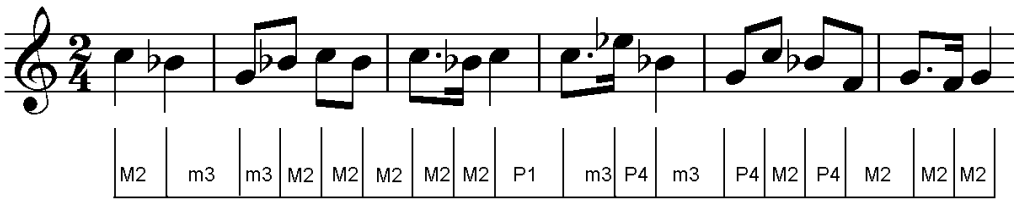


Figure 5.3: A piece in Minor Pentatonic mode

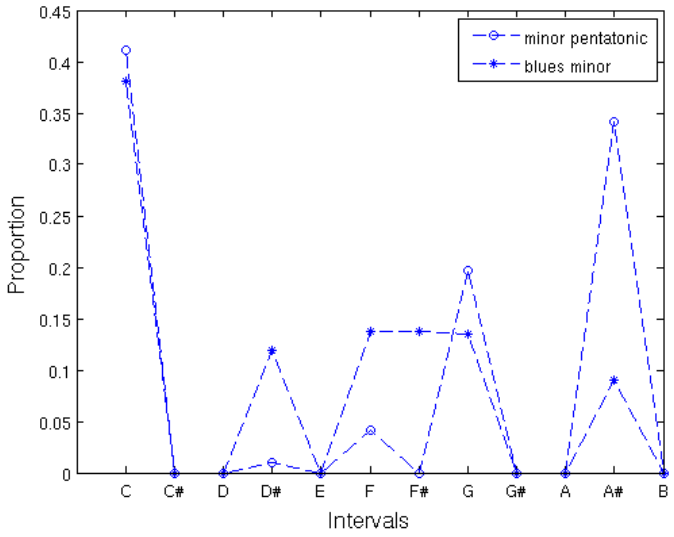


Figure 5.5: PCP of Blues and Pentatonic melody

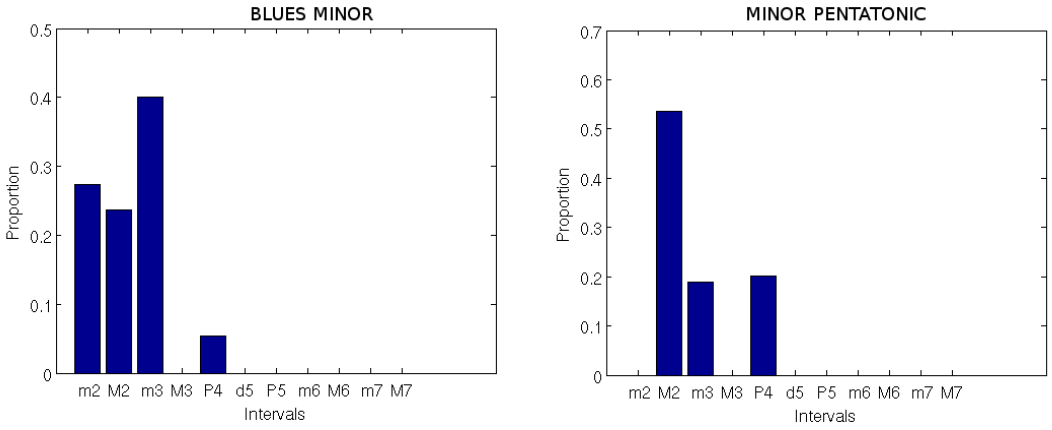


Figure 5.6: Interval distribution in respective melodies

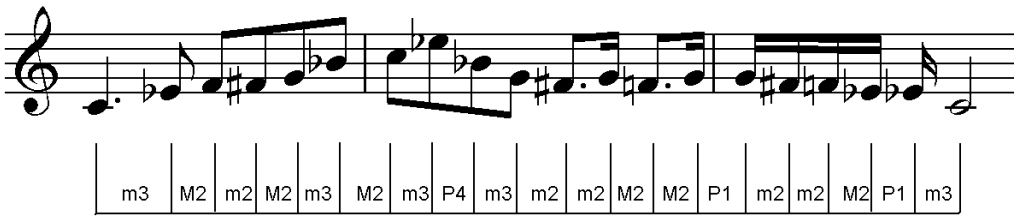


Figure 5.4: A piece in Minor Blues mode

5.2 Determining the key

Once the pitch class profile PCP_x and the interval distribution $\mathrm{interval}_x$ of the composition x are determined, the final step of the algorithm is finding the correct key by comparing the obtained feature vectors to the template profiles $\mathrm{PCP}_{\mathrm{T}}^k$ and $\mathrm{intervals}_{\mathrm{T}}^k$, precomputed for each key k. It is performed using the maximum correlation method as follows:

$$\mathrm{key}_x = \mathrm{argmax}_k \left(\mathrm{cor}(\mathrm{PCP}_x, \mathrm{PCP}_{\mathrm{T}}^k) + 0.6 \cdot \mathrm{cor}(\mathrm{intervals}_x, \mathrm{intervals}_{\mathrm{T}}^k) \right) \quad (5.3)$$

The value 0.6 was determined empirically. The particular choice of classification method is suggested by the previous work of Temperley [Tem07]. In addition, our preliminary experiments with some alternative state of the art machine learning techniques, such as SVM and the Naïve Bayesian classifier showed that the results of all methods are similar.

In the acoustical dataset, musical compositions are represented as analog sound wave signals, which makes it increasingly hard to precisely estimate the interval distribution. Therefore, for this dataset, we only used the first part of Equation 5.3:

$$\mathrm{key}_x = \mathrm{argmax}_k \left(\mathrm{cor}(\mathrm{PCP}_x, \mathrm{PCP}_{\mathrm{T}}^k) \right). \quad (5.4)$$

5.2.1 Computing template profiles

There can be several approaches to compute template PCP and interval profiles. Firstly, they can be determined experimentally, using a group of volunteers [Kru90]. Secondly, they can be derived from data.

In our work we derived the template profiles from the training dataset by averaging the profiles of the labeled compositions of the same key. As is impossible to collect a dataset containing sufficiently many representatives for each key (i.e. for each tonic×mode combination), we have transposed each composition into all possible 12 tonal positions.

Chapter 6

Evaluation

6.1 Data

Public datasets on key detection exist [dat], but they do not contain music for a sufficient variety of modes. Most of such datasets are based on classical Western music [Tem07], [Kru90], [MIR].

In our work we have reused the MIREX MIDI collection [MIR], which contained polyphonic piano pieces in major and minor modes. Music in remaining modes was collected by a musician and manually annotated with the key and mode.

The algorithm was evaluated on two datasets: one symbolic and one acoustical. Every item in the dataset was 30 seconds long, started and ended in the same key.

Name	**Format**	**Phonic structure**	**Number of items**	**Description**
Symbolic	MIDI	Polyphonic	206	Manually labelled MIDIs
Acoustical	WAV	Polyphonic	189	Audio, synthesized using *Timidity*[syn] from MIDI dataset

6.1.1 Modes

Eleven modes were selected (see Table 6.1). The choice was determined by the reasons described in Chapter 2 (Music Theoretical Background). Four modes come from Western music: major, minor, whole tone and minor blues modes. Four modes are pentatonic: two Japanese and two Chinese modes. Two modes are Arabic and the last one comes from the Jewish traditional music.

Mode	**Usage in music**	**Scale in C**
Major	Heptatonic diatonic scale. Wide usage: Western, Indian, Arab music	C D E F G A B
Minor	Heptatonic diatonic. Wide usage: Western, Middle-East, Indian, Arab music. Relative minor of major scale	C D E♭ F G A♭ B♭
Blues	Hexatonic. Characteristic of Blues, close relative to Minor Pentatonic.	C E♭ F F♯ G B♭
Whole tone	Hexatonic. Used in jazz	C D E F♯ G♯ A♯
Jewish	Heptatonic. Used in Jewish, Arab music	C D♭ E F G A♭ B♭
Phrygian	Heptatonic. Used in Arab (maqam kurd), Persian (dastgah shur), Jewish music, in flamenco, and jazz	C D♭ E♭ F G A♭ B♭
Double harmonic	Heptatonic. Used in Arab music (Maqam Hijaz Kar Kurd)	C D♭ E F G A♭ B
Pentatonic Major	Pentatonic. Chinese, Vietnamese music	C D E G A
Pentatonic Minor	Pentatonic. Chinese, Vietnamese music. Relative minor to Pentatonic Major	C E♭ F G B♭
Insen	Pentatonic. Japanese	C D♭ F G B♭
Hirajoshi	Pentatonic. Japanese	C D♭ F G A♭

Table 6.1: Modes used in the experiments

6.1.2 Symbolic dataset

The symbolic dataset consists of 206 MIDI files in eleven different modes. The proportion of different modes is not equal, as it is also not equal in real world. 32 % of MIDIs are in major mode, 31% in minor mode.

We divide all songs in the dataset into three categories:

1. Classical diatonic piano pieces from [MIR].
2. Songs by The Beatles, collected and annotated for this work.
3. All remaining, non-diatonic compositions.

In our experiments we shall test our method on the combination of the first and third category and the first and second category separately. This is to illustrate how the choice of genre for diatonic songs can influence the classification accuracy. It is known that classical compositions are highly tonal and contain a relatively small amount of chromatic notes and modulations to distant tonalities. Rock music is, on the contrary, less tonal and is often confused with blues modes, which is easily seen from our results.

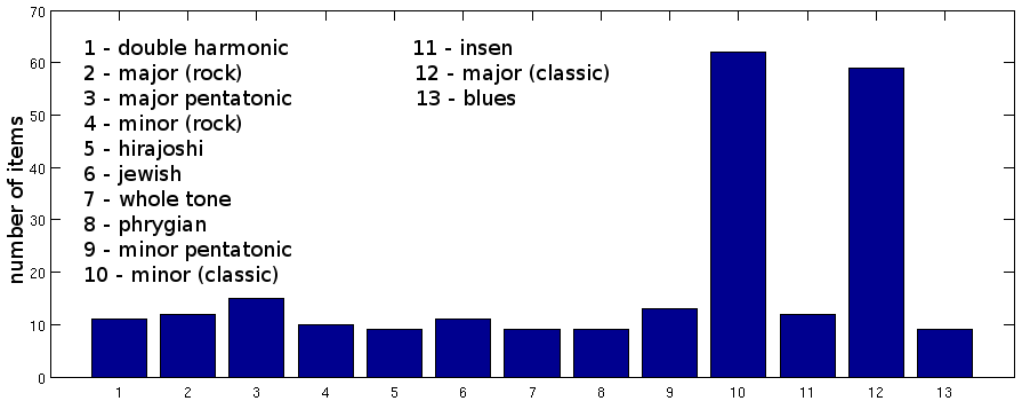


Figure 6.1: Proportion of music in different modes in symbolic dataset.

We extract the PCPs and interval distributions from the MIDI files, as described in Sections 5.1.1 and 5.1.2, and apply our maximum correlation algorithm.

6.1.3 Acoustical dataset

The acoustical dataset consists of audio signals in WAV format. It was generated from the symbolic one. We used Timidity [syn] software synthesizer to convert files MIDI files to WAV. We included first and third categories of songs fro the symbolic dataset. Thus the size of the dataset was 189 songs.

We extracted chromagrams, the acoustic equivalent to PCP profiles, from all songs in the dataset as it will be described below. In theory, it might have

been possible to also extract the interval distribution information, but turned out exceedingly hard to obtain satisfactory accuracy of those features, so we have left them out.

Preprocessing of audio

In order to extract pitch distribution information from the audio dataset, we needed to do several transformations with the audio signal (see Figure 6.3). Firstly, the sound (Figure 6.3a) is converted to the frequency domain (Figure 6.3b) using the Fast Fourier Transform [WW65]. In our work, the frequency range was confined to six octaves from C1 (32 Hz) to C7 (2093 Hz), as shown on Figure 6.2.

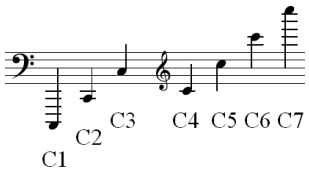


Figure 6.2: C1 to C7

For comparison, the playing range of a piano is from A0 (27.5 Hz) to C8 (4186 Hz).

Secondly, we divide the spectrum into 73 unequal, logarithmically increasing sections, as shown on Figure 6.3b, corresponding to all pitch classes between C1 and C7.

Finally, the amplitudes inside each section are averaged, resulting in a histogram (Figure 6.3c). The histogram is then octave-wrapped, giving a chromagram — a vector of 12 values. Thus, we obtain a representation, analogous to a pitch class profile.

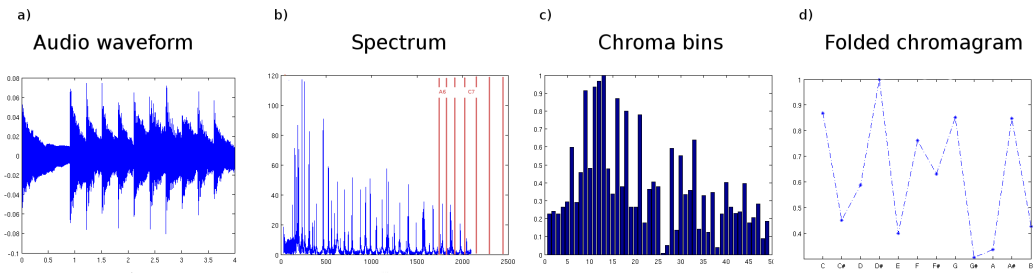


Figure 6.3: Conversion of audio signal to folded chromagram

The code for chromagram extraction was written in MatLab using MIRtoolbox [LT07].

6.2 Evaluating results

As the size of the dataset was rather small, the performance was validated using leave one out cross-validation (LOOCV). In particular, whenever we predict key for a composition, we make sure that this composition is not used in the computation of template profiles.

We present the results in terms of confusion matrices. These are matrices, showing how many items of each class have been (mis)classified for each other class. As there are 132 possible keys in total, the full size of the confusion matrices is 132×132. To compress the representation, we merge the rows and columns of the matrix corresponding to each mode. As a result we obtain a 11×11 matrix (given that there are 11 modes).

Finally, we summarize the results in terms of accuracy metrics. The overall accuracy of the key prediction for compositions in mode m is computed as

$$\mathrm{acc}^m_{\mathrm{key}} = \frac{n^m}{N^m}, \tag{6.1}$$

where n^m is the number of compositions in mode m that had their key predicted correctly, and N^m is the total number of compositions in mode m.

Often the mode would be predicted correctly, even if the key (i.e. the tonic) will be misclassified. To assess the accuracy of mode prediction, we measure

$$\mathrm{acc}^m_{\mathrm{mode}} = \frac{n^m_{\mathrm{mode}}}{N^m}, \tag{6.2}$$

where n^m_{mode} is the number of compositions in mode m that had their mode predicted correctly.

The proportion of cases when the tonic is classified correctly given that mode was predicted correctly is denoted by:

$$\mathrm{acc}^m_{\mathrm{key|mode}} = \frac{\mathrm{acc}^m_{\mathrm{key}}}{\mathrm{acc}^m_{\mathrm{mode}}}. \tag{6.3}$$

Finally, to obtain an overall accuracy rating, we average accuracies for particular modes, giving each mode equal weight:

$$\mathrm{acc}_X = \sum_{m=1}^{K} \mathrm{acc}^m_X, \tag{6.4}$$

where K denotes the total number of modes.

6.3 Experiments

We have performed 4 experiments. First we assess the method on the symbolic dataset. We study two aspects of the method.

1. The importance of the interval distribution.

2. The influence of the type of diatonic music on classification accuracy.

Finally, we test the method on the acoustic dataset.

6.3.1 Experiments on the symbolic dataset

The importance of interval distribution

As we can only use the interval distribution for symbolic (MIDI) data, it is important to estimate the loss in precision, that we incur by not using such features in our further experiments on the acoustic dataset. To estimate it, we compare the results obtained on the symbolic dataset with and without the interval distribution features.

Figure 6.5 presents the *confusion matrix* obtained on the "classical" set (first and third categories), when only the PCP features were used.

Predicted

Actual	1	2	3	4	5	6	7	8	9	10	11	
	3	0	0	0	0	1	0	0	3	0	0	| 1: Blues
	0	8	0	0	0	0	0	0	0	0	0	| 2: Double harmonic
	0	0	6	0	0	0	0	0	0	0	0	| 3: Hirajoshi
	1	0	3	5	0	0	0	0	1	0	0	| 4: Insen
	0	0	0	0	7	0	0	1	0	0	0	| 5: Jewish
	2	0	0	0	1	46	4	1	0	3	0	| 6: Major
	1	0	0	1	0	0	7	0	3	0	0	| 7: Major pentatonic
	1	0	0	2	2	3	0	39	11	1	0	| 8: Minor
	1	0	0	2	0	0	2	2	3	0	0	| 9: Minor pentatonic
	0	0	0	0	0	1	0	0	2	3	0	| 10: Phrygian
	0	0	0	0	0	0	1	0	0	0	5	| 11: Whole tone

Figure 6.4: Confusion matrix by mode classified using PCP (symbolic dataset)

From the confusion matrix we can see, that minor tonalities (both Western and pentatonic minor) are the most difficult to classify. The minor mode is confused with almost any other mode except double harmonic, major pentatonic and whole tone. Moreover, half of the blues songs were mistaken for minor pentatonic (of which it is, indeed, a close relative).

The overall accuracy acc_{key} was 74.5%, whereas mode prediction accuracy acc_{mode} was 74.8%. This small margin is caused by the fact that $\text{acc}_{\text{key|mode}}$ was 99.6%, i.e. whenever a mode is predicted correctly, so is most often the tonic. There were only three cases when the tonic of a correctly detected mode was misclassified. In two of these cases the tonic was confused with the dominant, and in one with a parallel minor key tonic. These mistakes are not grave. Confusion with the dominant key is sometimes regarded as semi-correct [MIR].

Adding the interval distribution improved the overall classification accuracy by 3%, resulting in $\text{acc}_{\text{key}} = 77.5\%$ (see Figure 6.5). This shows that the addition of the interval distribution features does help classification, but not overwhelmingly so.

We see, that both major and minor mode classification benefit from the interval distribution information. The worst predicted mode is still Minor Pentatonic, strangely enough, taking into account that its relative major

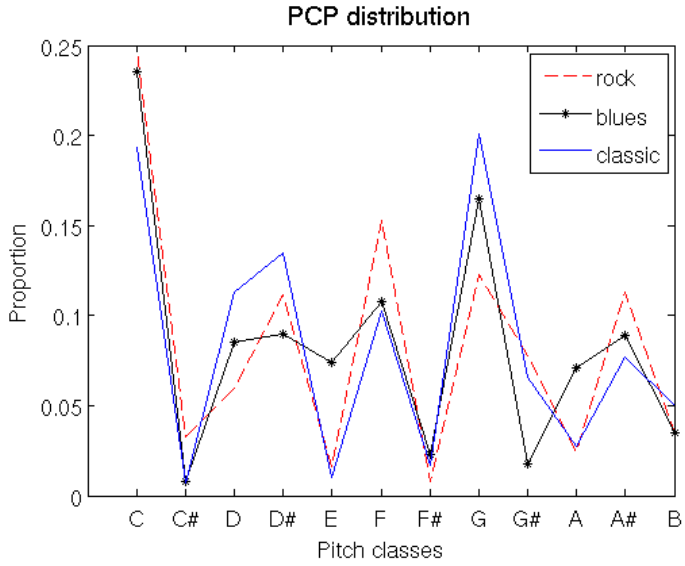


Figure 6.6: Profiles for rock, classic and blues music

is predicted quite well. Taking into account the dataset size, this may be an artefact, connected to the fact, that world music available in MIDI is orchestrated and quite westernised.

Actual \ Predicted	1	2	3	4	5	6	7	8	9	10	11	
1	3	0	0	0	1	1	0	0	2	0	0	1: Blues
2	0	8	0	0	0	0	0	0	0	0	0	2: Double harmonic
3	0	0	6	0	0	0	0	0	0	0	0	3: Hirajoshi
4	0	0	3	6	0	0	1	0	0	0	0	4: Insen
5	0	0	0	0	7	0	0	1	0	0	0	5: Jewish
6	2	0	0	0	0	50	4	0	0	1	0	6: Major
7	1	0	0	0	0	0	11	0	0	0	0	7: Major pentatonic
8	2	2	0	1	1	3	0	41	6	3	0	8: Minor
9	2	0	0	2	0	0	1	2	3	0	0	9: Minor pentatonic
10	0	0	0	0	0	0	0	0	0	6	0	10: Phrygian
11	0	0	0	0	0	0	1	0	0	0	5	11: Whole tone

Figure 6.5: Confusion matrix by mode classified using PCP and interval distribution (symbolic dataset)

The influence of the type of diatonic music

In this experiment we used PCP and interval distribution to classify compositions of the "rock" set (the second and third categories). Results are presented in Figure 6.7. We see, that 3 Beatles minor songs and 4 Beatles major songs were mistaken for blues. The absolute accuracy for the major mode $\text{acc}_{\text{key}}^{\text{major}}$ was 40% and for the minor mode $\text{acc}_{\text{key}}^{\text{minor}}$ was only 14%. In comparison, the same indicators on the "classic" dataset were 77% and 66% correspondingly.

This illustrates the point that diatonic popular rock music can be much more difficult to classify than diatonic classical pieces. Figure 6.6 shows the

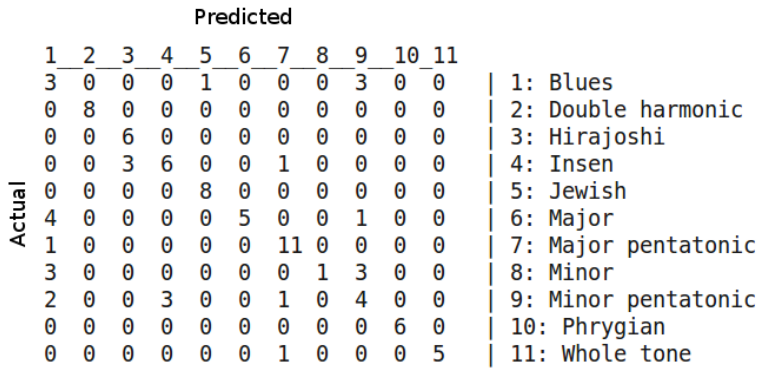


Figure 6.7: Confusion matrix by mode ("rock" subset)

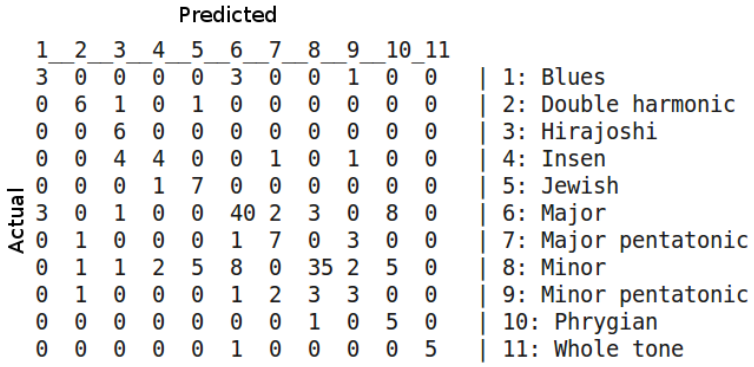


Figure 6.8: Confusion matrix by mode (acoustic dataset)

mean pitch profiles for the rock (The Beatles), blues and classical compositions. All of these genres are pretty similar by their PCPs, moreover, rock correlates slightly better with blues than with classical pieces.

6.4 Acoustic dataset

Figure 6.8 shows results, obtained on the acoustic dataset. The performance on WAV files was worse, than on symbolic data, as it should be expected. The overall accuracy on this dataset was 59%, which is by 17% worse, than on symbolic dataset, although still much better than chance. The mode prediction accuracy $\text{acc}_{\text{key|mode}}$ dropped from 77.5% to 66%.

The kind of misclassifications on this dataset was different. For a correctly predicted mode, the tonic was detected correctly only in $\text{acc}_{\text{key|mode}}$=86% of cases (compare to 99.6% on the symbolic dataset). Figure 6.9 presents the tonic confusion matrix for all cases where mode was predicted correctly. All except four tonic confusion mistakes correspond to confusion with the dominant key.

This is a natural consequence of the way we compute PCPs (Chromagrams) for the acoustic dataset. Indeed, the spectra of the tonic and the

dominant are highly similar, hence the PCP magnitude of the dominant is artificially intensified. Figure 6.10 compares the PCP vectors for *C major* from the symbolic and the audio datasets. Note that in case of audio, there is much less difference among the intensities of the diatonic notes.

Predicted

Actual	1	2	3	4	5	6	7	8	9	10	11	
1	26	0	0	0	0	0	0	8	0	0	0	1: C
2	0	9	0	0	0	0	0	0	2	2	0	2: C#
3	0	0	5	0	0	0	0	0	0	10	0	3: D
4	0	0	0	2	0	0	0	0	0	0	8	4: D#
5	0	0	0	0	11	0	0	0	0	0	0	5: E
6	4	0	0	0	0	3	0	0	0	0	2	6: F
7	0	0	0	0	0	0	4	0	0	0	0	7: F#
8	0	0	0	0	0	0	0	12	0	0	0	8: G
9	0	0	0	4	0	0	0	0	3	0	0	9: G#
10	0	0	0	0	4	0	0	0	0	9	0	10: A
11	0	0	0	0	0	0	0	0	0	0	8	11: A#

Figure 6.9: Confusion matrix by tonic on acoustic dataset

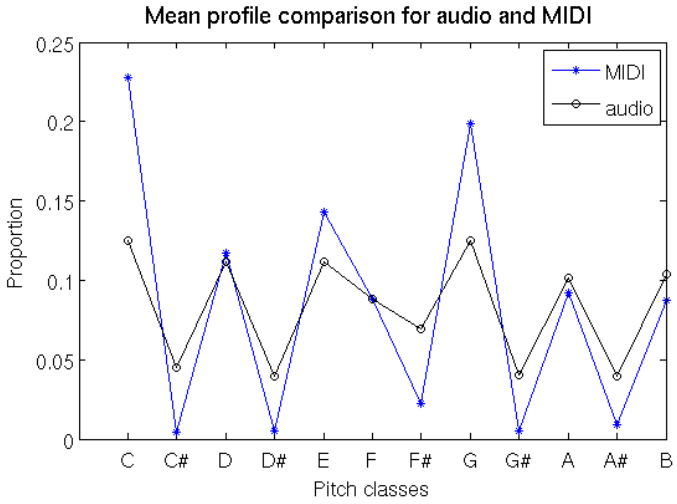


Figure 6.10: Mean profiles for major mode

6.5 Comparison by mode

Figure 6.11 shows accuracy for every mode. We can see that the mode having the least accuracy is minor pentatonic both for acoustic and the symbolic datasets. Overall, the minor modes were predicted with less accuracy. This may be explained by the fact, that most of the dataset modes are musically closer to minor modes (minor pentatonic, diatonic minor, double harmonic, jewish, blues and both japanese modes).

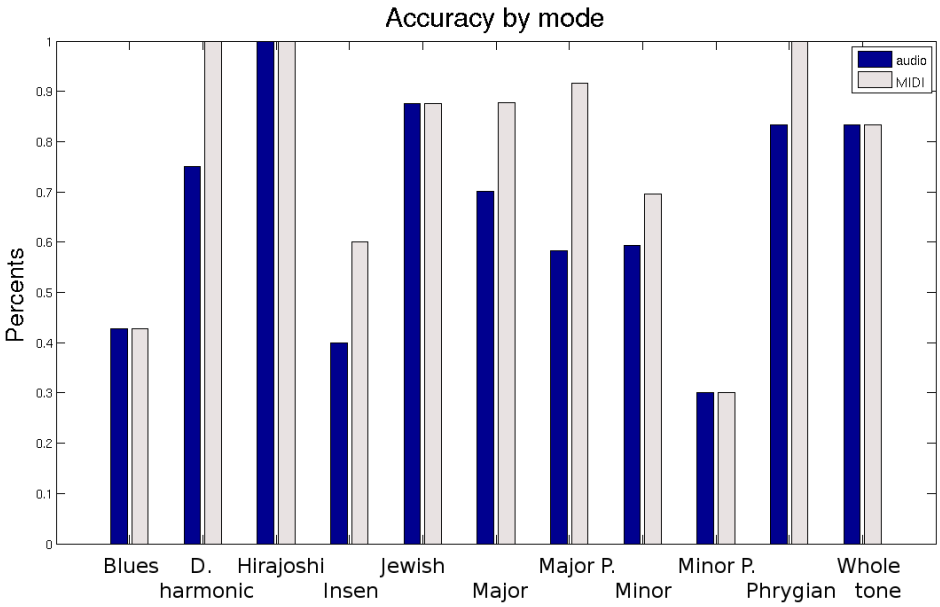


Figure 6.11: Classification accuracy by mode

The largest groups (major and minor classical pieces) demonstrate accuracies of 88% and 70% respectively in the symbolic dataset and 70% and 60% in the acoustic dataset.

Summary

How do we, humans, distinguish between different musical traditions and perceive their tonal hierarchy? And would it be possible for a computer? These questions are still unanswered by ethnomusicology, cognitive psychology and computer science. The studies on tonality perception usually only engage Western music. Hence, a knowledge gap exists in the studies of non-Western music. In this thesis we have proposed a model for tonality estimation, which is capable of handling music coming from various musical traditions and does not require their thorough analysis.

Every musical tradition has an underlying framework, comprising of a *tuning system*, a set of *modes* and *melodic patters*. Modes establish a hierarchy between pitches. In our model we have employed an assumption, that most musical traditions use duration to maintain pitch salience. This hypothesis has been previously proven for some world music [C.94, ABK84].

Proceeding from this assumption, we have proposed an algorithm for automatic key detection. Our method is based on a distributional approach [Kru90]. It involves calculation of pitch class profiles and interval distribution.

The proposed method was evaluated on both symbolic (189 MIDI files) and acoustic (the same MIDI files, converted to WAV) datasets. It achieved accuracy of 74.5% in case of MIDI and 59% in case of audio. Eleven modes were included in the dataset, four of them coming from Western music, four from pentatonic Sino-Japanese music, two Arabic modes and one Jewish.

In the theoretical part of the work we have provided a review on key detection methods and showed that most of existing approaches rely on harmonic theory of Western music. We have also provided an overview of existing musical traditions of the world and explained our musical choices.

This work could be developed further in several directions. The algorithm could be improved to incorporate some harmonic analysis. Our present method is approaching the problem of key detection from a purely structural point of view. Some functional features (such as melodic patterns, cadence, drone sound) could be useful. Such improvements would mean searching for meaningful musical concepts for non-Western music and quantifying them.

The dataset could be complemented with North-Indian music, bebop scales from jazz and the remaining maqamat.

Automaatne tonaalsuse avastamine

Magistritöö (30 ECTS)

Anna Aljanaki

Resümee

Kuidas meie, inimesed, suudame eristada maailma muusikalisi traditsioone ja tajume nende tonaalse hierarhia? Kas see oleks võimalik ka arvuti jaoks? Need küsimused on veel vastamata etnomusikoloogia, kognitiivse psühholoogia ja informaatika poolt.

Tavaliselt tegelevad uuringud tonaalsuse tajumisest ainult Lääne muusikaga. Seega, meie teadmised mitte lääne muusikast on puudulikud. Selles töös meie oleme pakkunud mudeli tonaalsuse avastamiseks, mis on võimeline tegelema muusikaga erinevatest muusikalisest traditsioonedest ilma, et nende põhjalik analüüs oleks nõutud.

Iga muusikalise traditsiooni aluseks on raamistik, mis hõlmab *häälestamis süsteemi*, *heliredeleid* ja *meloodilisi mustreid*. Heliredelid kehtestavad helide hierarhiat. Meie mudel põhineb eeldusel, et enamik muusikalisi traditsioone kasutavad hieraarhia kehtestaniseks helide kestust. See hüpotees oli varem tõestatud maailma muusikal [C.94], [ABK84].

Lähtudes sellest eeldusest, oleme pakkunud algoritmi automaatseks helilaadi avastamiseks. Meie meetod põhineb jaotuslähenemisel [Kru90]. Meetod kaasab heliklasside profiilide ja intervallide jaotuse arvutamist.

Meetod oli hinnatud nii sümboolse (189 MIDI faili) kui ka audio (esimene andmestik, konverteeritud audioks) andmestiku peal. Oli saavutatud 74.5 % täpsus MIDI puhul ja 59 % audio puhul. Andmestikku kuulusid üksteist helilaadi. Neist neli olid pärit Lääne muusikast, neli pentatoonilisest Hiina-Jaapani muusikast, kaks olid Araabia laadid ja üks Juudi laad.

Töö teoreetilises osas oleme andnud ülevaate tonaalsuse avastamise meetoditest ja näitasime, et enamik olemasolevatest lähenemisviisidest tuginevad Lääne muusika harmoonilisele teooriale.

Oleme ka andnud ülevaate olemasolevatest maailma muusikalistest traditsioonedest ja selgitasime meie laadide valikuid.

See töö võiks areneda mitmes suunas.

Praegune algoritm läheneb probleemile puhtalt struktuursest vaatepunktist. Võiks integreerida sellesse mõned funktsionaalsed omadused. See nõuab täiendavaid uuringuid erinevate rahvuste muusikast, nende sisuliste muusikaliste mõistete välja selgitamist.

Andmestik võiks olla laiendatud Põhja-India muusikaga, bebop heliredelitega ja ülejäänud maqam-heliredelitega.

References

[ABK84] Castellano M. A., J. J. Bharucha, and C. L. Krumhansl. Tonal hierarchies in the music of north india. 1984.

[Ben07] David J. Benson. *Music: Mathematical Offering*. Cambridge University Press, 2007.

[Boh02] Philip. V. Bohlman. *World Music. A very short introduction.* Oxford University Press, 2002.

[C.94] Krumhansl C. Tonality induction: A statistical approach applied cross-culturally. 1994.

[C.04] Stevens C. Cross-cultural studies of musical pitch and time. 2004.

[Coo59] D. Cooke. *The language of music*. Oxford University Press, 1959.

[D.82] Temperley D. *The cognition of basic musical structures*. The MIT press, 1982.

[D.89] Butler D. Describing the perception of tonality in music: A critique of the tonal hierarchy theory and a proposal for a theory of intervallic rivalry. 1989.

[D.99] Temperley D. Whats key for key? the krumhansl-schmuckler key finding algorithm reconsidered. 1999.

[dat] Million Song dataset. http://labrosa.ee.columbia.edu/millionsong (last accessed on 23.05.2011).

[DE08] Temperley D. and Marvin E. Pitch-class distribution and the identification of key. 2008.

[DFJ03] Rizo D., Moreno-Seco F., and Inesta J.M. Tree structured representation of musical information. 2003.

[DIP06] Rizo D., Jos I., and De Le P.J.P. Tree model of symbolic music for tonality guessing. 2006.

[DR93] Huron D. and Parncutt R. An improved model of tonality perception incorporating pitch salience and echoic memory. 1993.

[E.02] Chew E. The spiral array: An algorithm for determining key boundaries. 2002.

[EP04] Gomez Elaine and Herrera P. Estimating the tonality of polyphonic audio files: Cognitive versus machine learning modelling strategies. 2004.

[Gla] http://www.erpmusic.com/p_glasperlenspiel.htm (last accessed on 23.05.2011).

[H.88] Brown H. The interplay of set content and temporal context in a functional theory of tonality perception. 1988.

[HM71] Longuet-Higgins H.C. and Steedman M.J. On interpreting bach. 1971.

[Ism] Ismir. http://www.ismir.net/ (last accessed on 23.05.2011).

[KM09] Noland K. and Sandler M. Influences of signal processing, tone profiles, and chord progressions on a model for estimating the musical key from audio. 2009.

[Kru90] Carol L. Krumhansl. *Cognitive Foundations of Musical Pitch.* Oxford Psychology Series, 1990.

[LA04] Cuddy L.L. and Lunney C. A. Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. 2004.

[LJ82] Krumhansl C. L. and Kessler E. J. Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. 1982.

[LT07] Olivier Lartillot and Petri Toiviainen. A matlab toolbox for musical feature extraction from audio. 2007.

[MIR] MIREX. http://www.music-ir.org/mirex/wiki/2005:audio_and_symbolic_key_finding (last accessed on 23.05.2011).

[MPE04] Costa M., Fine P., and Bitti E.R.P. Interval distributions, mode, and tonal strength of melodies as predictors of perceived emotion. 2004.

[MS06] Terry E. Miller and Andrew Shahriani. *World Music: A global journey.* Routledge, 2006.

[R.94] Parncutt R. A perceptual model of pulse salience and metrical accent in musical rhythms. 1994.

[RD95] Berg R.E. and Stork D.G. *The Physics of Sound, 2nd ed.* Prentice-Hall, 1995.

[RR03] Lewis R. and Zbigniew R. Rules for processing and manipulating scalar music theory. 2003.

[S.04] Pauws S. Musical key extraction from audio. 2004.

[syn] Timidity synthesizer. http://timidity.sourceforge.net (last accessed on 23.05.2011).

[Tem07] David Temperley. *Music and probability*. Cambridge, MA:MIT Press., 2007.

[Wika] Wikipedia. http://en.wikipedia.org/wiki/circle_of_fifth (last accessed on 23.05.2011).

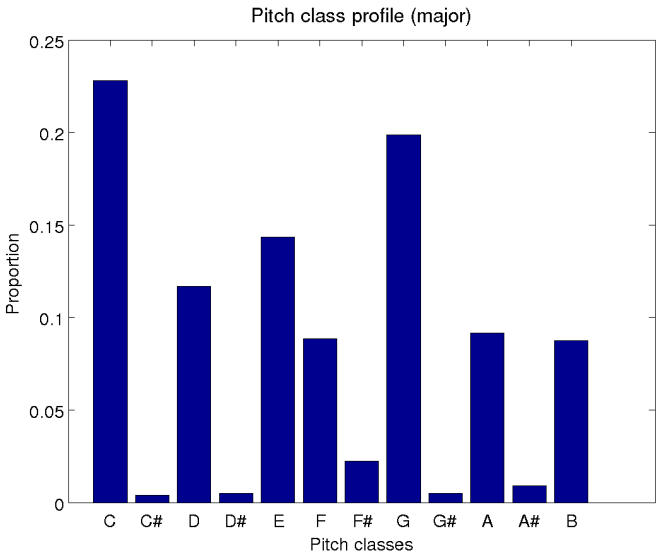
[Wikb] Wikipedia. http://en.wikipedia.org/wiki/interval_%28music%29 (last accessed on 23.05.2011).

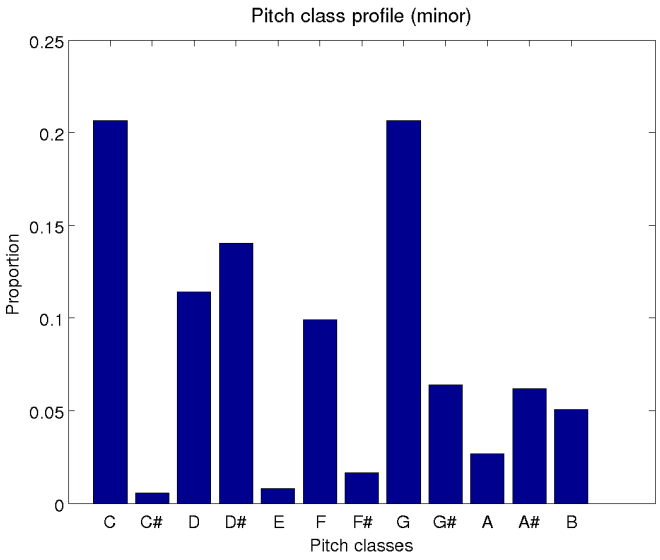
[WoMA] Dance World of Music and Art. http://womad.org/ (last accessed on 23.05.2011).

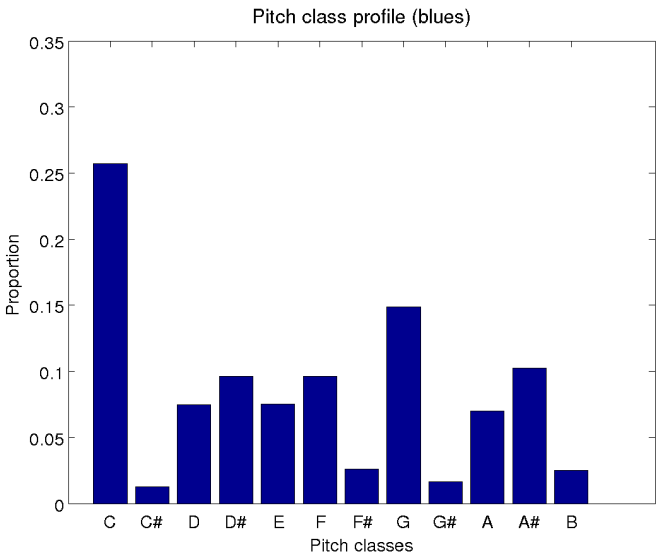
[WW65] Cooley J. W. and Tukey J. W. An algorithm for the machine computation of the complex fourier series. 1965.

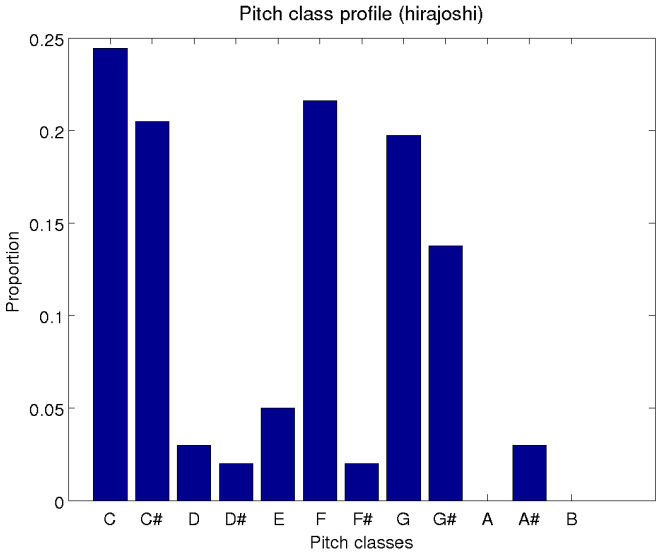
Appendices

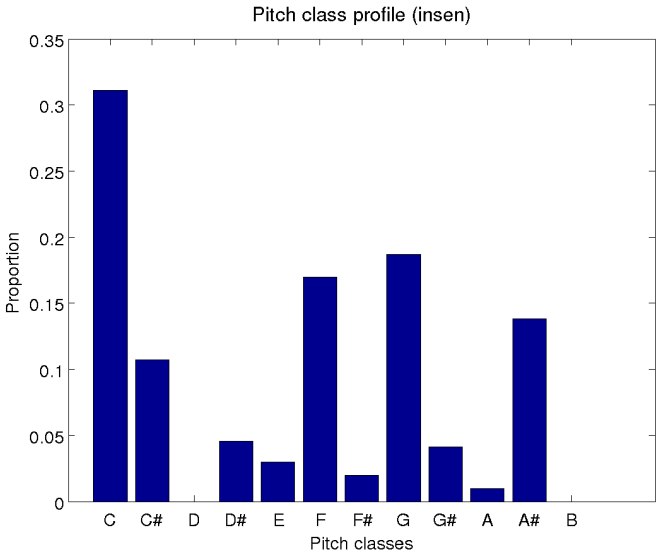
Appendix A. Program code (on a compact disc)
Appendix B. Figures

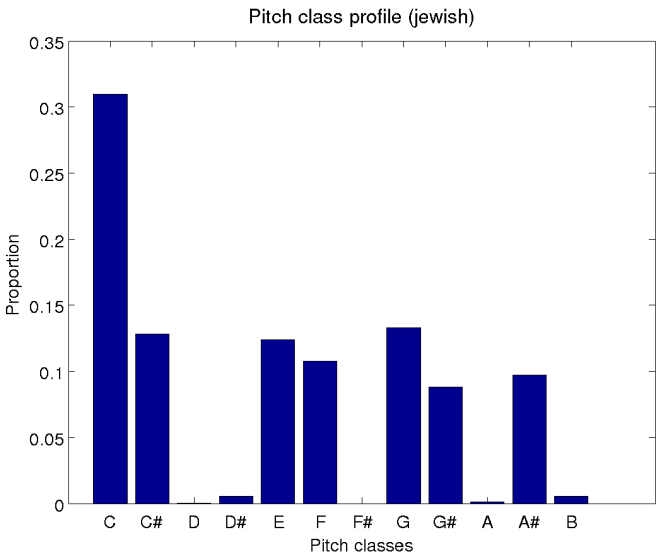


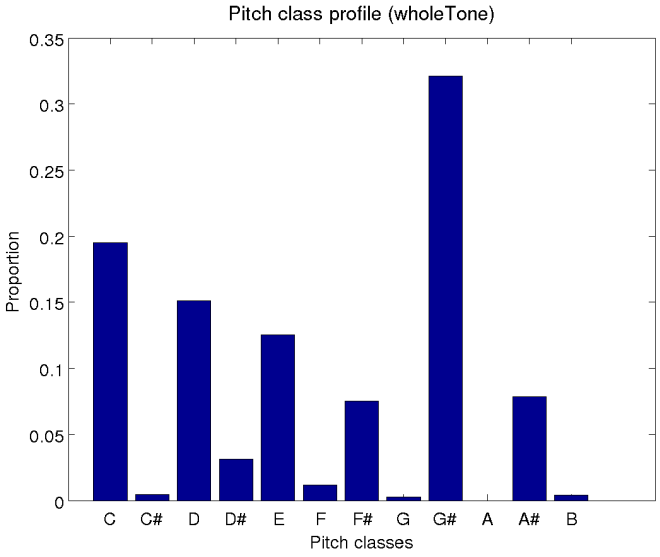


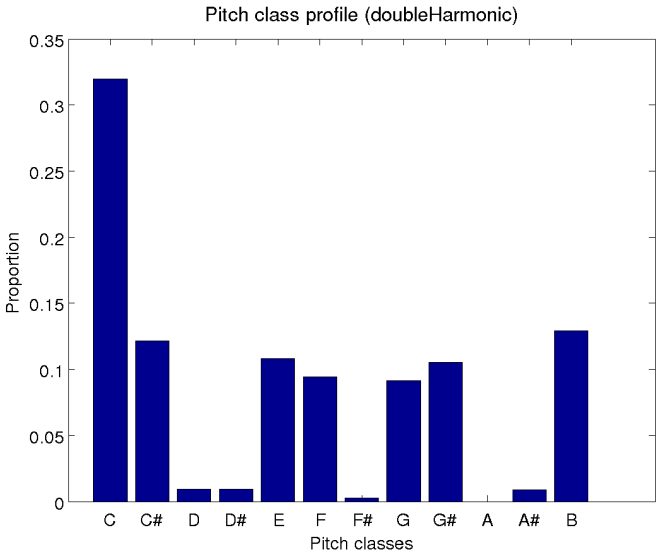

Pitch class profile (blues)
Proportion
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
C C# D D# E F F# G G# A A# B
Pitch classes

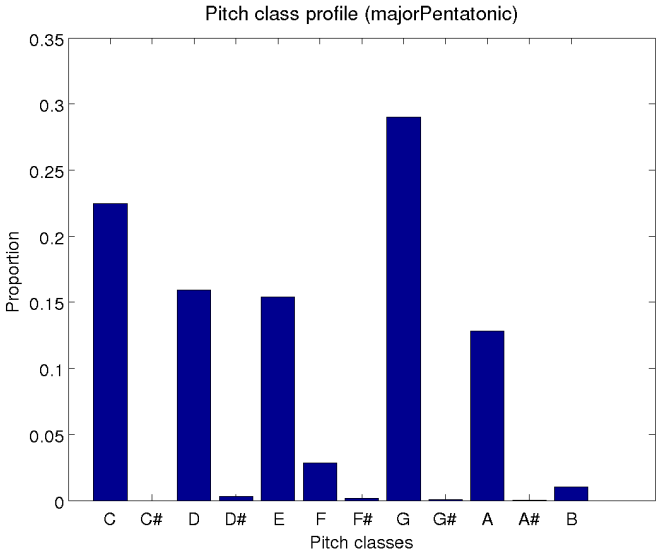

Pitch class profile (hirajoshi)
Proportion
0.25
0.2
0.15
0.1
0.05
0
C C# D D# E F F# G G# A A# B
Pitch classes

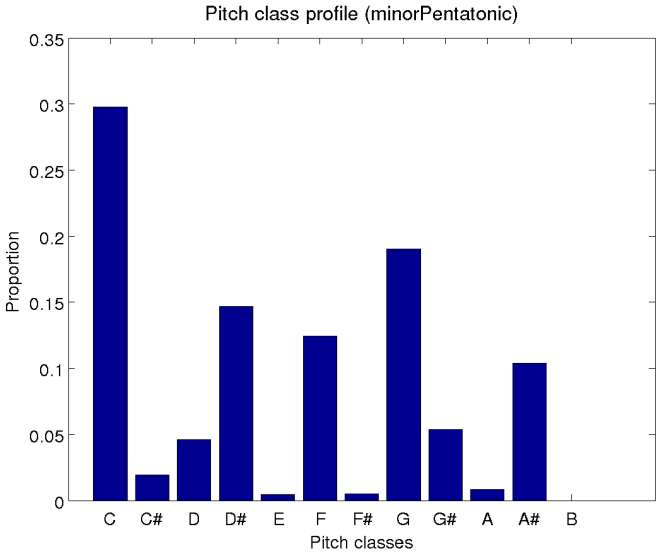

Pitch class profile (insen)
Proportion
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
C C# D D# E F F# G G# A A# B
Pitch classes

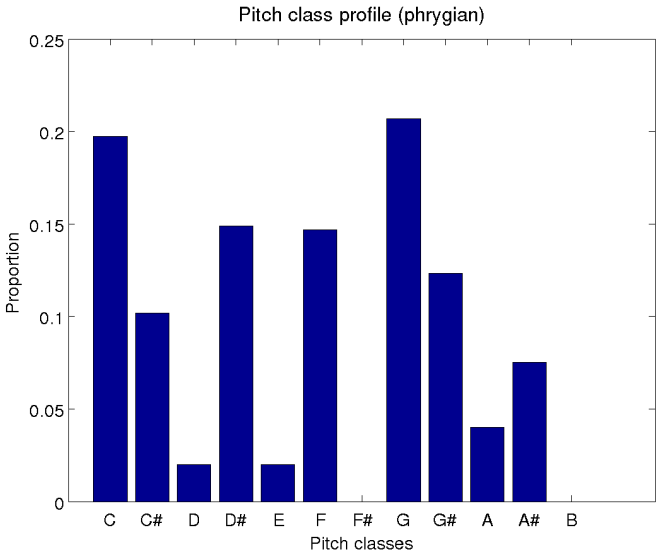

Pitch class profile (jewish)
Proportion
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
C C# D D# E F F# G G# A A# B
Pitch classes


Pitch class profile (wholeTone)
Proportion
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
C C# D D# E F F# G G# A A# B
Pitch classes


Pitch class profile (doubleHarmonic)
Proportion
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
C C# D D# E F F# G G# A A# B
Pitch classes


Pitch class profile (majorPentatonic)
Proportion
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
C C# D D# E F F# G G# A A# B
Pitch classes


Pitch class profile (minorPentatonic)
Proportion
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
C C# D D# E F F# G G# A A# B
Pitch classes


Pitch class profile (phrygian)
Proportion
0.25
0.2
0.15
0.1
0.05
0
C C# D D# E F F# G G# A A# B
Pitch classes