

Estimating the makam of polyphonic music signals: template- matching vs. class-modeling

Ioannidis Leonidas

MASTER THESIS UPF / 2010

Master in Sound and Music Computing

Master thesis supervisor:

Emilia Gómez

Department of Information and Communication Technologies

Universitat Pompeu Fabra, Barcelona



Abstract

This thesis provides an investigation on the problem of scale estimation in the context of non-Western music cultures and more specific the makam scales found in Turkey and Middle-East countries. The automatic description of music from traditions that do not follow the Western notation and theory needs specifically designed tools. We evaluate two approaches that have successfully dealt with same problem in Western music, but approaching the matter without considering Western pitch-classes. To accomplish the task of classifying musical pieces from the 'makam world', according to their scale, we used chroma features extracted from polyphonic music signals. The first method based on template-matching techniques compares the extracted features with a set of makam templates, while the second one uses trained classifiers. Both approaches provided good results (F-measure=0.69 and 0.73 respectively) on a collection of 289 pieces from 9 makam families.

Evaluation measures were computed to clarify the performance of both models, error analyses showed that certain confusions were musically coherent and that these techniques could complement each other in this particular context.

Contents

Contents.....	i
List of Figures	iii
List of Tables.....	iv
Acknowledgements	v
Chapter 1.....	1
Introduction	1
1.1. Motivation.....	1
1.2. Goals and expectations.....	2
1.3. Music Information Retrieval (MIR).....	3
1.3.1. Audio content description	3
1.3.2. Computer-aided ethnomusicology	4
1.4. Thesis structure	6
Chapter 2.....	7
Scientific background	7
2.1. Makam music.....	7
2.2. The makam scales	9
2.3. Feature extraction	10
2.3.1. Pitch-class profile estimation.....	11
2.3.2. Pre-processing.....	12
2.3.3. Estimator, frequency determination and mapping to pitch class .	13
2.3.4. Interval resolution.....	14
2.3.5. Post-processing.....	15
2.4. Scale estimation	15
Chapter 3.....	20
Methodology.....	20

3.1. Music material	20
3.2. Feature extraction	21
3.3. Template-matching model	23
3.4. Class-modeling model.....	28
3.5. Evaluation measures	29
Chapter 4.....	30
Results	30
4.1. Template-matching model	30
4.2. Class modeling.....	32
4.3. Error analysis	32
4.4. Cross-validation	34
4.5. Discussion	35
Chapter 5.....	37
Conclusion.....	37
5.1. Summary of contributions	37
5.2. Plans for future research	38
References.....	40

List of Figures

1.1	Conceptual framework for music content description.....	4
2.1	Tetrachords and pentachords	8
2.2	The makam scales	10
2.3	General block diagram for pitch-class estimation.....	12
2.4	Pitch Classes defined by Arel theory.....	15
2.5	Geographical regions where makam-music is found.....	16
3.1	Block diagram of the template model used.	24
3.2	Chroma vector extracted from a single song, aligned with respect to its tonic	25
3.3	Template of <i>nihavend</i> makam built with the model	25
3.4	Theoretical representation of <i>Hicaz</i> makam.....	26
3.5	Chroma vectors matched with their makam template for <i>segah</i> , <i>hicaz</i> and <i>nihavend</i> pieces.....	27

List of Tables

1.1.	Distinction between musicology and ethnomusicology	5
1.2.	Differences between Western and non-Western music material... ..	5
2.1.	Evaluation results of makam estimation from Gedik, Bozkurd (2009)	19
3.1.	Distribution of instances among the different makams	21
3.2.	Makam scale intervals of the 9 makam in <i>Arel</i> theory. Intervals given in <i>Holdrian</i> commas	22
4.1.	Evaluation measures for different parts of the sound files from the template-matching approach.....	31
4.2.	Comparison of the proposed models with and without tonic information.. ..	31
4.3.	Confusion matrix from the template-matching model	33
4.4.	Error percentage for all makam class.....	33
4.5.	Confusion matrix of the classification performed with SVM using our music library.....	34
4.6.	Accuracy rates from the cross-validation experiments.....	35

Acknowledgements

Firstly, I want to thank Dr. Xavier Serra and all the Music Technology Group for granting me the opportunity to study close to them and share their knowledge with me.

Then, I would like to gratefully thank my professor and supervisor, Dr. Emilia Gómez for guiding me throughout this thesis and for supporting me in every step of the master course. I also want to thank Dr. Perfecto Herrera for his valuable help and insights in my work.

Furthermore, I thank all the researchers at the MTG, who have helped me and shared their knowledge during this year. Many thanks to my classmates with whom we shared our time and experiences as they have been excellent colleagues and friends.

Finally, I would like to deeply thank my parents, Athina and Theodoro for supporting me in this trip all along and for always urging me to progress and learn.

Chapter 1

Introduction

In this chapter is stated the motivation of the project and the context is been briefly described along with the scientific and musicological background of makam art music. Finally the structure of this thesis is given.

1.1. Motivation

This thesis forms part of the current research carried out in the field of computationally driven ethnomusicology, which is a branch of the multi-disciplinary field of Music Information Retrieval (MIR). As the modern needs demand an easy and user-based access and search of huge databases which are becoming very popular, there has grown the need to have content-based description of musical audio. The need to characterize music in various ways has led to the development of high-level descriptors that give the ability to accurately detect and analyze information that could be difficult and time-consuming to perform by hand. Ethnic or traditional music styles form a special group of MIR research that often demands different approach than Western music styles.

As a consequence the musical scale, that is one of the higher level musical units, is of the key elements that characterize large families of music in various musical cultures. Even in many cases musicians use the scale name to communicate, in a musical sense. So an automatic labeling system for the scale type and, in particular, for makam music is the goal of this thesis. Several approaches, which include both from signal processing and machine learning techniques, have been proposed in the past by researches in order to

perform this task. Thus a secondary goal of this project is to build two alternative models and evaluate their performance.

1.2. Goals and expectations

As stated above the main goal of this work is to automatically detect and label the makam scale from a musical excerpt in audio format. The procedure followed to achieve the above is the following one:

1. Choose the most popular and commonly practiced makam scales and build a music library keeping an equal distribution among them. The choice of the makam depends a lot on similarity/dissimilarity issues. Makams will be explained more in depth in chapter 2.
2. Extract features that carry information that can help to identify the scale. In this project the feature space is formed only by chroma features and the tonic of each piece.
3. Build scale templates for each makam family first by applying music theory concepts and second by averaging extracted features over excerpts sharing the same makam.
4. Compute various distance measures with the obtained templates for new music excerpts in order to label them correctly.
5. Train a classifier with machine learning techniques that accomplish the same task.
6. Evaluate results from both steps 4 and 5.

All of the above will be made in order to achieve satisfying results for the makam labeling task as stated above. The task will have its constraints in terms of the music material that will be used and the theory used to define each scale nevertheless the ultimate goal is to build a model that can be extended, with the proper tuning and scale information, to estimate the scale from other music cultures that use micro-tonal intervals and a non-equal-temperament.

1.3. Music Information Retrieval (MIR)

Music Information Retrieval, as described in [Herrera 09], is a relatively young field that its roots can be traced back in the 60's. It was formed as a more concrete research discipline since 2000 when the first International Conference of MIR took place. The traditional ways of discovering music are rapidly changing with development of networks and social networking methods of accessing and exchanging audio and music, something that is already altering the music industry as well. According to [Casey 08], this need for music recommendation has been the main cause in the fast development of audio-mining techniques. It can involve disciplines from musicology, signal-processing and statistical modeling [Leman 02].

1.3.1. Audio content description

Content description is one of the main tasks MIR systems are meant for. The content of audio signal has different levels of abstraction. In [Leman 02] they are described in a top-down taxonomy shown in the framework of figure 1.1. The lowest level of abstraction is consisted of content close to the sensorial or acoustical properties of the signal, such as frequency, intensity, loudness etc. In the mid-level, objects with musical meaning start to appear as the time-space concept is introduced. In a time limit of approximately 3 seconds concepts like pitch, meter, timbre, etc can only be conceived. Finally in the higher level reaches concepts of musical human cognition related to musical form, genre and even psychological aspects such as emotions and understanding. Depending on the task MIR systems aim in the mid and high levels by trying to label audio signals according to them. Low-level features can be easily extracted and manipulated by computers with the latest signal processing techniques.

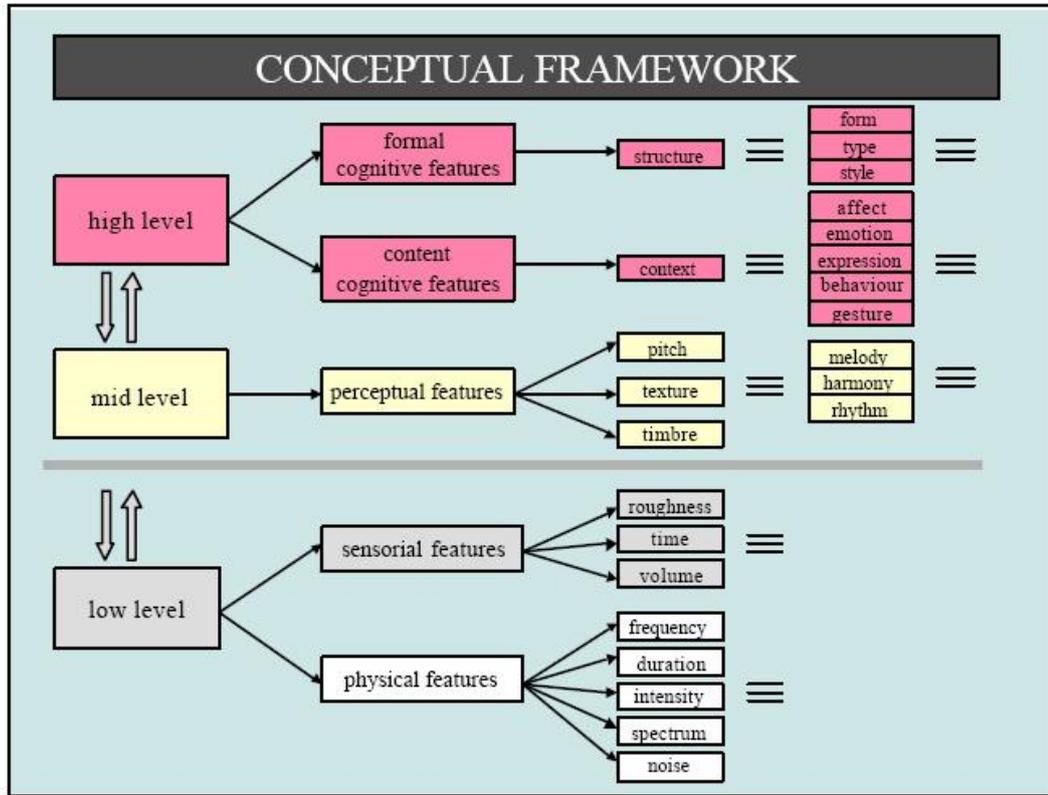


Figure 1.1. Conceptual framework for music content description, Leman (2002).

1.3.2. Computer-aided ethnomusicology

In [Tzanetakis 06] provides a survey named ‘*Computational ethnomusicology*’ that defines this multi-disciplinary field. Rapid evolution of MIR technology has given the opportunity to many ethnomusicologists to conduct their research using state-of-the-art signal processing tools. The prefix ethno- only puts the focus into traditional music styles and to music which does not rely on Western music theory. In [Tzanetakis 06] are also explained the differences between the term musicology and ethnomusicology claiming that for the latter research does not imply any particular methodology. In table 1.1, by Tzanetakis, this distinction is shown. Here the goals can be focused not only in music access and music recommendation systems, although this is the goal the current work presented here. Computational ethnomusicology can

differentiate when the aim is to capture and understand the diversity between musical cultures, or to engage in musical concepts that have not been documented by previous musicological research or are never annotated in scores, as usually there is lack of theoretical representation of ethnic music. In chapter 2.3 we will present the state-of-the-art on the problem of scale estimation in non-Western music.

Discipline	Musicology	Ethnomusicology
Music studied	Notated music of Western cultural elites	Everything else
How the music is transmitted	Notation	Orally transmitted
Methodology	Analysis of scores and other documents	Fieldwork, ethnography

Table 1.1. Distinction between musicology and ethnomusicology.

In table 1.2 we dictate differences between Western and non-Western music that are related to the use of tonal pitches. These properties are not incorporated in all non-Western music or not all at the same time. In order to conduct Computational Ethnomusicology using MIR tools, these tools should be tuned according to the needs of the each music.

Feature	Western	Non-Western
Octave deviation	Equal temperament	Non-tempered
Tuning	Standard diapason	No standard tuning
Practice vs theory	Can be studied both ways as theory defines the actual practice.	Lack of ground truth. Orally transmitted cultural music. Theory is descriptive.

Table 1.2. Differences between Western and non-Western music.

1.4. Thesis structure

Concerning the structure of this thesis project this chapter briefly defined the problem along with the goals and expectations of the method proposed to solve it. In chapter 2 we present the musicological background of makam music along with the scientific background of the signal processing and machine learning techniques that we will use throughout the thesis. In the same chapter we provide the current state of the art on the same research problem of scale – mode identification in Western and non-Western music. In chapter 3 we introduce the proposed model to solve the task and chapter 4 contains the evaluation and the results. In the last chapter we conclude and summarize our main findings and provide ideas for future research and a discussion on the topic.

Chapter 2

Scientific background

This chapter presents the musicological background of the music that we call '*makam music*'. It is a music culture that can be found in the Arab world, Ottoman classical music, and Persian music. It should be highly considered than in each place makam is practiced differently and there is not a ground truth or a common theoretical model, as it is an orally transmitted music. Here we focus on the scales that are studied in the context of this thesis.

2.1. Makam music

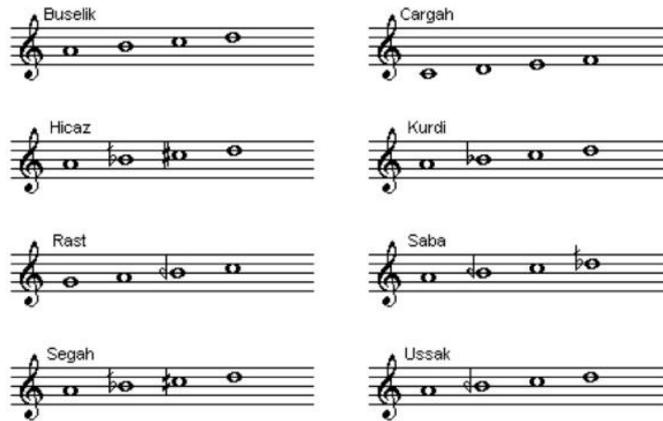
Makam (from Arab: *maqam*) is a unique musical concept found in musical cultures of the Arab world, East-Mediterranean, middle-East and it shares similarities with the '*echos*' of Byzantine music. The equivalent word in Western music would be mode or tonal scale, although that is just a simplification. Nowadays makams have been divided into three large families, the *Arab maqam* (pl. *maqamat*), the *Turkish makam* (pl. *makamlar*) and the *Persian maqam*. In this project we focus on the *Arel-Ezgi-Yekta* (Arel theory) and the Turkish Makam Music. Although makams differ from place to place, we claim that the work done here can easily be extended and applied into all of these musical cultures.

Makam is a compositional device that primarily defines the notes and their intervallic distances in a musical piece. It also indicates the note sequence (*seyir*), not in an absolute way but rather than in an ascending/descending behavior it has when developing in time. The makam also indicates the tonic as this is another way to categorize this music. Each makam scale has its own tonic centers and modulations on specific grades of the scale. Finally it is

a common belief among performers, listeners and theorists that it's connected to mood, and in a Western sense it can be considered as genre.

In Arab and Turkish music theory, the smallest named modal entity is the tetrachord and the pentachord. These building blocks are shown in figure 2.1. Their number is absolute in contrary to the number of makams which is claimed to be over 300; nevertheless the actual practice of this music is limited to approximately 30 makam scales, according to Harold S. Powers, et al. [2010]. A makam scale can be built by combining two or more of these tetra/pentachords and it is named after the lowest or most important one.

Tetrachords



Pentachords

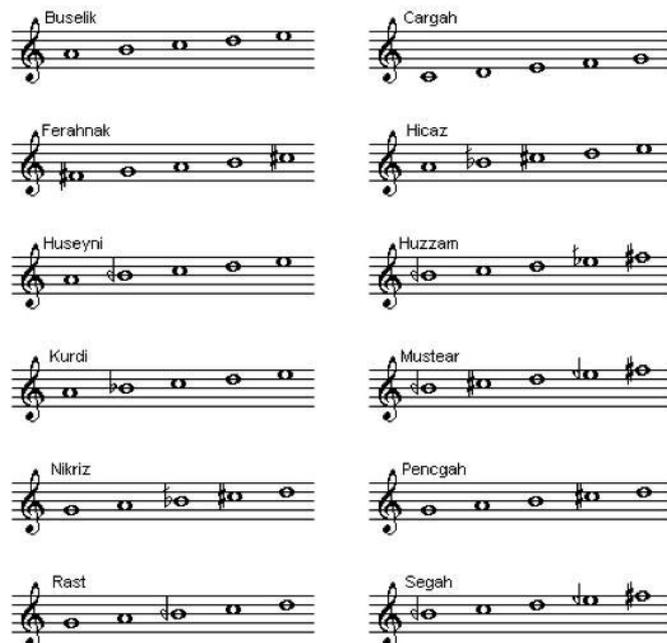


Figure 2.1. Tetrachords and Pentachords.

2.2. The makam scales

Makam is an Arabic word meaning place or position. As explained above, makam music is an orally transmitted culture and thus various theoretical models have been used in order to annotate [Walter Zev Feldman, 2010]. First efforts towards notation were in the 9th century, and many ways were proposed since then in order to notate rhythm, notes and form. Efforts to modernize this music genre in the 20th century resulted in the more self-consciously rigorous application of Western staff notation to the makam system. The system of notation that remains in use, in Turkish conservatories, is known as the Ezgi-Arel system, or simply Arel system. In [Yarman 08] there is a comparative evaluation among the historical and modern theoretical systems, claiming that the 53-equal commas, derived from the 24-Pythagorean deviation, has become widespread due to westernisation of the music theory. There have been proposed other systems afterwards without being as popular. For this study we will consider the 24-Pythagorean system, as being the most common in both practice and education and the one able to capture the plethora of intervallic distances in Turkish and Arab quarter notes.

As mentioned in section 2.1 the number of makam scales is always under discussion. The actual practice of this music is believed to be restricted to 30 – 50 scales, from experts of the style with the selection of these nine scales being approximately the 50% of the today's practice [Bozkurt 06].

The makam that we will classify and label are the ones used in other studies of the same problem by A. Gedik and Bozkurt. We then keep the same families in order to compare and extent previous researches. In figure 2.2 we present the Arel notation along with the tetra/pentachords that they are built each scale shows different evolution above and below the main octave and modulations to other tetrachords may happen. The accidentals are explained in figure 2.4.

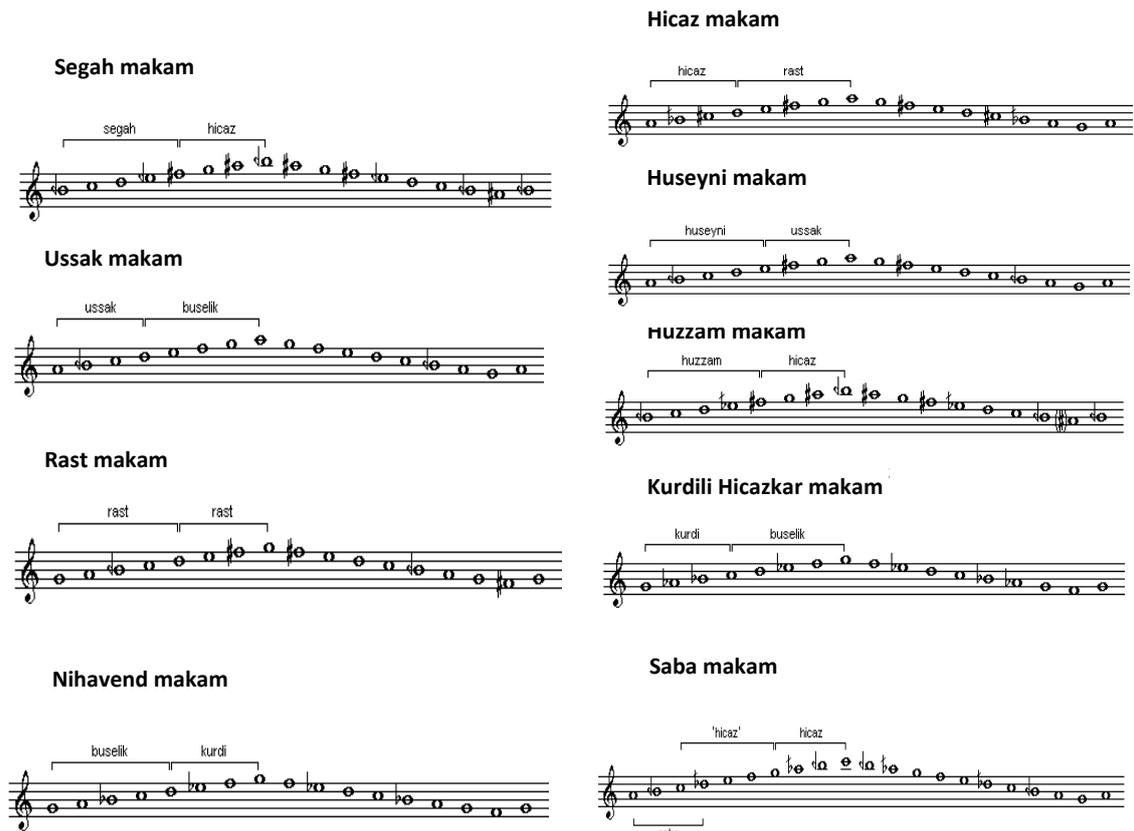


Figure 2.2. The makam scales.

2.3. Feature extraction

In section 1.3.1 we introduced the concept of content description. The key, or tonic scale/mode in the case of this thesis, is one of the descriptors that have been estimated from audio signals. The research on tonal features was motivated and applied to the automatic extraction of score information from audio signals, but we see how pitch class profiles can be useful for high semantic description, without the primer need of music transcription, as tonal description and transcription are highly related but not bonded together. The main feature for the description of tonal content of a music piece is the distribution of played notes. Gómez, [2006] provides a review of the methods proposed for fundamental frequency estimation and Pitch-Class Profile or chroma estimation. Fundamental frequency is more appropriate for monophonic sound excerpts, and the chroma features can be used in

polyphonic audio too. Both features assume the presence of harmonicity and are to be dealt with such signals. The steps in both fundamental frequency and chroma estimation follow the general schema shown in figure 2.3. The method proposed by Gómez [2006] is summarized here and is the one used in this project for the Harmonic Pitch Class Profile (from now HPCP) extraction, which was part of the feature space.

2.3.1. Pitch-class profile estimation

As mentioned, the need for multi-pitch estimation led to the development of tonal description. Herrera [2006] makes the distinction between transcription and description for different aspects of sounds, implying that various music semantics like chord, key or scale information can be estimated very accurately without the need of the music score. A reliable tonal descriptor, according to Gómez, [2006] should:

- i) Represent pitch class from both mono and poly-phonic audio.
- ii) Consider the presence of harmonic frequencies.
- iii) Be robust to noise (ambience noise, percussive, non-harmonic sounds).
- iv) Be independent of timbre characteristics and played instrument.
- v) Be independent of loudness and dynamics.
- vi) Independent of tuning, so that the reference frequency can differ.

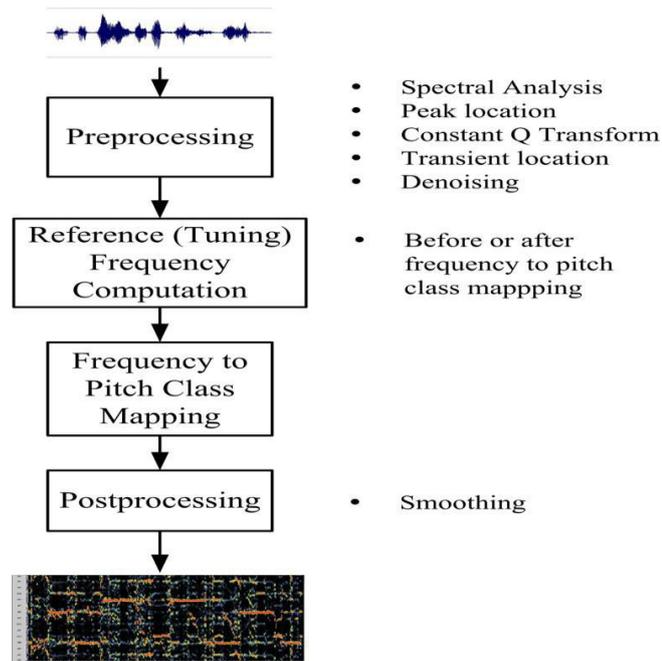


Figure 2.3. General block diagram for pitch-class computation.

2.3.2. Pre-processing

In the pre-processing step the task is to prepare the signal in order to facilitate an accurate and robust estimation and discard irrelevant information. It includes some steps that will enhance the features from which the computation will happen. Such steps are the spectral analysis of the signal, peak location of harmonic maxima, constant-Q transform [Purwins et al. 00] and various denoising techniques in order to reduce errors during extraction.

An important step in pre-process is the estimation of the reference tuning frequency. The middle A (440 Hz) is considered as the standard reference frequency, in both tuning and pitch class definition. Nevertheless this cannot be expected to be so in all cases, and even more when studying non-Western music, where the reference frequency can vary a lot. Pre-processing in HPCP consists the following steps:

- i. Detection of the local maxima of the constant Q-transform having its energy above threshold.

- ii. Grouping of the detected peaks into groups according to their deviation (modulus 12) from the 440 Hz reference frequency.
- iii. Choose the maximum group as the tuning frequency for this analysis frame, given that its value is prominent enough.
- iv. Build a histogram of the tuning frequencies of all frames and choose the maximum point as the reference frequency.

The evaluation of such methods is unknown and has been done using expert musical knowledge.

2.3.3. Estimator, frequency determination and mapping to pitch class

The main idea behind Pitch Class Profiles is measuring the intensity of each of the twelve semitones of the diatonic scale, and it is obtained by mapping each frequency bin of the spectrum to a given pitch class. The method proposed by Gómez introduces a weight into the feature computation and the presence of harmonics, and thus it is called Harmonic Pitch Class Profile (HPCP). The HPCP computation is done in the frequency region of 100 – 5000 Hz not considering high frequencies. The vector is computed by the following formula:

$$HPCP(n) = \sum_{i=1}^{nPeaks} w(n, f_i) \cdot a_i^2 \quad (1)$$

$$n = 1 \dots size$$

where a_i^2 and f_i is the linear magnitude and frequency values of the peak number of i , $nPeaks$ is the number of spectral peaks that are considered, n is the HPCP bin, $size$ is the size of the HPCP vector and $w(n, f_i)$ is the weight of the frequency f_i when considering the HPCP bin n .

2.3.4. Interval resolution

When dealing with chord, key or scale description, the resolution used in pitch class representation is an important parameter. In Western music the minimum resolution used is of one semitone (12 bins per octave, 100 cents). Increasing the resolution to 1/3 semitone (36 bins) can make the system more robust against tuning and other frequency oscillations, according to [Purwins 00], [Gómez 06, 04].

In the case of makam music the 12 pitch-class does not provide an appropriate representation of the intervals that are used, [Bozkurt 08]. [Gedik, Bozkurt 09] do an evaluation of the makam scale theory for the purposes of MIR research. As defined in Arel theory, an octave is divided in 24 pitch classes. This is only a subset of the 31 possible pitch classes that can occur according to the Arel theory, due to use of all possible intervallic alterations. All are shown in comparison to the Western clavier. The smallest intervallic distance observed is the Holdrian comma (22.64 cents), obtained from the division of the octave in to 53 logarithmically equal partitions. So at least a resolution of 53 bins is needed to represent properly makam scales. Gedik, in the same paper, claims that a 1/3 *Holdrian* (159 bins) comma can provide robustness and bigger accuracy the same way as stated above.

there is a thin line between what can be considered as traditional/ethnic music and what as non-Western music. There are many cases, that current MIR techniques are sufficient for ethno-musicological research.

Various efforts in key and mode estimation have been carried out, considering Western music. The problem can be considered as solved at least for classical music. The MIREX contest is held annually where algorithms for various MIR tasks are being evaluated. For the task of key estimation, and already from the year 2005, participants achieved accuracy from 79% to 90% for the same music library. The most successful method proposed [Ismirli, 2005] makes use of pitch distribution profiles which according to Krumhansl represent tonal hierarchies.



Figure 2.5. Regions where makam music is found.

These profiles are cognitive patterns as they have been extracted by listening experiments. As feature space, chroma profiles are extracted from the audio signal, and to estimate the key correlation coefficients are calculated between the summary chroma vector and the 24 pre-calculated profiles which correspond to the 12 major and minor modes. To increase robustness, Izmirli

proposes to consider the highest and second highest correlations, which is the key that has appeared more as an individual estimate in any of the windows of chroma estimates.

The non-temperedness of the musical mode as a feature has been studied by Gómez and Herrera [2008] where they define a set of tonal features and descriptors which can accurately inform about the 'equal-temperedness' or not of a musical excerpt. This descriptor is expected to be zero for Western-classical music and close to one in non-equal-tempered musical pieces. The latter was done for the task of labeling songs as Western or non-Western.

There are a few efforts towards scale/mode estimation as most ethnomusicological research is devoted in building automatic transcription systems for score notation and pitch analysis of non-Western musical instruments. Automatic analysis and classification of scales in music that is not organized according to the Western tonal system is much less developed [Cornelis, 2009]. We provide a brief description of some of the most recent efforts in classifying the scale.

Cornelis, does an exploration of the scales used in African music. He uses a database of 900 recordings consisted of solo performances and in some cases accompanied with singing voice. He performs pitch detection, and segments the notes by doing peak detection of the time-frequency representation. For the above step, he uses 1200 bin analysis where each bin corresponds to a cent value in the octave. To analyze its data, the chroma-vectors are cross-correlated in order to facilitate the retrieval of songs that are similar in terms of intervals. For further analysis, the scales obtained from the peak analysis are transformed into an array of all possible intervals that can be built with this scale. These interval histograms are capable of representing the equal temperedness since intervals not corresponding to the Western tuning are observed. An important observation in the above work is that African music gradually adopted "Western like" intervals (such as the perfect fifth). That was shown from an analysis that was considering the recording date (in three time periods) which confirmed the above observation.

Another different music culture, which is not based in the Western music theory, is the classical Indian music. Although there are a lot sub-categories and styles of that music, generally the concept of scale is characterized as *Raga* (*pl. Raag*), which as a musical concept shares a lot of similarities with the makam. As for the state of the art we refer, to the work done by [Parag Chordia, Alex Rae, 2007] in Raag recognition. They achieved high accuracy for this task by using training algorithms and as feature space Pitch class Distributions (PCD) and Pitch class Dyad Distributions (PCDD), which were extracted from a music library with a big variety of Raags. The PCD and PCDD are extracted by doing pitch detection and then by applying an onset detection algorithm to label the notes into Western pitch-classes.

On the problem of makam recognition there has been a series of papers published by Ali C. Gedik and B. Bozkurt dealing with the problem. This research included also other aspects such as tonic detection and note transcription. Various experiments are conducted using as material a library of monophonic recordings of 9 different makams, performed by undisputable experts of the style, as they propose that the use of theory should be not be underestimated, but cannot be expected to be followed in an explicit way in all performances.

The technique used in this work template-matching with the use of pitch-histograms based on fundamental frequency, extracted from the Yin algorithm [de Cheveigne, Kawahara, 2002]. A robust tonic detection algorithm is proposed, which takes advantage of the Arel theory to find the tonic from an annotated song. Each makam is represented in a template as a sum of Gaussians with the center of each Gaussian being on the intervals dictated in the *Arel* theory. By shifting, to each bin, and cross-correlating each recording's histogram with the corresponding theoretical template, the tonic is detected in the maximum correlation. The next step is to shift the histogram with the tonic bin first. To label a song with its makam name they use different approaches such as computing the minimum of various distances between the recording's histogram and the theoretical representation, or with templates computed by averaging the histograms extracted from their database. Their collection of music was parted from 118 recordings of 9 different makams:

rast (15), *segah* (15), *huzzam* (13), *saba* (11), *hicaz* (11), *ussak* (11), *kurdili hicazkar* (17), *nihavend* (12), [Gedik, Bozkurt, 09]. The same audio library was used for creating the template and testing the model, by using leave-one-out cross-validation in the construction of the templates.

To summarize the results from this work we provide in table 2.1 the evaluation results obtained from their model.

Makam Type	Recall	Precision	F-measure
Hicaz	70	88	78
Rast	73	88	79
Segah	85	85	85
Kurdili Hicazkar	63	48	55
Huzzam	71	63	67
Nihavend	78	56	65
Huseyni	50	63	56
Ussak	63	60	62
Saba	76	94	84
Mean	68	68	68

Table 2.1. Evaluation results of makam estimation from Gedik, Bozkurd, [2010].

Chapter 3

Methodology

This chapter presents the methodology and steps followed for building a makam scale estimation system.

3.1. Music material

We have collected a music library of 289 pieces from the selected 9 makkam types, which are *hicaz*, *huseyni*, *rast*, *nihavent*, *kurdili hicazkar*, *saba*, *segah*, *huzzam*, *ussak*. Songs from various countries were considered (Turkey, Greece, Egypt and Iran). Nevertheless, the majority comes from Turkey, where it is a common practice to annotate the makam in the title of the piece. That made easy to establish the ground truth information for each piece. All the songs used in our study were extracted from commercial CDs. We use the AEU (from now *AreI*) theory when needed, as it is claimed to be capable of representing the plethora of intervals in Turkish makam music and also being suitable to capture the intervals in Arab makams [Yarman, 2007]. We tried to keep an equal distribution among the makams and to include pieces from diverse morphological manifestations of that music, such as *taksim* (improvisational performance), religious chants, *sirto*, *longa* or *semai*. We address to Tanrıkorur [1996] for further information on the musical forms of Ottoman music. The plurality of styles and origins had the intention to make a non-biased model to a specific region or morphological style. The beginning, the end and a part from the middle of the piece were extracted from all the analyzed pieces. In all cases the excerpts were of 60 seconds long. This length has been used in similar research on Western music [MIREX 2005].

A second library was used to test and cross-validate the proposed models for makam estimation, which are presented in this chapter, along with its dependency on the trained database. Gedik and Bozkurt have carried out similar research which is summarized in section 2.3. In their work they use a library of 118 monophonic pieces of *taksim* performances from ‘undisputed masters’ of Turkish art music. They made available online the *Yin* output files extracted from each song for research use. From this library we used a subset of 112 songs that were from the same makam types studied here. The distribution among different makams can be seen in figure 3.1. In section 3.2, we show how vectors corresponding to chroma representation were built from *Yin* output files.

Makam	Count	Percentage (%)
hicaz	49	16,0
huseyni	28	9,6
rast	32	11,0
nihavend	30	10,3
kurdili hicazkar	22	7,6
saba	39	13,5
segah	23	7,9
huzzam	34	11,7
ussak	32	11,0

Table 3.1. Distribution of instances among the different makams.

3.2. Feature extraction

As described in chapter 2, HPCP vectors (of 159 bins per octave) form our feature vector. This resolution was chosen as a single octave is divided into 53 equally spaced intervals according the AEU theory and each one forms a *Holdrian* comma. So a 1/3 comma resolution is considered to provide tuning robustness and good resolution [Gedik, 08, Gomez, 06]. Each one of these bins represents the relative intensity within a given 1/3 comma octave region.

Chroma features were computed on a frame basis and a global vector was obtained by averaging the frame values within the consisted excerpt. We used a 46.5 msec frame window and we considered the frequency range between 40 – 5000 Hz.. Figure 3.3 provides an example of chroma feature extracted from a musical excerpt of *rast* makam. The dashed line indicates the intervals with respect to the tonic as dictated from the *Arel theory*, which are available also in table 3.1. In figure 3.4 we provide, for the makam *nihavend*, the template build from the model described in 3.3 by averaging chroma features over all *nihavend* musical excerpts of our music collection. The peaks agree with the theoretical representation and we also observe which scale degrees are more frequent. We claim that these histograms could measure the same type of information as some templates obtained for Western classical music by different strategies (listening ratings [Krumhansl, 1999] or statistical analyses [Wei Chai PhD thesis, 2005]).

	1	2	3	4	5	6	7	8
hic	5	17	22	31	35	39	44	53
hus	8	13	22	31	39	44	-	53
ras	9	17	22	31	40	48	-	53
nih	9	13	22	31	35	44	-	53
kurd	4	13	22	31	35	44	-	53
sab	8	13	18	31	35	44	49	-
seg	5	14	22	31	36	45	49	53
huz	5	14	19	31	36	49	-	53
uss	8	13	22	31	35	44	-	53

Table 3.2. Makam scale intervals of the 9 makam in *Arel* theory. Intervals given in *Holdrian* commas.

For the case of the library from [Gedik, 2009] used to further evaluate the different approaches, we needed to build analogous feature vectors measuring energy within a single octave of 159 bins. The provided *Yin* files contain fundamental frequency values in cents. Thus they were first converted

into Hertz and then into octave reference numbers, using as tuning frequency the standard A '440Hz', the same used for the chroma profiles. Then, the octave information was mapped into a single octave represented in a 0 to 1 scale where 0 and 1 are the same pitch class. Finally we average the obtained vector over the whole song so that the sum of occurrences is equal to 1.

3.3. Template-matching model

This model is inspired from [Gedik, 2008] and it is summarized in figure 3.2. It is based on pattern recognition and its goal is to identify patterns, in this case chroma vectors, which are similar in terms of distance. Makam templates are defined and describe the intervals of each makam and the weight of each degree in the scale. A makam template of a *nihavend* makam is shown in figure 3.3. These templates are built by extracting the chroma information from each song annotated under the same makam name. Then the vector is shifted towards the tonic bin which is detected using the model of [Gedik, 2008], which is briefly explained in section 2.3. Finally all the vectors from one makam category are averaged in order to obtain each makam template.

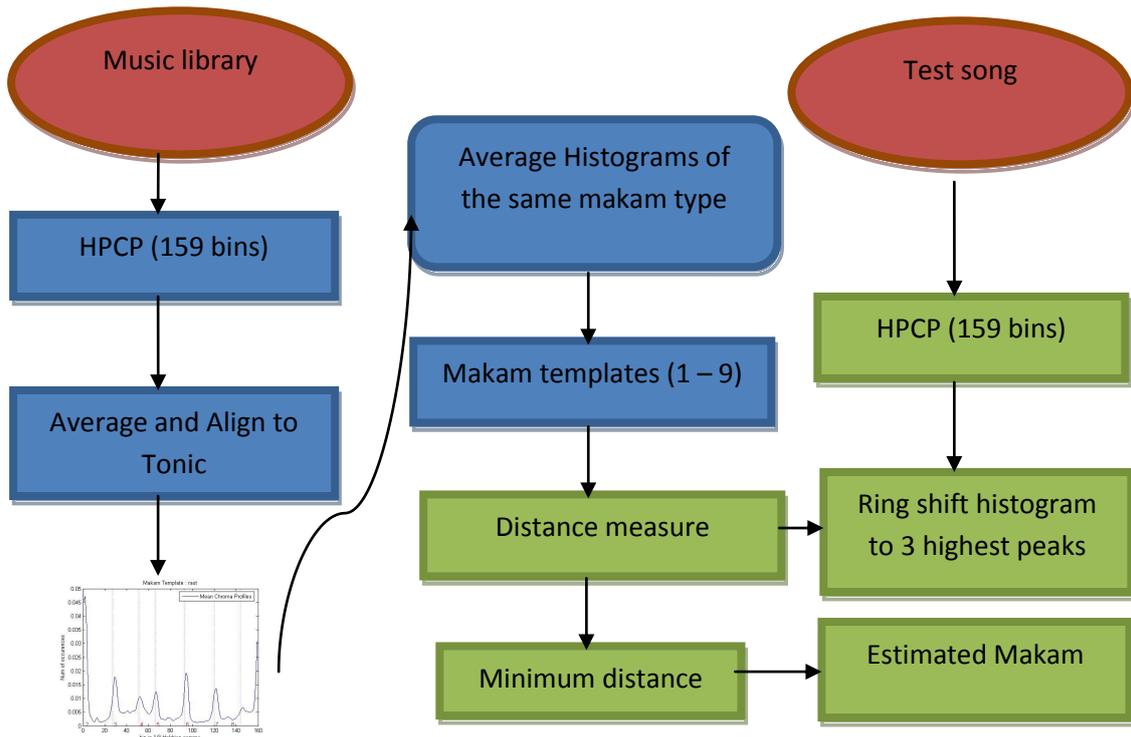


Figure 3.1. Block diagram of the template model used.

The tonic detection procedure is been done by computing the cross-correlation between the chroma vector extracted from the song and its' theoretical representation. To theoretically represent the makam we used the intervals dictated from the *Are!* theory, in table 3.1 and constructed vectors of 159 bins with a Gaussian sum where the peaks of the Gaussians were centered on each interval with comma precision. Such a template is shown in figure 3.5. Thus, we note that to automatically detect the tonic we use ground truth knowledge, meaning that the makam scale is known. In chapter 5 we discuss further the limitations we faced and possible future directions concerning the tonic detection method. To detect the tonic of a recording its chroma vector was slid over the theoretical template with 1/3 comma steps. In each step the correlation was computed and finally the tonic was assigned to the first bin of the shift amount where the maximum correlation was found.

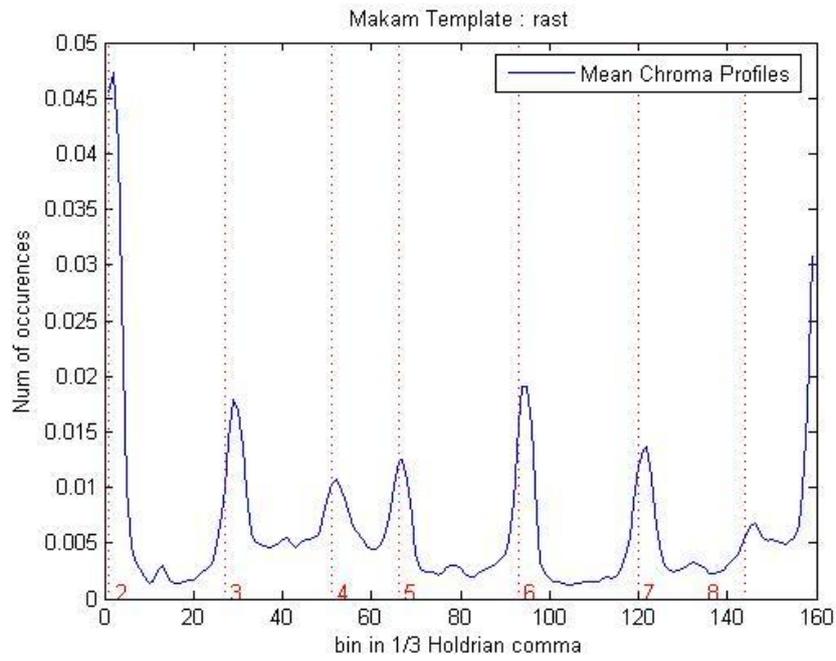


Figure 3.2. Chroma vector extracted from a single song, aligned towards its tonic.

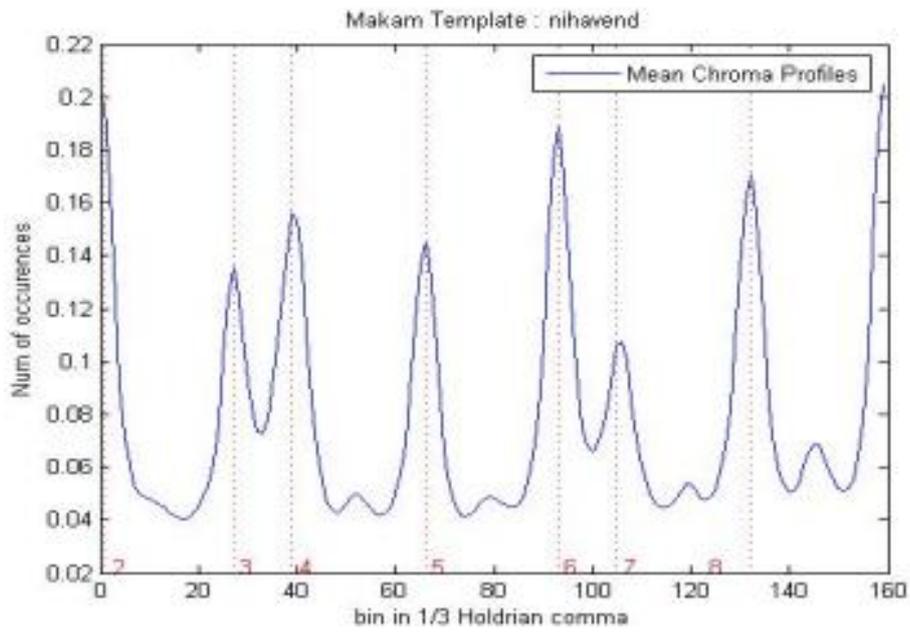


Figure 3.3. Template of *nihavend* makam built with the model.

distance was found. Chapter 4 presents the evaluation results for all the described experiments and a discussion about their limitations.

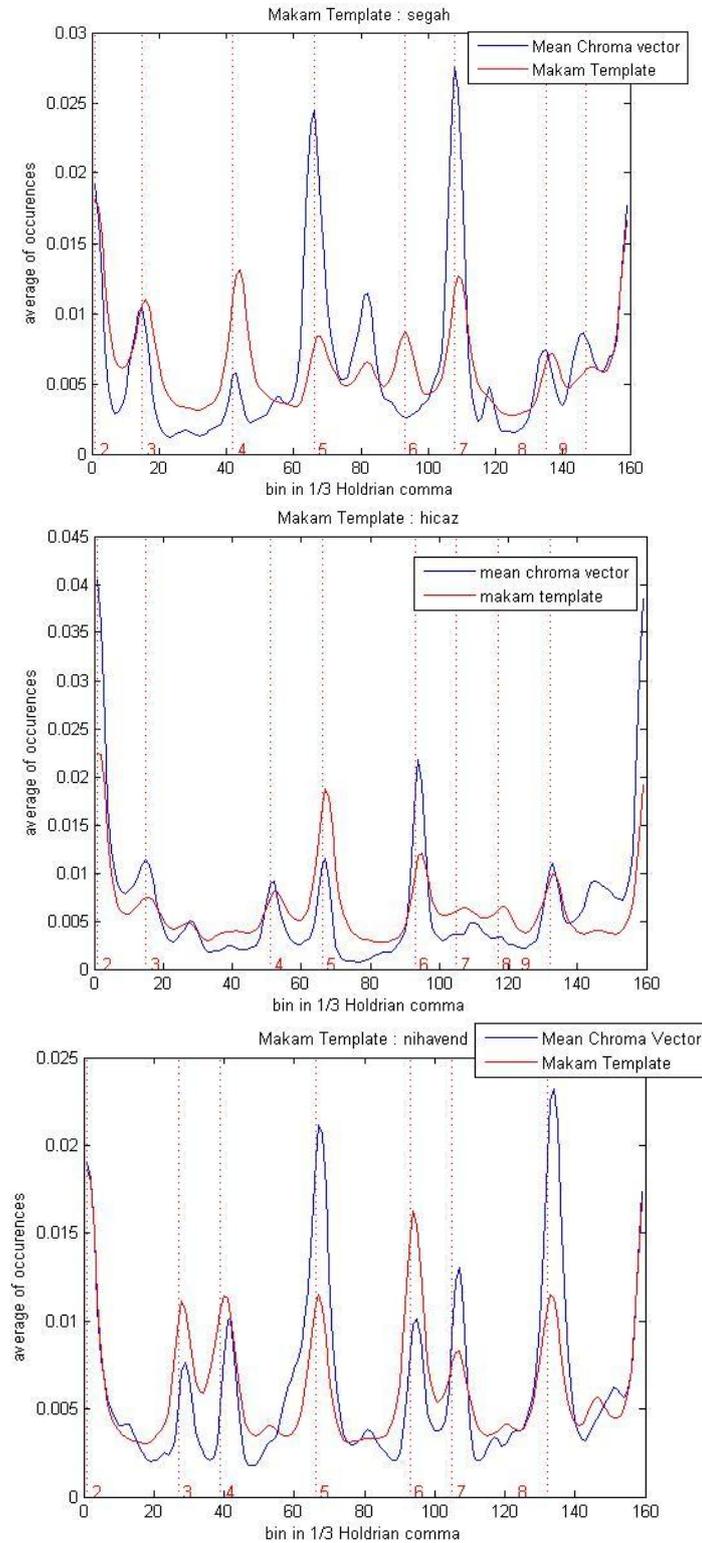


Figure 3.5. Chroma vectors matched with their makam template for *segah*, *hicaz* and *nihavend* pieces.

3.4. Class-modeling model

To investigate the success of machine learning techniques to solve the same problem from the same feature set, we set up several experiments using the machine learning tool WEKA¹. Weka is an open-source program with a collection of machine learning algorithms for data mining tasks. We performed two experiments. In the first one, chroma vectors were used as feature set. In the second one we provided these vectors shifted with respect to the detected tonic using the algorithm described in section 3.3. Various classification techniques were evaluated but we report results obtained with Support Vector Machines (SVM) as they yielded the best overall performance.

At first we trained the classifier using the same feature set as in template-matching approach in order to have a direct comparison between the two models, thus we provided only the chroma values from the 159 bins vector. To boost the performance we applied various numerical transforms to the vector and concluded that by computing the log of the vector was the best one, as it added a 2.6% to the overall performance. As the feature space was very long, more than the half the size of instances, we applied an attribute selection procedure, using the *BestFit* function, which shortened the number of attributes to 46. This also provided the final boost in the classification accuracy. A 10-fold cross-validation was used in all cases to ensure enough generalization power to model and to not get stuck into an overfitted method.

To cross-validate the model with the external library from Gedik [2009] we trained the model using the chroma features extracted from our database and the vectors created from the yin output files as a test set. Results from all the experiments, along with the obtained confusion matrixes are summarized and compared to the template-matching model in chapter 4.

¹ <http://www.cs.waikato.ac.nz/ml/weka/>

3.5. Evaluation measures

Standard evaluation measures were used to study the success of the two models proposed in this work. Precision, Recall and F-measure were calculated for each of the 9 classes corresponding to each of the investigated makams. The formulas that give the above measures are the following:

$$Precision = \frac{true\ positive}{true\ positive + false\ positive} \quad (2)$$

$$Recall = \frac{true\ positive}{true\ positive + false\ negative} \quad (3)$$

$$F - measure = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (4)$$

- *Recall*: measures how many of “positive” class are correctly classified.
- *Precision*: measures how many classes classified as “positive” are indeed “positive”.
- *F-measures*: Is a weighted average between *precision* and *recall*.

Where, for a makam type, e.g. *hicz*, “true positive” are the instances correctly classified as *hicz*, “false positive” are the instances that belong to a different makam but were classified as *hicz*, and “false negative” are instances that are classified as a different makam although they were *hicz*.

Chapter 4

Results

Here we summarize the results obtained with the two proposed models. As mentioned in chapter 3, standard evaluation measures were computed to study the success of both approaches. The class-modeling model was cross-validated with the use of an external library. The evaluation database consists on the two libraries described in section 3.1.

4.1. Template-matching model

One of the first tasks was to shorten the music library in order to have a uniform set of songs, in terms of duration. Thus we decided to select 60 seconds from each music excerpt. Our assumption was that the presence of the tones dictated from the theory would be mostly present at the beginning or at the end of the pieces, as it is highly expected that a long performance will ‘pass’ from various scales and will modulate to different tonics as well. To validate that, we performed a test with the template-matching model using 60 seconds excerpts from the beginning, end and middle (second minute of the piece) of the files. We took care that in all three cases the makam templates were built with the corresponding parts of begin, middle or the end of files. The results in terms of overall F-measure are summarized in table 4.1, which shows that the best performing excerpt is the end. This musically means that the presence of a makam scale is stronger in the ending of a performance. From now on we report the results using just the end of files from the music collection.

	Precision	Recall	F-measure
Begin	0.50	0.40	0.37
Middle	0.39	0.39	0.36
End	0.50	0.45	0.45

Table 4.1. Evaluation measures for different parts of the sound files for the template-matching approach.

In the confusion matrix we observe in table 4.3 that some makam types are most confused between each other, such as *ussak* with *huseyni*, and *segah with huzzam*, and this is due to the similarity of these scales in terms of intervals. For the template-matching model we can directly compare with previous research done by [Gedik, 2009], provided in section 2.3. The obtained results show close accuracy rates to that previous research, F-measure 0.68 and 0.69 in this work, and the confusion matrices show the same type of errors between certain classes, finding the musical significance to these errors. As a conclusion, using chroma vectors instead of fundamental frequency histograms did not improve neither worsen the performance of this model. Finally we can assume that these evaluation rates reflect the limitation of the template-matching approach and it could be improved by introducing other features that could help the system better discriminate the makams.

	Template-Matching			Machine-Learning		
	P	R	F	P	R	F
HPCP	0.50	0.45	0.45	0.29	0.31	0.29
Tonic-shifted	0.69	0.69	0.69	0.72	0.73	0.73

Table 4.2. Comparison of the proposed models with and without tonic information.

4.2. Class modeling

Table 4.2 shows the overall accuracy obtained by training a classifier. There were two experiments: first using the raw chroma vector as feature space without applying any kind of data transformation. Its weakness to perform in a satisfying way is obvious (F-measure: 0.29). In the second experiment we achieved better performance by providing the chroma vectors shifted towards its tonic bin and by applying the log transform to the chroma vectors (F-measure: 0.73). The confusion matrix is also provided in table 4.4. We observe the same type of confusion than for the template-matching model between two pairs of makams, *ussak - huseyni*, and *segah - huzzam*. But there are classes such as *nihavend* and *kurdili hicazkar* that were well classified here, in comparison to the template-matching model.

4.3. Error analysis

The erroneous instances were examined and the following reasons were observed for the wrong makam estimations:

- a) Low quality of recording and thus very noisy chroma information with spurious peaks.
- b) Wrongly detected tonic, confusion with the dominant or other scale degrees.
- c) Confusion between scales sharing similar intervals, as presented in table 3.1.

makam	Bad quality	Wrong tonic detection	Confusion makam
hicaz	0	71.42	28.57
huseyni	0	45.45	54.54
rast	15.38	76.92	7.69
nihavend	10.52	36.84	52.63

kurdili hicazkar	5.26	42.10	52.63
saba	26.08	69.56	4.34
segah	0	40.0	60.0
huzzam	19.04	42.85	38.09
ussak	23.07	42.30	36.61
overall	14.10	50.64	34.61

Table 4.3. Error percentage for makam classes

Table 4.3 demonstrates the reasons in percentage for erroneous estimation for the template-matching model. It shows that half of the overall errors were due to bad tonic detection and that 34% were wrongly classified because of confusion with a similar scale, while a small percentage of the instances had noisy chroma information which led to wrong estimation. Moreover we observe that classes like *segah*, *huseyni*, *kurdili hicazkar* and *nihavend* had the biggest confusion between other scales, something that was also observed by Gedik, [2008].

makam	hic	hus	ras	nih	kurd	sab	seg	huz	uss
hic	43	1	2	1	1	0	0	0	2
hus	7	7	1	1	4	1	1	0	7
ras	8	2	23	2	0	0	1	0	0
nih	8	4	2	11	3	0	0	0	2
kurd	2	3	2	6	5	0	0	0	4
sab	9	3	2	2	2	17	4	0	2
seg	2	1	1	0	0	1	21	1	0
huz	9	2	1	0	0	1	6	15	2
uss	3	9	1	5	5	2	1	0	6

Table 4.4. Confusion matrix from the template-matching model.

makam	hic	hus	ras	nih	kurd	sab	seg	huz	uss
hic	45	1	0	0	3	0	0	0	1
hus	1	13	0	2	2	2	0	0	8
ras	0	0	38	1	0	0	0	0	0
nih	0	1	0	31	0	1	0	0	3
kurd	1	2	0	0	15	2	0	0	2
sab	0	2	0	4	1	27	0	0	4
seg	1	0	0	0	0	0	16	8	0
huz	1	0	0	0	0	0	5	28	0
uss	2	8	0	3	4	2	0	0	11

Table 4.5. Confusion matrix for machine learning approach.

4.4. Cross-validation

Cross-validating a model to ensure its generalization power is an important task in this type of research and makes clearer the effectiveness and robustness not only of the model itself but also the quality of the data and how representative it is of the music property that tries to describe. When only a single dataset is used, to what is called 'Self-Classification', it does not reflect the classifier ability after it has learned from the same library and may fall into an overfitted model. Evaluation results using a single database are not necessarily an indication of the generalization abilities of the classification process and its suitability for practical applications of sound classification can be easily questioned [Livshin, 2003]. We describe the external library used for cross-validation in chapter 3.1, as explained in section 3.2 how the *Yin* output files are converted into chroma representation vectors. We run a series of tests in *WEKA* with the *SVM* function using our dataset and the external test library and the mixture of all instances as well. Table 4.5 shows the accuracy rates.

	Accuracy (%)
mixed instances	72.56
mixed instances (log)	73.56

music library / classified by	Library	Gedik's Library
Library	70.58	56.05
Gedik's Library	74.10	73.21

music library / classified by (log of vectors)	Library	Gedik's library
Library	73.12	50.86
Gedik's Library	77.67	67.85

Table 4.6: Accuracy rates from the cross-validation experiments.

In Table 4.6 we see the accuracy rates from the experiments held with the external test library. We observe that by mixing the two datasets and perform a self-classification the accuracy raises only from 73.12 to 73.56, meaning that adding more instances didn't provide much in the learning process. Moreover, we see that, when trying to classify our music library when training with the external one, the results are significantly worsen and the log transform of the chroma vectors, which in all other cases boosted the performance, here it yielded even worse results from 56.05 to 50.86.

4.5. Discussion

As a summary of the overall results we managed to achieve good accuracy rates. In both models the error rates were in acceptable ranges and especially

for the case of class-modeling approach there were classes that were almost perfectly classified.

The best results were obtained with the class-modeling approach after having applied log transformation of the feature space and using the Support Vector Machines classification technique.

Another issue that led to reasonable conclusions was the cross-validation procedure where the importance of the data in a training process was confirmed. Only by augmenting the library by 30% more instances (mix of instances) provided 0.5% to the overall accuracy. When trying to classify the smaller external library while training with our larger one the results were even higher (77.67%), when for the opposite case the results were significantly worsen (50.86%). This can only mean that the quality in combination with the quantity of the dataset plays an important role in such experiments.

Chapter 5

Conclusion

In this final chapter of the thesis, a summary of the task undertaken is presented along with conclusion, reflections and outcomes. Moreover, suggestions for future work are given on the topic of non-Western scale estimation for improving its understanding and implementation ideas are provided.

5.1. Summary of contributions

The goals of this thesis as stated in the introduction part 1.2 were:

- a) To estimate the makam scale of polyphonic makam music pieces.
- b) To accomplish the task with two different methods that have successfully been used for Western music and evaluate the success of both of them.

The study of the state of the art on the same topic led to the implementation of the first model described in chapter 3.2, using template-matching techniques to detect the scale from the chroma information. In order to have a direct comparison to previous research we restricted the number of makam types to 9, as explained in section 2.1. An important contribution in the task was to show that it is possible to work with polyphonic music signals and still achieve similar rates of success for the same makam types. This is more general for modern needs of music retrieval via large online databases. Although we cannot confirm it yet, this work showed that different morphological manifestations, tempi or rhythmic structures of this music style are not capable of confusing the system to correctly detect the scale. In

previous work on makam identification, a monophonic music library of only improvisational performances (*taksim*) was used, on the assumption that ‘clearer’ notes and intervals are used and that a better feature space can be built in such a way.

Our second goal was to evaluate the success of a class-modeling approach to the problem since it is of the most popular trends in MIR-related research. A classifier was trained using the same feature space and the results were from slightly to significantly higher than the first model. That shows the possibilities that current training algorithms have in such tasks.

As a final conclusion cross-validating the machine-learning model with an external library, we conclude the used training set has an important role in such type of research. The number of occurrences and the quality of information can boost the performance of a training system. By trying to classify the small external library while the system was trained with our dataset, the results were even higher than in auto classification process. While inverting the train-test libraries the accuracy was quite low (55%). We conclude that a broader library, in terms of instrumentation, performers and morphologies was used in our work and was more capable of characterizing the scale.

5.2. Plans for future research

The problem of scale estimation has been well studied for Western music and during the last years there have been efforts toward traditional music styles that do not use the Western equal temperament system. Something that was not carefully considered in this work is that we used techniques that are mostly suitable for Western music and they were not properly adapted to the needs of makam music, as chroma features are implemented and based on octave equivalence and Western-type pitch-classes. Although the relevant success of the two approaches for scale detection, we can be sure that if all

the known and used makam scales were taken into account the system performance would be low. To extent this work to the majority of makams (approx. 30), new features and descriptors should be introduced to yield equivalent results. An important step to make the research go further is the introduction of note sequences as a new feature that could help the classifier discriminate between makams that share similar intervals.

Different models tend to bring different kind of errors; hence an assumption is that both template-matching and class-modeling techniques can be used in a combined way to shorten the number of misclassified instances. Combination of the two confusion matrixes, provided as features in a training system, can maybe help achieve better results, something that we expect to try in the future.

References

- Harold S. Powers, et al. "Mode." *Grove Music Online. Oxford Music Online*. 21 Apr. 2010<<http://www.oxfordmusiconline.com/subscriber/article/grove/music/43718pg5>>.
- Walter Zev Feldman. "Ottoman music." *Grove Music Online. Oxford Music Online*. 21 Apr. 2010<<http://www.oxfordmusiconline.com/subscriber/article/grove/music/52169>>.
- Gómez E., Herrera P. "*Music and geography: content description of musical audio from different parts of the world*", Kobe, Japan, 2009.
- Herrera, P. Serrà, J. Laurier, C. Guaus, E. Gómez, E. Serra, X., "*The discipline formerly known as MIR*", International Society for Music Information Retrieval (ISMIR) Conference, special session on The Future of MIR (fMIR) Kobe, Japan, 2009.
- Casey et al., "*Content based Music Information Retrieval: Current Directions and Future Challenges*", Proceedings of the IEEE Vol. 96, No. 4, April 2008.
- Leman, M., Clarisse, L., De Baets, B., De Meyer, H., Lesaffre, M., Martens, G., Martens, J., and Van Steelant, D. "*Tendencies, perspectives, and opportunities of musical audiomining*", Forum Acusticum Sevilla, 2002.
- George Tzanetakis, Ajay Kapur, W. Andrew Schloss, Matthew Wright, "*Computational Ethnomusicology*", Journal of interdisciplinary Music Studies, Vol. 1, issue 2, fall 2002.
- Yarman, Ozan, "*79-tone Tuning & Theory for Turkish Maqam Music as a Solution to the Non-Conformance between Current Model and Practice*", PhD Thesis, Istanbul Technical University: Institute of Social Sciences, 2007.
- Gómez E., "Tonal Description of Music Audio Signals", PhD Thesis, Department of Information and Communication Technologies, Universitat Pompeu Fabra, 2006.
- Ismirli O., "*An Algorithm for Audio Key Finding*", Music Information Retrieval Evaluation Exchange (MIREX), 2005.
- Cornelis O., Leman M., Moelants D., "*Exploring African Tone Scales*", 10th International Society for Music Information Retrieval Conference (ISMIR), Kobe Japan, 2009.
- Chordia P., Rae A., "*Raag Recognition Using Pitch-Class And Pitch-Class Dyad Distributions*", 8th International Society for Music Information Retrieval Conference (ISMIR), 2007.
- Gedik Ali C., Bozkurt B., "*Pitch-frequency histogram-based music information retrieval for Turkish music*", Signal Processing Volume 90, Issue 4, Pages 1049-1063, 2010.

Bozkurt B., "An Automatic Pitch Analysis Method for Turkish Maqam Music", Journal of New Music Research, Vol. 37, issue 1, Pages 1-13, 2008.

Gedik Ali C., Bozkurt B., "Automatic Classification of Turkish Traditional Art Music Recordings by Arel Theory", Proceedings of the fourth Conference on Interdisciplinary Musicology (CIM08), Greece, 2008.

Gómez E., Herrera P., "Estimating the tonality of polyphonic audio files cognitive versus machine learning modeling strategies", International Society for Music Information Retrieval Conference (ISMIR), 2004.

Alpkoçak, A. and Gedik, A.C., "Classification of Turkish Songs according to makams by using n grams", Proceedings of the 15. Turkish Symposium on Artificial Intelligence and Neural Networks, TAINN, Mugla, 2006.

de Cheveigne, A. & Kawahara, H. "YIN, a fundamental frequency estimator for speech and music". Journal of the Acoustical Society of America, 111(4), 1917–1930, (2002).

C. Tanrıkorur, "The Ottoman Music", 10 Jun 2010, http://www.turkmusikisi.com/osmanli_musikisi/the_ottoman_music.htm

Livshin A, Rodet X., "The Importance of Cross Database Evaluation in Sound Classification", International Conference on Music Information Retrieval, 2003.

Krumhansl, C. L. "Cognitive foundations of musical pitch". Oxford University Press, pp. 16-49, New York, 1999.

Wei Chai, Vercoe B. L., "Automated Analysis of Musical Structure", Phd Thesis, MIT Press, 2005.