AUTOMATIC TRANSCRIPTION OF TRADITIONAL TURKISH ART MUSIC RECORDINGS: A COMPUTATIONAL ETHNOMUSICOLOGY APPROACH

A Thesis Submitted to
the Graduate School of Engineering and Sciences of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in Electronics and Communication Engineering

by
Ali Cenk GEDİK

January 2012
İZMİR

We approve the thesis of **Ali Cenk GEDİK**

Prof. Dr. F. Acar SAVACI
Supervisor

Assoc. Prof. Dr. Barış BOZKURT
Committee Member

Assoc. Prof. Dr. Bilge KARAÇALI
Committee Member

Prof. Dr. Ayhan EROL
Committee Member

Prof. Dr. Efendi NASİBOĞLU
Committee Member

27 January 2012

Prof. Dr. F. Acar SAVACI
Head of the Department of Electrical
and Electronics Engineering

Prof. Dr. R. Tuğrul SENGER
Dean of the Graduate School of
Engineering and Sciences

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my previous adviser Dr. Barış Bozkurt for his, patience, guidance, constant support and continuous encouragement throughout this research. Except the last 4 months I found the chance to study with him not only on computational music research but make music with him as well in many stages of Izmir, even at the streets for more than 4 years. I would also like to thank to my current adviser Dr. Acar Savacı not just for accepting me as a PhD candidate about to complete the thesis, but also as a colleague I could find the chance to discuss professional issues from the beginning of my PhD studentship. I feel very lucky for finding the chance to attend the courses of Dr. Ayhan Erol in department of musicology where I earn my MSc. degree long ago. He gave me the most crucial ideas about the ethnomusicological side of the thesis. The role of Dr. Bilge Karaçalı about my thesis was no doubt contribute in raising the academic standarts of my research in many ways. Finally, although I could not apply the theory and methods of Dr. Efendi Nasiboğlu's famous lectures on Fuzzy Set Theory, it is one of the nearest future plan for me. Finally, students of my lecture on music laboratory in department of musicology transferred the manual transcriptions to the computer. I am grateful to each of them.

The first three years of this research was financially supported by Scientific and Technological Research Council of Turkey, TÜBİTAK (Project no: 107E024, the automatic music transcription and automatic *makam* recognition of Turkish Classical music recordings).

Mesude'yle ortak hayat yoldaşlığımız olmasaydı bu teze başlama şansım bile olmayacaktı.

Mehlika ve Sadettin'in katkıları ise her zaman olduğu gibi bir anne ve baba olmanın fersah fersah ötesindeydi.

ABSTRACT

AUTOMATIC TRANSCRIPTION OF TRADITIONAL TURKISH ART MUSIC RECORDINGS: A COMPUTATIONAL ETHNOMUSICOLOGY APPROACH

Music Information Retrieval (MIR) is a recent research field, as an outcome of the revolutionary change in the distribution of, and access to the music recordings. Although MIR research already covers a wide range of applications, MIR methods are primarily developed for western music. Since the most important dimensions of music are fundamentally different in western and non-western musics, developing MIR methods for non-western musics is a challenging task. On the other hand, the discipline of ethnomusicology supplies some useful insights for the computational studies on non-western musics. Therefore, this thesis overcomes this challenging task within the framework of computational ethnomusicology, a new emerging interdisciplinary research domain. As a result, the main contribution of this study is the development of an automatic transcription system for traditional Turkish art music (Turkish music) for the first time in the literature. In order to develop such system for Turkish music, several subjects are also studied for the first time in the literature which constitute other contributions of the thesis: Automatic music transcription problem is considered from the perspective of ethnomusicology, an automatic *makam* recognition system is developed and the scale theory of Turkish music is evaluated computationally for nine *makamlar* in order to understand whether it can be used for makam detection. Furthermore, there is a wide geographical region such as Middle-East, North Africa and Asia sharing similarities with Turkish music. Therefore our study would also provide more relevant techniques and methods than the MIR literature for the study of these non-western musics.

ÖZET

GELENEKSEL TÜRK SANAT MÜZİĞİ KAYITLARININ OTOMATİK OLARAK NOTAYA DÖKÜLMESİ: BİR HESAPLAMALI ETNOMÜZİKOLOJİ YAKLAŞIMI

Müzik Bilgi Erişimi (MBE) müzik kayıtlarına dair erişim ve dağıtımda gerçekleşen devrimci değişimlerin sonucu ortaya çıkan yeni bir araştırma alanıdır. MBE araştırmaları şimdiden geniş bir uygulama alanını kapsamasına rağmen, yöntemleri temel olarak batı müziği için geliştirilmiştir. Batı müziği ve batı-dışı müzikler arasında ise müziğin en önemli boyutlarında temel farklılıklar olduğu için, batı-dışı müzikler için MBE yöntemleri geliştirmek oldukça güçtür. Diğer yandan etnomüzikoloji disiplini batı-dışı müzikler üzerine hesplamalı çalışmalar yapmak için önemli araçlar sunmaktadır. Bu anlamda bu tez yeni ortaya çıkan disiplinlerarası bir araştırma alanı olan hesaplamalı etnomüzikoloji çerçevesi içinde bu güçlüğün üstesinden gelmektedir. Sonuç olarak bu tezin ana katkısı literatürde ilk kez Geleneksel Türk Sanat Müziği (Türk müziği) için otomatik bir notaya dökme sistemi geliştirilmesidir. Bu sistemin geliştirilebilmesi için yine literatürde ilk kez çalışılmış olan çeşitli konular ele alınmıştır. Bu çalışma konuları da tezin diğer katkılarıdır. İlk olarak otomatik notaya dökme problemi etnomüzkoloji disiplininin perspektifinden tartışılmıştır. İkinci olarak bir otomatik makam tanıma sistemi geliştirilmiştir. Üçüncü olarak da Türk müziğinin dizi kuramı, makam tanımada kullanılıp kullanılamayacağını anlamak üzere dokuz makam için hesaplamalı olarak değerlendirilmiştir. Ayrıca, Orta-Doğu, Kuzey-Afrika ve Asya gibi çok geniş bir coğrafyanın müzikleri Türk müziği ile önemli benzerlikler göstermektedir. Çalışmamız bu batı-dışı müziklerin çalışılması için de varolan MBE yöntemlerine göre daha kullanışlı araçlar sunacaktır.

To Mesude and all my other comrades...

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Automatic music transcription (AMT) is roughly defined as the conversion of acoustic music signals into symbolic music format (e.g. MIDI) in the literature and mainly applied for music information retrieval (MIR). However, the problem definition, in other words the meaning of transcription is not well-defined within the AMT literature. Automatic transcription is usually considered as the automatization of manual transcription procedure. However, while music is visually represented by staff notation for performance or analysis in manual transription, AMT applications generally are not developed for either performance or analysis and thus do not require staff notation.

Ellingson (2011) lists conventional meanings of transcription as follows:

- Transfer of a work from one notation system to another.
- Arrangement such as adaptation of a score from orchestra to piano.
- Writing down a musical piece from a live or recorded performance.

The common point of all three meanings is the visual representation of music for either performance or analysis. On the other hand, main focus of AMT as a research domain within MIR is developing systems for the retrieval of musical pieces from large music databases. These systems require symbolic representation of musical information which mainly consists of pitch, onset time, and duration information both for the query and the database. Therefore symbolic representation of music need not to be in visual form for music information retrieval. Since the conventional meanings of transcription is based on visual representation of music for the performance or analysis, it can be said that a new meaning of transcription occurs by the AMT where the music is neither represented visually nor used for performance or analysis.

Similar to MIR studies, AMT studies also cover a wide range of applications. Thus the meaning of transcription and the output vary depending on the kind of application. Naturally, the representation of reference data for evaluation varies accordingly. In this sense, applications of AMT can be roughly grouped as follows:

- Query-by-humming (QBH)/singing/whistling/playing an instrument

- Melody and/or bass line extraction from polyphonic recordings
- Automatic transcription of polyphonic/monophonic recordings
- Automatic music tutors/ Audio to score alignment

The form of transcription ranges from simple pitch track such as f0 curve to western staff notation depending on the kind of application. However, only very few of these applications try to obtain western staff notation which requires additional information such as note names, tonality and rhythm. In this sense, automatic music tutors and few of the studies on automatic transcription of polyphonic/monophonic recordings try to obtain western staff notation. The meaning of transcription in such studies are close to the third conventional meaning of transcription in the sense the music is represented visually for the performance.

Transcription applications for automatic music tutors aims to match the performance of the user with the original notation in order to help the music student to align her/his performance visually which is also called as audio to score alignment (Mayor et al. 2009). Few of the automatic transcription of polyphonic/monophonic music applications also aim to help amateur musicians without proper music education to write down their musical compositions (Wang et al. 2003).

Despite the varying meanings of transcription in AMT, the transcription is usually defined in the literature as if the conventional meanings are used without mentioning the specific aim of the application. It is clear that the meaning and output of an automatic transcription task are quite different for retrieval applications and music tutor applications. While the representation of music makes no sense for the user in the former case, the representation of music should be conventional (eg. western staff notation) for the latter case.

However, AMT studies mostly deal with a general automatic transcription problem as the conversion of acoustic music signals into symbolic music format (e.g. MIDI) and presents only their method leaving the decision of application domain to the reader; QBH, music tutor, musicological analysis, audio coding etc. (e.g. Bello et al. 2000; Monti and Sandler 2000; Ryynanen and Klapuri 2004; Kriege and Niesler 2006; Typke 2011; Faruqe 2010; Argenti et al. 2011 etc.).

Ambiguity in the problem definition of AMT reveals itself especially when the evaluations of transcription systems are considered. Automatic transcription of a musical performance independent from the kind of application is usually compared with either original notation or manual transcription. Furthermore many studies even did not specify the source of the reference data (e.g. original notation or manual transcription) used for evaluation, also (e.g. McNab and Smith 2000; Wang et al. 2003; Paiva et al. 2004; Bruno and Nesi 2005; Fonsesca and Ferreira 2009 etc.). The problem is whether the original notation or the manual transcription can exactly match with performance due to personal interpretations of both performer and transcriber. However this point especially becomes a problem when automatic transcription is defined as obtaining original notation from performance as a kind of reverse-engineering (Klapuri 2004).

Only very few of the studies accept that original notation and transcription of a performance significantly differs (Dixon 2000, Orio 2010) and define automatic transcription as obtaining a human readable description of performance (Cemgil et al. 2004; Hainsworth 2003) which is more reasonable. Hainsworth (2003) within MIR literature figure out that manual transcription strategies can be quite different resulting various degrees of divergence from the original performance. Similarly the study of Cemgil (2004) shows that there is no unique ground truth for manual transcription even among well-trained musicians.

Finally, this thesis presents automatic transcription of monophonic instrumental audio recordings of traditional Turkish art music (shortly Turkish music). Output of our transcription system is conventional staff notation which can be used for performance and education. Therefore, our study can be considered within the context of conventional meaning of transcription. However, it should be mentioned that our aim is not to obtain original notation from performance as formulated by Klapuri (2004) as reverse engineering, rather we try to obtain a human readible description of performance as stated by Cemgil (2004).

Besides the ambiguity in the definition of automatic music transcription problem, there are serious challenging problems for developing a AMT system for Turkish music. The most challenging problem is about the fact that current techniques and methods of AMT studies are mainly developed for western music. In this sense the quality and quantity of AMT studies on non-western musics can be neglible in comparison to studies on western music. Therefore application of current techniques and methods of AMT directly to Turkish music, as a non-western music, is a challenging task based on following factors:

- Differences between western music and Turkish music in terms of pitch space, rhythm and tonality/modality.
- Divergence of theory and performance in Turkish music.

- Problems of notation system in Turkish music.
- Lack of robust MIR methods on non-western musics.

The first subsection, "*Problems of developing an AMT system for Turkish music"***,** discuss these factors, briefly. Following subsections sketch the framework of the study which also presents the outlines of the thesis, as follows:

1.2 *A framework of computational ethnomusicology (CE):* CE supplies necessary approaches for AMT of non-western musics which current MIR literature lacks.

1.3 *Automatic makam recognition and tonic pitch detection: makam* and tonic pitch of a given recording are crucial for automatic transcription. It is not possible to find a reference pitch without the determination of tonic pitch, and in order to find tonic pitch it is necessary to find the *makam* of the piece in Turkish music.

1.4 *Evaluation of scale theory of Turkish music:* Western music theory plays a crucial role in current MIR methods. Therefore we investigate whether the scale theory of Turkish music can provide a basis for MIR studies on Turkish music in a similar way western music theory provides for the current MIR studies.

1.5 *Automatic transcription of Turkish music:* Segmentation and quantization of f0 curve, determination of pitch intervals, note labelling and quantization of duration.

1.1. Problems of Developing an AMT System for Turkish Music

A number of recent studies discuss the challenging aspects of applying current MIR methods to non-western musics. With a focus on musics of Central Africa, Moelants et al. (2006; 2007) mentions three differences of African musics from western music in terms of pitch space: absence of a fixed tuning system, variable and distributional characteristic of pitches and absence of octave equivalence. Such aspects which are similar to Turkish music are also discussed by Gedik and Bozkurt (2010) in detail in a recent special issue on "ethnic music". In the same issue, Cornelis et al. (2010) and Lidy et al. (2010) discuss the challenges in a broader MIR spectrum considering the access and classification issues of non-western musics, in turn.

More specifically, the problems of applying current MIR methods to Turkish music can be shortly summarized, since they are considered in detail by Gedik and Bozkurt (2009; 2010). Figure 1.1 enables to compare the pitch-classes defined in Turkish and western music theories. While 24 pitch-classes are defined in Turkish music theory, there are 12 pitch-classes defined in western music theory as can be seen from the figure.

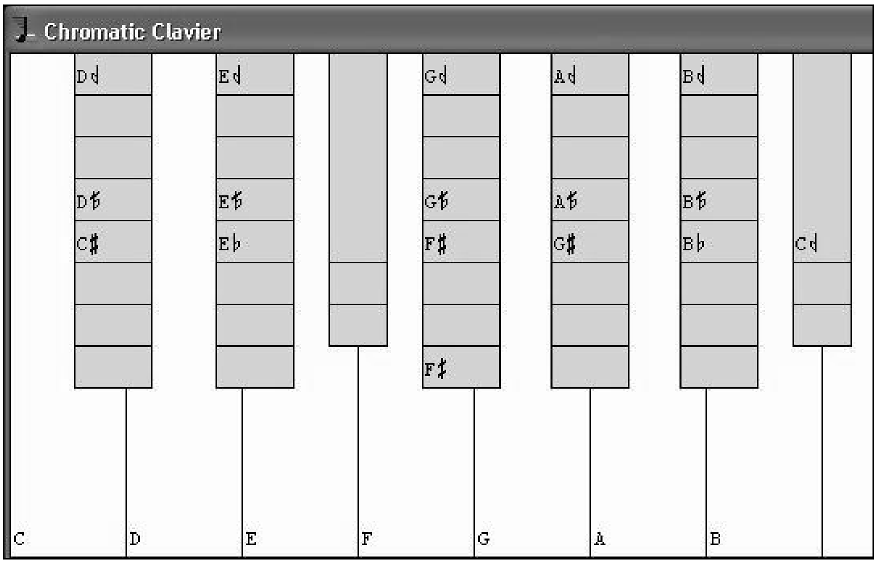


Figure 1.1. The pitch-classes defined in Arel Theory are represented at a chromatic clavier obtained by Scala software (T24 Turkish notation system of Arel-Ezgi). [1]

However, in contrast to western music, there is a divergence between theory and practice in Turkish music. The pitch interval values and the number of pitch-classes between the practice and theory of Turkish music are not in a complete accordance. It is still an open debate how many pitches per octave –propositions vary from 17 to 79 – are necessary to conform to musical practice in Turkish music. Therefore the proper representation of the pitch space is an important problem for Turkish music. Bozkurt (2008) proposed a pitch-frequency histogram representation of pitch space of Turkish music.

An example of pitch-frequency histogram is presented in Figure 1.2. Although the *cent* (obtained by the division of an octave into 1200 logarithmically equal partitions) is the most frequently used unit in western music analysis, it is common practice to use the *Holderian comma (Hc)* (obtained by the division of an octave into 53 logarithmically equal partitions) as the smallest intervallic unit in Turkish music

[1] http://www.xs4all.nl/~huygensf/scala/, Version 2.24j, Command language version 1.86i, Copyright Manuel Op de Coul, 2007

theoretical parlance. Therefore a pitch-frequency histogram of a recording is represented in terms of Hc as shown in Figure 1.2.

Instead of a tonal structure as in western music, Turkish music has a modal structure. While simple transpositions of two tonalities, major and minor, constitute the basis for the MIR studies on western music, there are most frequently used 30 distinct modalities (historically 600 *makamlar*) called as *makam* in Turkish music. The pitch frequencies in Turkish music are not based on fixed tuning as in western music (e.g. A4= 440 Hz). However only the knowledge of modality of a piece supplies a relative reference pitch name (tonic name) and thus the pitch intervals with respect to tonic. In other words, a piece from a certain modality can have different performances with different reference pitch frequencies, but the pitch intervals may remain the same. The knowledge of modality also supplies accidental signs which are necessary for the automatic transcription.

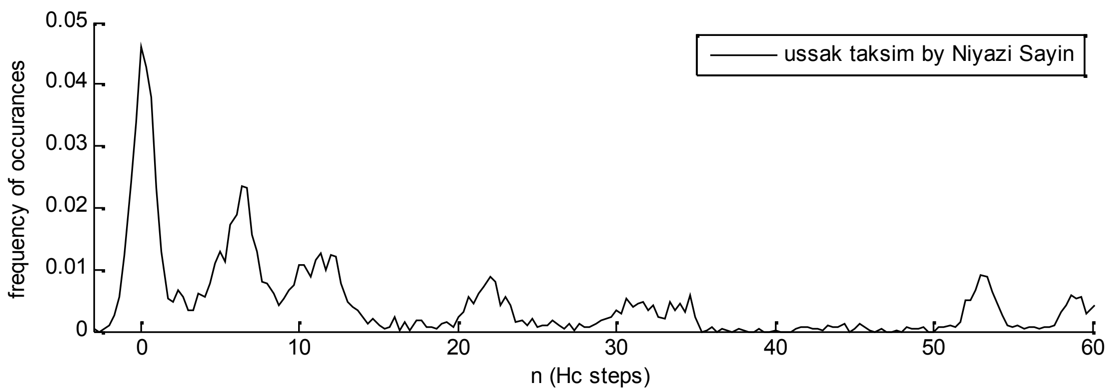


Figure 1.2. Pitch-frequency histogram of an uşşak performance by Niyazi Sayın.

Therefore, important differences in pitch spaces between western and Turkish music can be simply observed by an example of pitch histogram from Turkish music as shown in Figure 1.2. The figure presents pitch-frequency histogram of an *uşşak* performance by Niyazi Sayın. The number of pitches and the pitch interval sizes are not clear. The pitch intervals are not equal, implying a non-tempered tuning system. The performance of each pitch shows a continous space in contrast to western music where pitches are performed in fixed frequency values.

The rhythmic structure of Turkish music, involving such rhythms 7/4, 9/8, 10/4, 15/8 etc., is also much more complicated than the rhythmic structure of western music. Another important difference between Turkish and western music is about the notation system. Since notation system of Turkish music is a direct reflection of theory, the relation of notation and performance is highly problematic even for the manual transcription of Turkish music. A final important difference between Turkish and western music is the frequent use of ornamentations and performance styles as one of most the important characteristics of Turkish music which makes pitch space musch more complicated than western music. Furthermore these characteristics are not represented in notation which makes transcription more challenging for Turkish music.

Although there are few MIR studies on AMT of non-western musics, they are also far from presenting a solution for the challinging aspects of applying current MIR methods to non-western musics. A recent study reported that although there is a slight increase in the number of papers on non-western musics presented at the most important symposium of MIR community, ISMIR, within last 9 years, the percantage of non-western studies is only 5.5 % in total (Cornelis et al. 2010: 1011). Among them, only 6 papers are about the transcription of non-western musics which corresponds to less than 1 % of the papers in total. Only one paper (Nesbit et al. 2004) presents transcription of Australian Aboriginal music, consist of two simple accompaniment instruments, while other 5 papers explore specific facets of transcription problem.

These studies usually either converge the pitch space to western music or simply do not mention the characteristics of pitch space of non-western music considered. Nesbit et al. (2004) presents a very simple case of transcription of Australian Aboriginal music without facing any pitch space problem. A percussion instrument, clapstick and an accompaniment instrument producing only fundamental and several harmonic pitches, didjeridu are transcribed in this study. Since this traditional music of Indigenous Australians has no written notation, the study aims to provide a tool for ethnomusicological study.

Out of ISMIR, there are not much studies on automatic transcription of non-western musics. Al-Tae et al. (2009) considers 2 types of woodwind flute-like instrument, nay nawa and nay shabbaba from Arabian music, for a MIR system of query-by-playing within a database of Jordanian music. Although the pitch space is quite different from the western music, the system is based on approximation of all pitches to nearest pitch-classes in western music. Similarly a pitch tracking study on

Sout Indian music (Krishnaswamy 2003a) reduces the pitch space to 12 pitch-classes in western music. Kapur et al. (2007) presents a different paradigm by presenting a transcription of North Indian fretted string instrument sitar for education by the help using visual data obtained from sensors placed on the frets. However the pitch space peculiar to North Indian music is not considered in this study.

There is also a folk music research domain within MIR, which is usually considered under the “ethnic music” title which reminds “non-western musics” (Cornelis et al. 2010; Orio 2010). There are many MIR studies on folk music based on European song collections, but they are represented by western music notation sharing the same pitch space with western classical music (e.g. Huron 1995; Toiviainen and Eerola 2001; Juhász and Sipos 2010; Kranenburg et al. 2010). Among these studies there are only 2 studies dedicated to the automatic transcription task: Duggan et al. (2009) present the automatic transcription of traditional Irish tunes and Orio (2010) presents automatic transcription of Balkan and Italian songs. However, both studies deal with 12-pitch-classes of western music and consider the transcription task within a retrieval system.

As a result, current MIR literature seems to be insufficient for the development of AMT system for non-western musics. On the other hand, the discipline ethnomusicology supplies some useful insights for the computational studies on non-western musics.

Instead of considering the problems briefly presented in this subsection as an independent chapter in the thesis, each problem is considered within relevant chapters, as follows; ethnomusicological approach to the ambiguity in the definition of automatic music transcription problem, divergence of theory and practice, and problems of notation system in Turkish music are considered within Chapter 2, computational approach to differences of pitch space between western music and Turkish music, and divergence of theory and practice in Turkish music are considered in Chapter 3 and Chapter 4, respectively. Finally, lack of robust MIR methods on non-western musics is considered within Chapter 3 and Chapter 5.

1.2. Computational Ethnomusicology for AMT of Non-Western Musics

Due to the infancy of MIR studies on non-western musics, current methods developed for western music are usually applied blindly to non-western musics by engineers or computer scientists with little or no musicological considerations (Tzanetakis et.al. 2007). On the other hand, the volume of research using computational methods on non-western musics is much larger and has a much longer history within ethnomusicology than the MIR studies on non-western musics. Tzanetakis et al. (2007) review these studies and introduce a new term, *computational ethnomusicology* (CE), "to refer to the design, development and usage of computer tools that have the potential to assist in ethnomusicological research". Although Tzanetakis et al. (2007) underline the benefits of integrating MIR methods into ethnomusicological research, they use the term CE rather to emphasize an interdisciplinary collaboration of MIR and ethnomusicology.

In this sense, the problem of "transcription" of non-western musics, as well as western music, is also as old as the ethnomusicology itself. The issue was subject to hot discussions for the founders and leading figures of the discipline such as Ellis (1814-90), Stumpf (1848-1936) and Hornbostel (1877-1935), and Seeger (1886-1977). The distinction between original notation and transcription has already been defined fifty years ago by Charles Seeger in 1958 (Ellingson 1992a: 111). While prescriptive notation (original notation) defines how a specific piece should be performed, the descriptive notation (transcription) defines how a specific performance actually sounds.

Furthermore it is interesting to note that the technology for the "automatic transcription" of non-western musics within ethnomusicology is also much older than the MIR as a result of the invention of autotranscription machines by 1870s (Ellingson 1992, p. 134). Several devices were developed either for the measurement of pitch intervals or autotranscription of non-western musics such as Appunn's Tonometer (1879), Miller's Phonodeik (1916), Metfessel's Phonophotography (1928), Seashore's Phonophotograph (1932), Stroboconn (1936), Obata and Kobayashi's Direct-Reading Pitch Recorder (1937) as reported by Cooper and Sapiro (2006). However, it has been the Seeger's Melograph (1951, 1958) most widely used in ethnomusicological research for the automatic music transcription. More recently, a software mainly developed for speech analysis, PRAAT, has been used for the automatic transcription by the ethnomusicologists as suggested by Cooper and Sapiro (2006) in their survey.

On the one hand, techniques and methods of MIR for AMT are currently more advanced, compared to PRAAT in the computational sense. On the other hand ethnomusicology as a musicological programme rooted in the research on non-western musics, has already solved methodological problems long ago such as avoiding the use of western musical concepts for non-western musics, an example of ethnocentrism, in the emerging years of the discipline. The problem of ethnocentrism is exactly what the MIR research experiences almost whenever non-western musics are considered even by the “insiders”.

The qualitative methods of ethnomusicology and quantitative methods of MIR could be another collaboration point between the two disciplines. Especially the quantitative approach of MIR toward evaluation makes the details of the process inaccessible. On the contrary the methods of ethnomusicology are mainly qualitative which supplies details of a procedure for any musical event. As a result, the perspective of ehnomusicology presents a solution for the ambiguity of the problem definition in AMT literature. Furthermore, the perspective of ethnomusicology also supplies necessary approaches to many facets of this problem related with Turkish music such as divergence of theory and practice, and problems of notation system in Turkish music. In this sense our study tries to establish this interdisciplinary connection between ethnomusicology and MIR for the automatic transcription of traditional Turkish art music.

Finally, Chapter 2 presents this ethnomusicological framework which subsequent chapters are based on. Briefly, the ethnomusicological perspective toward the transcription problem, the divergence of theory and practice in Turkish music and the problems of Turkish notation system are presented in Chapter 2. A brief ethnomusicological case study on manual transcription is also presented at the end of this chapter.

1.3. Automatic *Makam* Recognition

As aforementioned, *makam* and tonic pitch of an audio recording are crucial for automatic transcription in Turkish music. Furthermore knowledge of *makam* also provides the accidentals to be used in the transcription. It is not possible to find a reference pitch without the determination of tonic pitch and in order to find tonic pitch it is necessary to find the *makam* of the piece in Turkish music. However, firstly f0 data should be extracted for any operation on pitches. Representation of pitch space for Turkish music recordings were presented by Bozkurt (2008) for the first time. F0 data is extracted by the YIN algorithm (de Cheveigne and Kawahara 2002) with post-filters designed by Bozkurt (2008) to correct octave errors and remove noise on the f0 data. Then Bozkurt (2008) presented the pitch-frequency histogram representation of Turkish music and automatic tonic detection.

In the MIR literature on western music, tonality of a musical piece is found by processing pitch-class histograms which simply represent the distribution of 12 pitch-classes performed in a piece. In this type of representation, pitch-class histograms consist of 12 dimensional vectors where each dimension corresponds to one of the 12 *pitch-classes* in western music. The pitch-class histogram of a given musical piece is roughly compared to templates of 24 tonalities, 12 major and 12 minor, and the tonality whose template is more similar is found as the tonality of the musical piece.

The construction of the tonality templates is mainly based on three kinds of models in the literature: music theoretical (e.g. Longuet-Higgins and Steedman 1971), psychological (e.g. Krumhansl 1990) and data-driven models (e.g. Temperley 2008). These models were also initially developed in the studies based on symbolic data. However, neither psychological nor data-driven models are fully independent from western music theory. In addition, two important approaches of key-finding algorithm based on music theoretical model use neither templates nor key-profiles: the rule-based approach of Lerdahl and Jackendoff (1983) and the geometrical approach of Chew (2002).

As a result, we apply template matching for finding the *makam* of a given Turkish music recording. However, a data-driven model is chosen for the construction of templates due to the lack of either psychological models or a relaible theory in Turkish music. Similar to pitch-class histogram based classification studies, we use a template matching approach for *makam* recognition using pitch-frequency histograms (see Figure 1.2). We used pitch-frequency histograms for the representation of pitch space of Turkish music. The template for each *makam* type is simply computed by averaging the pitch-frequency histograms of audio recordings from the same *makam* type after aligning all histograms with respect to their tonics. Figure 1.3 shows 2 histogram templates of *makamlar rast* and *uşşak*.

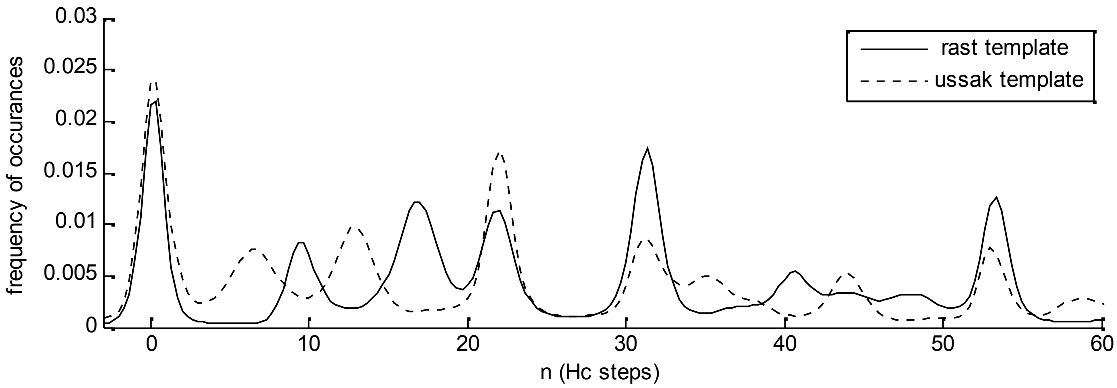


Figure 1.3. Pitch-frequency histogram templates for the two types of melodies: *rast makam* and *uşşak makam*.

Thus, each recording's histogram is compared to histogram templates of the *makam* types and the *makam* type whose template is more similar is found as the *makam* type of the recording for automatic *makam* recognition. As an example, pitch-frequency histogram of a hicaz recording shown in Figure 1.2 is compared to the two makam templates, *rast* and *uşşak* shown in Figure 1.3. The most similar *makam* template is found as *makam uşşak* which gives name of the *makam* of the recording. Since both makam recognition and tonic detection base on matching a histogram with a template, these two steps are indeed performed by a single histogram matching operation. Therefore, since the tonic of each makam template is given, automatic *makam* detection also supplies automatic tonic detection.

The distance between pitch frequency histograms are measured by City-Block (L1 norm) distance. 172 recordings of 9 *makamlar* which represent 50% of the current Turkish music repertoire are used in the study. Leave-one-out cross validation method is applied for evaluation and success rate is found as 68 % in terms of F-measure for automatic *makam* recognition.

Finally, the details of the automatic *makam* recognition and tonic detection and a comprehensive review on the use of pitch-class histograms in MIR studies both for western and non-western music in comparison with Turkish music are presented and lack of robust MIR methods on non-western musics is discussed in Chapter 3.

1.4. Evaluation of Scale Theory of Turkish Music for MIR

In this part of the study, our main motivation is to investigate whether the scale theory of Turkish music can provide a basis for automatic makam detection in Turkish music in a similar way western music theory provides a basis for the current modality deection studies. Western music theory plays a crucial role in current MIR methods, especially for the representation of the pitch space as equal tempered 12 pitch-classes. In this sense, we try to understand whether scale theory of Turkish music can provide such valid pitch-class definitions for MIR studies on Turkish music. However, there are several different theories of Turkish music where the number of pitch-classes varies from 17 to 79 (Yarman 2008).

As a result, we consider the most influential theory in Turkish music developed mainly by Hüseyin Sadeddin Arel (1880-1955). Arel theory is an official theory for music education, and musical notations and transcriptions are also written according to Arel theory in Turkey. On the other hand, the discussions about the divergence between the theory and the practice are also mostly held with respect to Arel theory, especially about the defined *makam* scales. Therefore, both for the research in MIR and ethnomusicology, Arel theory is worthy of investigation. Consequently, we have evaluated the *makam* scale theory of Arel.

The automatic *makam* recognition method and the data set summarized in Subsection 1.3 are used for the evaluation. Since the theory defines fixed pitch intervals for each *makam* scale, we have represented theoretical pitch intervals for each *makam* as a sum of Gaussian distributions as shown in Figure 1.5. The mean of each Gaussian distribution was set at the fixed pitch interval values defined in the theory for each *makam*, and their standard deviations were selected as 2 Hc, nearly half a semitone, heuristically.

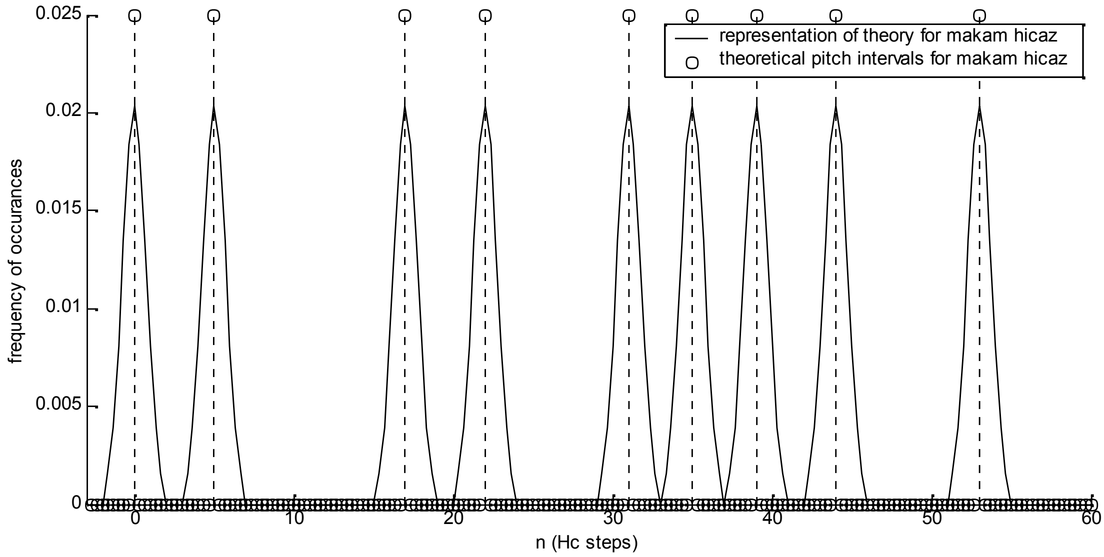


Figure 1.4. Representation of hicaz *makam* scale defined in Arel theory as sum of Gaussian distributions.

Several pitch intervals was found to be lacking in the theory in comparison to pitch intervals in practice. As a result, the success rate of 64 % in terms of F-measure is found which is 4 % less than the success rate of data-driven model summarized in Subsection 1.3.

Another *makam* recognition model is applied where new templates are constructed by using the pitch intervals and weights obtained from the templates of the data-driven model for new Gaussian distributions. This new automatic recognition model outperformed data-driven model. The success rate of automatic *makam* recognition based on this new model was found as %75 in terms of F-measure, %7 better than the success rate of data-driven model.

Finally, both the divergence of theory and practice is evaluated and a more successful automatic *makam* recognition model is designed for our automatic transcription system. The details of this study are presented in Chapter 4.

1.5. Automatic Transcription of Turkish Music

Automatic transcription of Turkish music as a problem, mostly demonstrates resemblance with automatic transcription of singing, humming or performance of fretless pitched instruments such as violin within MIR studies, due to the resulting continuous pitch space. As Ryynanen (2006) mentioned most of the singing transcription applications are designed as the front-end of QBH systems in contrast to our study. The most challenging task in singing transcription is converting a continuous f0 curve to note labels (Ryynanen 2006: 362). However, despite the resemblance of pitch-spaces in singing and Turkish music, it should be kept in mind that it is always a matter of quantization of the f0 curve to the nearest pitch-class in western music. Of course a simple rounding operation gives poor results for quantization of f0 curve, depending on the following two important characteristics of singing:

- The performance of a singer can result with deviation of its frequency from the reference frequency in time.
- Performance of ornamentations such as vibrato, legato and glissando which are not possible in fretted instruments.

  Since we are interested in instrumental recordings in Turkish music, the first characteristic is out of our scope. The second characteristic is one of the most important characteristic of Turkish music as aforementioned. However ornamentations also take little attention in the litrature.

Automatic transcription task is roughly consist of three steps: extraction of f0 information, segmentation of f0 curve and labeling each segment with note names. There are various methods for the extraction of f0 information: methods based on time-domain, frequency domain or auditory model. Methods for segmentation and labeling of f0 curve mainly follow two approaches: cascade approach where f0 curve is first segmented and then labeled, and statistical method where segmentation and labeling are jointly performed (Ryynanen 2006: 363).

The most popular statistical method for automatic transcription is Hidden Markov Modeling (HMM). However, as mentioned by Orio (2010) the use of HMM for automatic transcription requires collection of scores for training HMM which are hardly available for non-western musics. In order to obtain training data for HMM the use of manual transcriptions is also problematic for non-westen musics. Manual transcription of non-western musics either requires existence of a notation system or a notation system in accordance with performance as in western music.

Therefore we preferred cascade approach in our AMT system as shown in Figure 1.5. The system accepts monophonic audio recordings of instrumental Turkish music. After the extraction of f0 data, pitch-frequency histogram is calculated in order to find the *makam* (modality) and tonic pitch of the piece. Both the knowledge of *makam* and tonic pitch are crucial for transcription, since without the determination of tonic pitch, it is not possible to find a reference pitch in Turkish music. It is obvious that pitch intervals can be only found with respect to a reference pitch. However, in order to find tonic pitch it is necessary to find the *makam* of the piece, since each *makam* has a relative tonic pitch and definite note name for that tonic picth. Therefore, automatic *makam* recognition supplies both f0 value and name of the tonic pitch. Knowledge of *makam* also provides the accidentals to be used in the transcription.

F0 extraction is applied as presented in Subsection 1.3. Automatic *makam* recognition and tonic detection are applied as presented in Subsection 1.4. Therefore, it is possibble to express f0 curve with respect to tonic pitch and then to obtain pitch intervals. This operation is applied after converting the f0 curve to Hc. Then the value of tonic pitch is substracted from the f0 curve. In order to label resulting f0 curve by note names, firstly it is necesarry to segment the f0 curve. Segmentation corresponds to finding the onset of the notes. Secondly f0 curve within each segment is quantized which corresponds to eliminating ornamentations such as appagiatura, acciaccatura, vibrato and glissandos. Rule-based approach is applied for segmentation and quantization where parameters are heuristically determined depending on the musicological knowledge peculiar to Turkish music.

After segmentation and quantization, representation of pitch intervals in terms of Hc gives a resolution of 53 Hc/octave for f0 curve which is much bigger than the number of pitch classes defined in theory as 24 pitch-classes/octave. Since notation system of Turkish music is a direct reflection of theory and in order to obtain a readable notation, pitch intervals are converted to the nearest pitch-classes which have distinct names for 2 octaves in theory. As the last step before transcription, note durations corresponding to the segment lengths are quantized by using duration histogram. Finally note names, onset time and note durations are used as an input to a notation software

MUS2[2] which is specifically designed for Turkish music and outputs conventional Turkish music staff notation. Since each block has a definite success rate, GUI enables user to correct any faulty information such as makam name, tonic pitch etc. İn order to obtain a more robust transcription result.

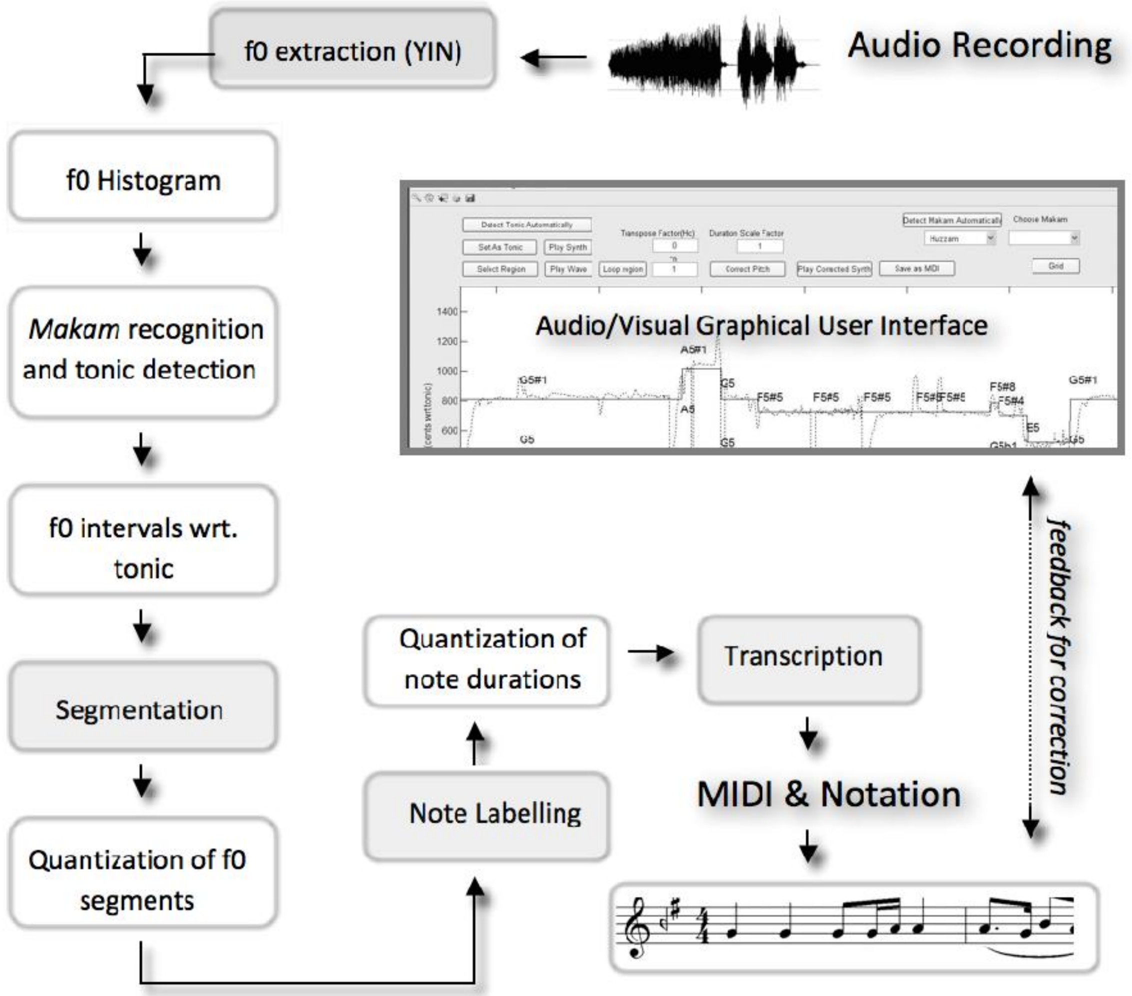


Figure 1.5. Block diagram of AMT system for Turkish music.

As a result, while our automatic transcription system outputs conventional notation which corresponds to prescriptive notation, the GUI supplies descriptive notation where details of a recording can be observed on f0 curve in comparison to

[2] http://www.musiki.org/

parameters of prescriptive notation such as note names, duration information and onset/offset times.

Finally, 5 recordings are used for evaluation. Manual transcriptions of 2 musicians and automatic transcriptions are evaluated with respect to original notation. While automatic transcription outperforms manual transcriptions for 2 recordings, success rates of automatic transcription for the rest of 3 recordings are found close to the success rates of manual transcription. The study and qualitatively evaluation of the results are presented in detail in Chapter 5.

1.6. Contributions

Main contribution of the thesis is the design of an AMT system for Turkish music for the first time in the literature. Secondary contributions of the study are the approaches, methods and techniques developed also for the first time throughout the research for automatic transcription of Turkish music. These contributions can be listed as follows:

- An interdisciplinary approach for the study of automatic transcription of non-western musics which synthesize qualitative methods of ethnomusicology and quantitative methods of MIR.
- Automatic *makam* recognition.
- Evaluation of scale theory of Turkish music for nine *makamlar* in order to understand whether it can be used for *makam* detection.

Finally, output of our AMT system corresponds to the conventional meaning of transcription, since we try to obtain conventional staff notation from recordings of Turkish music for the purposes of performance and education. Since this kind of AMT application covers the most comprehensive information, output of our study would enable other kinds of applications for Turkish music also such as retrieval and ethnomusicological analysis. Furthermore there is a wide geographical region such as Middle-East, North Africa and Asia where the musical cultures shares close similarities with Turkish music. Therefore our study would also provide more relevant techniques and methods than the MIR literature for the study of these non-western musics.

CHAPTER 2

ETHNOMUSICOLOGICAL FRAMEWORK

2.1. Basic Concepts of Turkish Music[3]

Traditional musics of wide geographical regions, such as Asia and Middle East, share a modal system in their musics instead of tonal system of western music. In contrast to tonal system, the modal systems of these non-western musics cannot be only described by scale types such as major and minor scales as in western music. Modal systems lie between scale-type and melody-type descriptions in varying degrees peculiar to a specific non-western music. While the modal systems such as *maqam* in Middle East, *makom* in Central Asia and *raga* in India are close to melody-type, the *pathet* in Java and the *choshi* in Japan are close to the scale-type (Powers 2008). In this sense, the *makam* practice in Turkey, as a modal system, is close to the melody-type, and thus shares many similarities with *maqam* in the Middle East.

Turkish traditional art music is basically classified into several *makamlar,* both in theory and in practice. Each *makam,* having a distinct name, generally implies a set of rules for composition and improvisation. These rules are roughly defined in theory in terms of the scale type and the melodic progression (*seyir*). Although there is a general consensus about the names of *makamlar,* at least in practice, the rules that define them remain problematic.

However, the definitions and the number of the *makamlar* have greatly changed throughout the history. While the number of *makamlar* is stated as 27 in the treatise of Dimitrie Cantemir (17th c.), Arel (1993) defines 113 *makamlar*. The defining rules of *makamlar* are also considerably altered in the Arel theory, such as the abandonment of the traditional concepts and classification categories *avaze*, *şube* and *terkib*. On the other hand, Öztuna (2006) reports that historically, there have been as many as 600

[3] This section is adapted from Gedik, A. C. and Bozkurt, B.(2009). Evaluation of the Makam Scale Theory of Arel for Music Information Retrieval on Traditional Turkish Art Music, *Journal of New Music Research*, 38(2): 103-116.

makamlar, but only one sample for each of the 333 *makamlar* are left today, and approximately seventy percent of the current repertoire consists of only 20 *makamlar*.

Form provides an additional classification for Turkish music in theory and in practice. Each composition has a distinct *makam* name such as *hicaz*, *saba*, *nihavend* and a distinct form such as *peşrev*, *sazsemai*. So each composition is referred to as *hicaz peşrev*, *saba sazsemai* etc, where the *makam* name is followed by a form name. The *usul*, the rhythmic structure of a composition such as *aksak* (9/8), *semai* (3/4) etc., is also mentioned in the naming of compositions. Improvisation is considered as a free-rhythmic form and classified as instrumental (*taksim*) and vocal (*gazel*).

2.2. The Divergence of Theory and Practice in Turkish Music[4]

The divergence between theory and practice, by no chance, is common[5] in the traditional art musics of the Middle East, where practice is mainly based on oral tradition and theory is meant to be speculation and science of music (Bohlmann 2008). Nonetheless, this fact appeared as a "problem" due to the westernization and nationalization by the 20th century, which also try to bring standardization in music. However, the lack of standardization in the production of instruments seems to lead to the discussions about the divergence of theory and practice.

Two representative examples of the westernization and the nationalization of music are Egypt and Turkey. The Congress of Arab Music held in Cairo in 1932 is an historical attempt to standardize the theory and practice of traditional art music[6] (Racy 1991:68). Although nationalism was not very explicitly present in the congress, the term "Arab music" was clearly implying a distinction from Turkish and Persian musics (Thomas 2007:2). The cultural policies of the government in Egypt intended both to define an "Arab music" and raise it to the "level" of western music (Racy 1991:70) in accordance with the general top-down direction of westernization and nationalization processes.

On the other hand, the same processes followed a different course in Turkey. Few years after the 1923 revolution, educational institutes of the traditional art music

[4] This section is adapted from Gedik and Bozkurt (2009)
[5] Even as early as in the 13.c., the theory of Urmevi slightly diverted from the practice of his time (Marcus, 1993:50).
[6] The term "traditional art music" is used to refer the relevant musics of Egypt and Turkey.

such as official schools, religious lodges and cloisters were closed (Tekelioğlu 2001:95). This music was regarded as a symbol of Ottoman past, which implies a primitive, morbid, non-rational, non-western and non-Turkish heritage, blurred with Arab, Persian and ancient Greek effects (Signell 1976:77-78). Thus, the new Turkish music was defined as the synthesis of "pure" Turkish folk music and western classical music. Still this neither led to the disappearance of traditional art music nor to the prevention of its own westernization and nationalization. This can be considered as a characteristic of late modernization: the concurrent existence of modernity and traditionality and/or hybrid structures.

The music theorists did not follow the cultural policies of the state, and developed new discourses and theories based on the "Turkishness" and "westernness" of traditional art music. Despite the ideological and physical interventions of the state, even the radio broadcasting of traditional art music was banned between 1934 and 1936, these theories and discourses were started to prevail among the theorists and the musicians. However, the political climate of Turkey after 1940s changed and seemed to become more tolerate towards traditional art music (Öztürk 2006b:153). The journal, *Musiki Mecmuası* (founded by Arel in 1948) and semi-official and unofficial schools of traditional art music played a crucial role in the appreciation of these theories and discourses. Nevertheless, traditional art music is not officially recognized until 1976 by the foundation of the first Conservatory of Turkish Music. Only after this event, the current theories and discourses were also officially recognized and appreciated, and thus constituted the basis of national education of the traditional art music. Therefore, these theories and discourses have been much more prevailed and established after 1976.

It should be added that neither the Arab music congress nor the Turkish revolution was a sudden turning point for the westernization of traditional musics. Westernization dates back to the 19th cenutry, both in Egypt and Turkey: Khedive Isma'il (1830-1895), a reformist ruler of Egypt, and Selim III (1761-1808), a reformist Ottoman emperor, were both patrons of music, interested in western and traditional musics and took important steps toward westernization of musical life. So the new theories and discourses in Turkey can be considered as a continuation of the trends started in the 19th century. Furthermore, two of the most influential modernist theorists of the 20th century, Rauf Yekta Bey (1871-1935) and Arel were also the "students" of the heads of dervish lodges (Akdoğu 1993:xii).

The study of Yekta on the westernization of the theory provides a historical turning point. The term “Ottoman music” is replaced by “Turkish music”, and the traditional number of intervals is increased from seventeen to twenty-four (Öztürk 2006a: 213-214). However, his colleague Arel went much further in trying to “prove” both the Turkishness[7] of the traditional art music and its resemblance to western music. He invented new instruments (soprano, alto, tenor, bass and double-bass *kemençe*) and gave *makam çargah*, which has only one piece in repertoire, a central role in his new theory due to its equivalance to scale of C major in western music.

Feldman (1990:100) compares the positions of Yekta and Arel as follows: while Yekta appears to be more involved with musicological works, Arel plays the main role in the ideological struggle against the cultural policies of the state which rejects traditional Turkish art music. Nevertheless, it should be noted that Yekta had already written an explicit answer against the arguments of the cultural policies in his 1925 articles (Yekta 1997a:5-7; 1997b:33-34) twenty years before Arel. However, Arel seems to exceed the logical limits of past trends both theoretically and discursively in the 20th century.

Arel theory was first published as a book in 1968 after its earlier publication as articles in 1948, though Zeki Yılmaz’s book, published in 1977, which is a simplified and somewhat distorted version of the Arel theory, has prevailed as if it was an official textbook. Shiloah (2008) describes a similar tendency in Egypt after the second half of the 20th century as an shift of interest from theory to practical theory. Therefore, Arel theory is not much known in detail today, except among theorists and few musicians.

The main problems of Arel theory can be listed as follows (Öztürk 2006a:214-216):

- *makam çargah,* has been given a central role and attributed as a general scale, which is identical to the C major scale and tonality in western music. The hierarchical tonal functions are attributed to the specific scale “degrees” and a new notation system similar to western staff notation is introduced.
- One of the most important aspects of Turkish music, the melodic progression (*seyir*), is underestimated. Therefore, the *makam* concept is reduced to a tonal scale as in western music.

[7] All past theorists are considered as ethnic Turks, although many of them were non-Ottoman or even non-Turkish.

Stokes (1996) also refers to these attempts as the "Arel project" in reference to its strong relations with nationalization and westernization. However, there is an increasing tendency toward criticizing Arel theory today, especially among the theorists because of its divergence from the practice.

As a result, the westernization and the nationalization of the theories and discourses have become more established by the official institutions founded in Turkey and in Egypt after the second half of the 20th century. Thus, the divergence between the theory and the practice became more apparent and problematic in countries due to the officially institutionalized common discourse: "the theory should generate practice" (Thomas 2007:4). Especially the standardization of tuning system as equal-tempered quarter-tone scales in Egypt and as division of the octave into 24 unequal intervals in Turkey generates similar new discourses among musicians: Pitch interval values are performed differently than the ones defined in theory, and musicians describes this flexibility with respect to the theory by using such terminology as "a little higher", "a little lower" or "minus a comma" (Marcus 1993). Unstandardized fret positions in the production of instruments such as *kanun* and *tanbur* explicitly provide evidence for these flexible pitch preferences of performers in Turkey (Yavuzoğlu 2008:12).

On the other hand, although the performances diverge from the theory, the Arel theory is highly respected among performers, and they hesitate to contradict the theory when the pitch intervals of their performances are measured by musicologists[8].

2.3. Perspective of Ethnomusicology towards Transcription Problem

Since existence of a notation system is a prerequisite for any transcription, it is necessary to define the concept of notation first. Notation is shortly a communication system between musicians either in written or in oral form. However, oral notation is out of our scope, since our focus is transcription. Besides communication, notation also helps musicians to remember a much greater repertoire which otherwise not possible to memorize. (Bent et al. 2011)

In this sense, first transcription attempts were for the purpose of preserving musical cultures without notation at the beginings of 19th century (Nettl 1982: 67). The

[8] Karl Signell and M. Kemal Karaosmanoğlu (quoted from Can Akkoç) shared their measurement experiences with foremost performers Necdet Yaşar and Niyazi Sayın, respectively. (personal communication with Signell and Karaosmanoğlu, 6-8 March 2008, İstanbul)

first folk song collections in Europe with the same motivation also encounter the first problems of transcription about using a notation system not designed for the transcribed music. Therefore, these folk song collections, also used in MIR studies consist of distorted versions of the original songs (Burke 2009: 44-45). Transcription for the purpose of analysing non-western musics and comparing it with western music emerges by the foundation of the discipline ethnomusicology. By the end of 19th c. it was widely accepted that use of European notation for non-Eropean music cultures was inadequate (Ellingson 1992a: 117).

Transcription, from the ethnomusicological point of view rather corresponds to the description of a musical piece. On the other hand, notation corresponds to representation of musical features for the purpose of prescription (Ellingson 1992a: 153). Therefore, transcription and notation are interrelated concepts since transcription is only possible for a definite notation system. As a result, both are crucial concepts for the automatic transcription which take little attention within literature of AMT.

One of the milestones for the discussions of transcription in ethnomusicology is the distinction suggested by Seeger. Notation is classified either as prescriptive or descriptive by Charles Seeger in 1958 (Ellingson 1992b: 111): prescriptive notation defines how a specific piece should be performed and the descriptive notation defines how a specific performance actually sounds.

Nettl (1982: 69) also suggests a similar approach; the prescriptive notation provides information about only the piece, not the style, to the native of that musical culture (insider) even in western music; in other words in order to perform a mazurka of Chopin from notation, one has to be familiar with the literature of Chopin and gain the knowledge of how Chopin sounds. On the other hand the descriptive notation tries to provide an "objective" analytical insight of the piece to the researcher (outsider) who is not native of that musical culture. Thus, prescriptive notation provides information only sufficient for a native to perform. This fact implies impossibility of a complete correspondence between notation and performance as suggested by the perspective of MIR.

However the concept of transcription as used in AMT corresponds to descriptive notation since the procedure as applied aims to obtain original notation from recordings of performance. Klapuri (2004) summarizes the aim of automatic transcription in MIR as reverse-engineering which try to obtain the original prescriptive notation or "source code" from recordings of performance. Therefore the perspective of MIR clearly results with disappearance of the important distinction between a notation of a piece and a transcription of a performance even for western music.

As Ellingson (1992a: 154) discussed, performance need not be strictly the same as the dictations of notation: " 'Prescriptive' seems to be too strongly normative and hierarchical a term to characterize some significant communications to performers about musical sounds, communications that might be better conceived as 'suggestive', 'advisory', 'interactive', and even 'inspirational', rather than prescriptions dictated to performers."

Nettl (1983: 69) also discussed that "It is 'insiders' who write music to be performed, and they write it in a particular way. Typically, outsiders start by writing everything they hear, which turns out to be impossible." Instead of understating the distinction, Nettl rather tries to reveal that a fully "objective" descriptive transcription is not possible, since any visual representation of music is an abstraction. Similarly, Turkish ethnomusicologist Erol (2009: 190) mentions that anyone whom was not familiar with a specific musical culture would be helpless in either interpreting a notation or transcribing a musical sample.

Finally, the emprical studies of List (1974) and Stockmann (1979) discussing the reliability of manual transcriptions shows that different participants gives out transcriptions with a certain amount of difference primarily for the durations and secondarily for the pitches of notes.

2.4. The Notation System of Turkish Music

One of the obstacles against developing an AMT system for Turkish music is the meaning of notation in this musical culture which is still mainly based on an oral tradition. Oral tradition in Turkish musical culture, called as *meşk* system, is historically the learning process of a music student face to face with a master musician based on memorization of the repertory without any use of notation.

Although the indroduction of western like or western notation for the representation of Turkish music dates back to 17th century, these first attempts of Albert Bobowski/Ali Ufki and Dimitrie Cantemir/Kantemir were mainly served as preservation of the repertory. The first use of the western notation for performance was only available at the beginings of 19th century as a result of the westernization processes by the Ottoman court limited with the court musicians. These musicians were already familiar with another notation system called Hamparsum a decade ago, derived from Armenian neumes, which consists of only sequence of letters denoting pitch names and duration information without any staff, and thus quite different from the western notation.

Therefore the use of western notation was seemed to be a simple matter of learning the corresponding symbols found in Hamparsum but resulted with a dichotomy in practice: existence of *meşk* and western notation side by side (Ayangil 2008: 416). This attempt was limited with the court musicians, a small group compared to the much larger community of musicians out of the court. It was only possible by the end of 19th c., western notation adapted to Turkish music started to be used more widely (Ayangil 2008: 418). After various attempts of adapting western notation system to Turkish music, western-based notation system of Arel-Ezgi-Uzdilek (shortly Arel theory) became an official system by the 1970s and began to be thought in public and private schools extensively.

Nevertheless, the divergence of theory and practice about the pitch space stand at the center of discussions about notation and thus *meşk* system has never been left completely. This fact lead to a hybrid education system from the 19th century up to today. In other words the pitch space represented in practice and in the notation system does not converge completely, requiring verbal explanations and musical demonstrations where the *meşk* system takes role. Another reason for the indispensable role of *meşk* system results from one of the most distinctive character of Turkish music, the quite intense use of ornamentations and performance styles which are not represented in notation, again (Ayangil 2008: 441).

As shown in Figure 2.1 the interval of a major second or whole tone (204 cents) is divided into 9 equal parts ("Comma value" row) in theory. In other words an octave is divided into 53 equal parts where each part is called as an Holdrian comma and a subset of 24 notes are used among these 53. The resulting tuning system is 24 tone non-tempered system. In contrast to two types of accidentals in western music, there are four kinds of sharps and corresponding four kinds of flats which are used to represent 24 pitch-classes in Turkish music. Nevertheless, these accidentals fail to cover all pitch intervals performed in practice.

Type of Interval	Koma	Bakîye	Küçük mücenneb	Büyük mücennep	Tanini	Artık ikili	Naturel
Sharp	♯	♯	♯	♯	×	{♭ ♯}	
Flat	♭	♭	♭	♭	♭♭		
Letter	F	B	S	K	T	A	♮
Comma Value	1	4	5	8	9	12	
Ratio	81/80	256/243	2187/2048	65536/59049	9/8	576/486	
Cent Value	23	90	114	180	204	271	

Figure 2.1. Accidentals in notation system of Arel theory
(Source: Ayangil 2008: 426)

As a result the western-based notation system of Arel theory is simply the application of these accidentals to the western staff notation as shown in Figure 2.2.

Figure 2.2. Staff notation of a composition by Arel, showing only the first line.

Only G-clef is used in AEU notation as shown in Figure 2.2 Furthermore the pitch D4 (*neva*) is tuned to 440 Hz in common practice, instead of A4 as in western music. There are 13 standart tunings, called as *ahenk* in Turkish music defined by the 13 *neyler* (sl. *ney*) (flute-like woodwind instrument) with different but standard sizes.

The key signature as shown in Figure 2.2 also does not indicate either tonality as in western music or modality in Turkish music, since there are sevral modalities sharing the same accidentals. However, the modality of the piece is indicated at the title of the song, "*Hüseyni*", also implies the tonic of the piece *dügah* (A4). The information about form is also written at the title, "oyun havası". Similarly, although time signature 7/8 is written as in western music, the rhytmic structure (*usul*) is also written by words as devr-i turan, since the same time signature can have different beats. Other signs about form and tempo such as segno and metronome, and about dynamics such as crescendo, decrescendo and mezzoforte (mf) and about articulation such as staccato and ties are used as in western music. Finally name of the composition, “Düğün Evinde” and name of the composer, Hüseyin Sadeddin Arel are also mentioned in the notation as shown in Figure 2.2.

2.5. Comparison of Notation and Performance in Turkish Music

Asaforementioned, notation system of Arel theory naturally reflects the theory of Arel theory. Nevertheless, the theory of Arel theory does not reflect the practice of Turkish music appropriately in terms of pitch space as discussed by Gedik and Bozkurt (2009) in detail. The divergence of theory and practice in Turkish music can be observed from Figure 2.3.

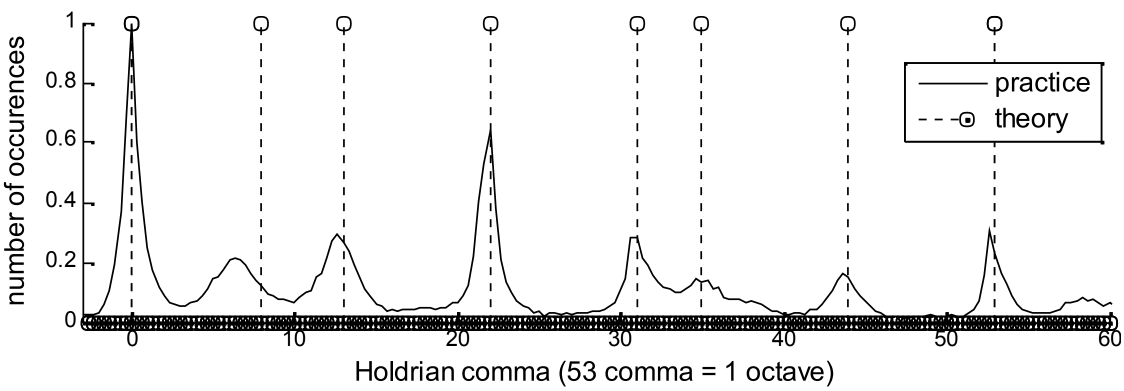


Figure 2.3. Comparison of pitch spaces defined in theory and performed in practice for the makam *uşşak*.

It has also been shown that the 25 pitch-classes defined by the Arel theory are lacking two pitch intervals, and six pitch intervals also diverges from the defined pitch-classes in theory. The reasons of the divergence of theory and practice in terms of pitch space can be listed as follows (Gedik and Bozkurt 2009:107):

- The freedom of musicians in performance of a specific *makam* by varying the pitches for certain pitches of the *makam* scale.

- The small variations of pitches performed depending on the direction of melodic progression, either descending or ascending.

Ayangil (2008: 443-445) also discussed the problems of notation system of Turkish music in detail from the musicological point of view:

- Inaccuracy of representing pitch classes: “Yet, the musicians who have a good understanding of the system of makams and pitches attain almost absolute accuracy in the performance of makams and pitches, inspite of the relativity and inaccuracy of the notation system and its alteration signs.” (Ayangil 2008: 445)
- *Ahenk* system: Although there are 13 posssible transpositions, performers frequently use only 2 of them for practical reasons such as pitch range of vocal and instruments. However the notation system does not reflect any transposition and performers had to apply the transposition by using their musical skills, not by the notation.
- Performance styles and ornamentations: While the performance styles such as melodic and rhythmic variations and ornamentations constitutes one of the most peculiar characteristic of Turkish music, they are not represented in notation.

Kaçar (2005) discussed this last item by comparing the notation of pieces and the the performances of pieces by master musicians. According to Kaçar notation system of Turkish music leaves much more freedom to the performer for the interpretation of a composition, in comparison to western music where a composition is more strictly defined by the notation by 19th the century. Even the composers of Turkish music performs their own compositions different than the notation. Notation functions as if it is a framework of composition in Turkish music (Kaçar 2005: 216).

According to The New Harvard Dictionary Of Music ornamentations are classified as follows (Kaçar 2005: 216):

- Insertion of additional notes into melody
    - Insertion of small durational notes
    - Ornamentations such as changing note durations
    - Insertion of notes into tonal pitches
- Ornamentations based on various variations
- Ornamentations based on tempo and note duration changes such as ritardando, rubato and cadence

Figure 2.4. Two bars from a composition of Tanburi Cemil Bey, "Şedaraban Saz Semaisi". The first line is the notation of the composition and the second line is the performance of the composer. (Source: Kaçar 2005: 223)

Kaçar (2005) classifies the source of differences between the notation and performance in Turkish music under two main titles: ornamentations and non-note based performances which is mentioned as performance styles by Ayangil (2008). Ornamentations detected by Kaçar are as follows: acciaccatura, mordent, trill, grupetto and tremolo. The ornamentations used in Turkish music also includes vibrato, glissando and portamento as mentioned by Ayangil (2008). However, these ornamentations are not represented in the notation as shown in Figure 2.4. As can be seen from the figure, the performer applies grupetto as an ornamentation which is not present in the notation.

Non-note based performances or performance styles can be roughly listed as follows (Kaçar 2005: 224):

- Performance of notes with long durations as notes with small durations.
- Additional notes other than ornamentations.
- Application of double notes.
- Arpeggios.

Figure 2.5 shows application of the second item, performing additional notes other than ornamentations.

Figure 2.5. One bar from a composition of Tanburi Cemil Bey, “Muhayyer Saz Semaisi”. The first line is the notation of the composition and the second line is the performance of Yorgo Bacanos. (Source: Kaçar 2005: 226)

Finally Kaçar concluded that the notation in Turkish music is only a reminder for the performer (2005: 226).

2.6. Manual Transcription of Turkish music: A Case Study

In order to understand the manual transcription procedure and the relation of musicians with original notation, performance and transcription in Turkish music we have applied two qualitative methods of ethnomusicology: interview and participant observation. Interviews are made with two local figures from the Turkish music community of İzmir, Turkey. C. was 20 years old locally well-known professional *tanbur* performer recently educated from the state conservatory of Turkish music. E. was 40 years old *ney* producer, performer and educator without a formal music education. Interviews with C. and E. were made at 11.07.12 and 11.07.06, respectively in İzmir. While C. earns his life by professional performances, E. earns his life mainly by selling *neyler* produced by himself and private *ney* education. E. mainly performs in amateur choruses in İzmir (a city of Turkey). As a result both interviewers represent two facets of Turkish music community in İzmir; *alaylı* (performer without a formal education) and *okullu* (performer graduated from music school). Although community of Turkish music has much more facets, these two categories form the two main division of musical life in Turkey. Therefore it was reasonable to interview and observe these two figures about the transcription procedure of Turkish music.

When I asked E. about his transcription experiences, he replied that he seldomly transcribes Turkish music. One memory of his transcription experience was about helping a friend. His friend found a notation of a piece of *Yansımalar* (music group performing synchretic compositions consist of western harmonization accompanied by guitar to melodies of traditional instruments *tanbur* and *ney*) which did not match with the performance. Therefore, E. transcribes the piece more "accurately". Another experience of E. with transcription is studying the *ney taksimler* of master musicians such as Aka Gündüz Kutbay and Sülayman Yardım from their manually transcribed performances. In order to observe the transcription procedure, I asked him to transcribe a recording of a piece composed by Sadettin Kaynak and performed by *neyzen* (*ney* player) Salih Bilgin with the notation supplied publicly in a web site[9]. He had followed the following steps for transcription without looking at original notation:

i. He listened to each segment (corresponding to one measure usually) repeatedly (3 or 4 times at leats) and try to play the segment by *ney* while listening and then write the corresponding notes to the staff sheet by pencil.
ii. He detected the *usul* as *sofyan* (4/4).
iii. While transcribing each segment he erased and rewritten several note groups.
iv. After transcribing several measures I asked him the *makam* of the piece and he replied that it is probably *segah* makam. Thus he tried to put the accidentals of this *makam* and declared that the tonic is *segah*. However he actually put the accidentals of *hüzzam makam.* There is a 3 Hc difference between the accidentals of these two makamlar are as follows: while *hüzzam makam* has an accidental of b4 for E5*, segah makam* has an accidental of b1 for E5.
v. When I realized that he did not transcribe the ornamentation notes, I asked the reason. He replied that ornamentations are seldom represented on notation and while performing from notation they do not look much at ornamentations, perform the piece as how they knew and heard.
vi. After completing the transcription he listened to the piece one more time to make some corrections.

C. was more experienced in transcriptions and thus our interview takes longer time than E.. He said that instead of transcription of Turkish music, transcription of western music is thought at school but they studied solfege with Turkish music.

[9] The web site neyzen.com provides vaious notations accompanying performances of *neyzen* Salih Bilgin. Including the notation used in this section, all notations and performances used for automatic transcription are taken from this web site. An important detail about the notations and the performance of Salih Bilgin used in this study is that he had selected the most appropriate notations among a number of notation which are slightly different but used for the same piece among musicians.

However he said that this was useful for him when transcription of Turkish music is necessary out of school such as in studio works and composing. I asked him the differences between transcribing a compositional and improvisational form (e.g. *taksim)*. He said that compositions (melodies with *usul*) can be transcribed with success rate of %60-70 depending on the knowledge of instrument and composer. He gave an example that in order to transcribe a composition of Çinuçen Tanrıkorur, a prior knowledge of his style is necessary; an example which reminds the Nettl's example of performing Chopin from notation as mentioned before.

On the other hand according to C. transcription of a *taksim* (improvisation without *usul*) could be more subjective resulting with a success rate of 30-40 %. According to him three transcriptions of the same *taksim* performance could hardly match. He gave an example of this situation based on one of his experiences. One of his friend from school, a class-mate, had asked him to check his transcription of a composition. C. had found many inaccuracies in the transcription due to wrong detection/perception of *usul*. Therefore his friend had failed to discriminate ornamentations from "actual notes" depending on the wrong rhythmic accents.

Another important point about notation and transcription mentioned by C. was the central role of listening to a performance of the piece:

> "In order to perform a piece from notation, listening to the piece is essential. When I transcribed a taksim of İzzet Öke, even I perform the *taksim* from my notation according to the recording I remember. If the transcription was not mine even I could not perform the *taksim*." (interview with C., 2011)

C. also made a distinction between notations as simple and stylistic. While simple notations are transcribed by non-masters of this music which are mostly in use, stylistic notations are transcriptions of master musicians which are rarely found. According to C. simple notations reflects only 20-30% percent of the piece. However stylistic notations transcribed by master musicians such as Çinuçen Tanrıkorur and Alaaddin Yavaşça reflects their style giving the notation more accurate representation of the piece. In fact these transcriptions are rather rewritten notations of compositions instead of transcription of recordings. I asked C. to transcribe the same recording. He asked me what kind of transcription I preferred, transcription of the composition or the style. I preferred transcription of the composition and he followed the following steps:

i. He listen the whole piece repeatedly (3-4 times) and detected the usul of the piece first as *düyek* (8/8).

ii. Then he detected the makam as *hüzzam* and tonic as *segah*.

iii. He started transcription measure by measure while listening each repeatedly without help of any instrument, although his tanbur was with him.

iv. When I realized that he did not transcribe the ornamentation notes, I asked the reason. He said that intentionally he tried to keep the notation simple in order to keep it easily readable.

v. After completing the transcription he listened to the piece one more time but did not need to make any corrections.

Before visual comparison, the differences of the two transcriptions are already clear from the makam and usul detection of transcribers. While E. detected makam and usul as segah and sofyan, C. detected them as hüzzam and düyek in accordance with the original notation. As a result, Figure 2.6 presents a comparison of the original notation and corresponding transcriptions. As can be seen from the figure especially the 1st and 3rd measures of the transcriptions are different from the original notation both in terms of durations of notes and added or deleted notes.

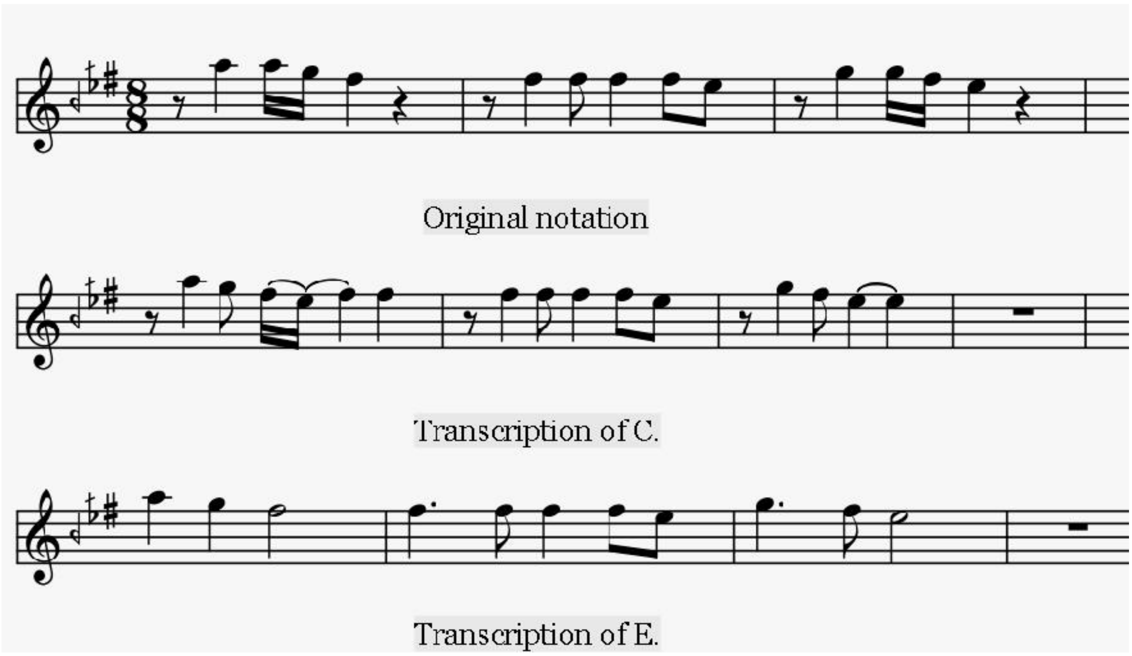


Figure 2.6. First 2- 4 measures of the piece, "*Alma Tenden Canımı*" composed by Sadettin Kaynak shown in the first line, and its corresponding transcriptions by C. in the second line and by E. in the third line. Transcription of the recording of the piece was performed by *neyzen* Salih Bilgin.

2.7. Discussion and Conclusion

Although the notation system has many problems as dicsuued in this chapter, it is a fact that experienced musicians develop an ability to cope with these problems in their relation with current notation system as mentioned by Ayangil (2008). Therefore, the target of an AMT system for Turkish music should be conventional notation of Turkish music, since transcription is by definition is possible only for existing notation system. Of course it is also possible to `invent` a notation system in accordance with practice such as the ethnomusicologists usually follow in their relevant researches. However, the main target of our study is to produce transcriptions that can be read by musicians either for performance or education which leaves a unique way to represent transriptions as conventional staff notation.

Therefore, output of our system will be a prescriptive notation in terms of ethnomusicological definition. However, in contrast to either intentional or unintentional trend of current AMT studies, which aim to obtain original notation from performance as formulated by Klapuri (2004) as reverse engineering, we try to obtain a human readible description of performance as stated by Cemgil et al. (2006).

As a result, our AMT system firstly obtains a detailed transcription of recordings and then eliminates ornementations such as appagiatura, acciaccatura, vibrato and glissandos that are seldomly represented in Turkish staff notation. However, detection of some ornementations and performance styles are not possible considering the state-of-art. There is no method to decompose a recording of a performance as the notes inserted into composition by performance styles and some ornementations on the one side and the notes dictated by notation on the other side. Figure 2.4 and Figure 2.5 present two example of such ornementations and performance styles, respectively.

Finally, even this fact alone demonstrates that it is not possible to obtain original notation from performance which supports the argument about obtaining a readable description of performance. However, since our system follows a direction from a detailed description of performance to a simple notaion of it, our system also supplies a descriptive transcription which represents the details of performance.

CHAPTER 3

AUTOMATIC *MAKAM* RECOGNITION[10]

Due to the divergence of theory and practice presented in the previous chapter, we prefer direct processing of the audio data with data-driven techniques and to utilize very limited guidance from theory. One of the important differences of our approach compared to the related MIR studies is that we do not take any specific tuning system for granted.

As aforementioned, the proper representation of the pitch space is an essential prerequisite for most of the MIR studies for non-western musics. Therefore, our study focuses on the representation of pitch space for Turkish music targeting information retrieval applications. More specifically, this study undertakes the challenging tasks of developing automatic tonic detection and *makam* recognition algorithms for Turkish music.

Makam and tonic pitch of an audio recording are crucial for automatic transcription in Turkish music as discussed in the introduction. It is not possible to find a reference pitch of a recording without the determination of its tonic pitch in Turkish music. In order to find the tonic pitch it is necessary to find the *makam* of the piece in Turkish music.

This chapter firstly presents a comprehensive review of pitch histogram use in MIR studies both for western and non-western music in comparison to Turkish music. Then we discuss more specifically the use of pitch histograms in Turkish music analysis. Following this review part, we present the automatic *makam* recognition and tonic detection based on pitch-frequency histograms.

---

[10] This section is adapted from Gedik, A. C. and Bozkurt, B. (2010). Pitch Frequency Histogram Based Music Information Retrieval for Turkish Music, *Signal Processing*, 90: 1049-1063.

3.1. A Review of Pitch Histogram based MIR Studies

Although there is an important volume of research in MIR literature based on pitch histograms, application of current methods for Turkish music is a challenging task, as briefly explained in the introduction. Nevertheless, we think that any computational study on non-western music should try to define their problem within the general framework of MIR, due to the current well-established literature. Therefore, we review related MIR studies in this section by relating, comparing and contrasting with our data characteristics and applications. Both the data representations and distance measures between data (musical pieces) are discussed in detail since most of the MIR applications (as well as our *makam* recognition application) necessitate use of such distance functions.

We first present our representation of Turkish music pitch space. Musical data is represented by pitch-frequency histograms constructed based on fundamental frequency(f0). f0 data is extracted from monophonic audio recordings. Thus, we apply methods based on pitch histograms. Secondly, necessary methods to process such representation are presented. Thirdly, automatic recognition of Turkish audio recordings by *makam* types (names) is presented.

3.1.1. Pitch Spaces of Western and Turkish Music

A considerable portion of the MIR literature utilizing pitch histograms targets the application of finding the tonality of a given musical piece either as major or minor. In the western MIR literature, tonality of a musical piece is found by processing pitch histograms which simply represent the distribution of pitches performed in a piece as shown in Figure 3.1. In this type of representation, pitch histograms consist of 12 dimensional vectors where each dimension corresponds to one of the 12 *pitch-classes* in western music (notes at higher/lower octaves are folded into a single octave). The pitch histogram of a given musical piece is compared to 24 tonalities, 12 major and 12 minor templates, and the tonality whose template is more similar is found as the tonality of the musical piece.

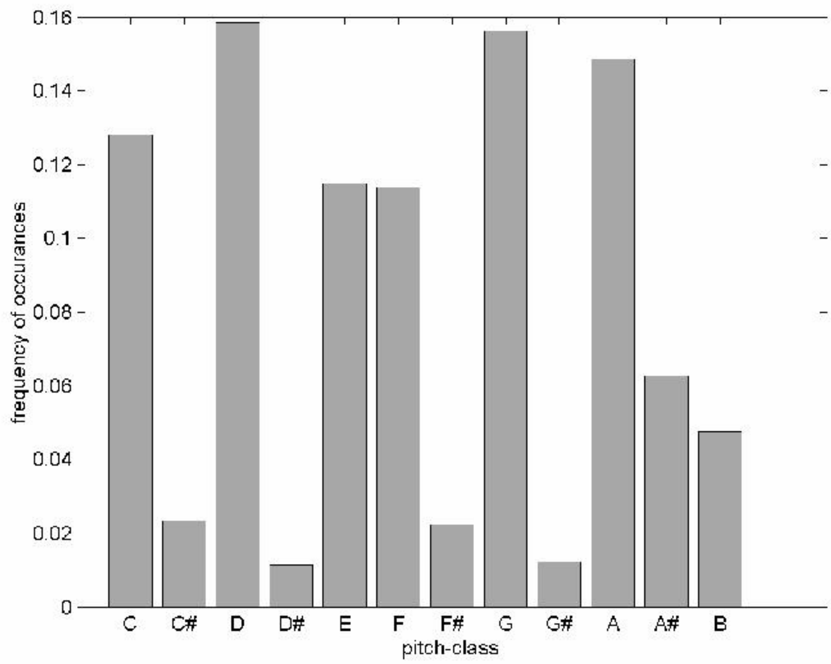


Figure 3.1. Pitch-class histogram of J.S. Bach's C-major Prelude from ...*Wohltemperierte Klavier II* (BWV 870).

As illustrative examples of Turkish music we present two pitch-frequency histograms in Figure 3.2. Two histograms are aligned according to their tonics in order to compare the intervals visually. The tonic frequencies of the two performances are computed as 295 Hz and 404 Hz, hence they are not in a standard pitch. This is an additional difficulty/difference of Turkish music in comparison to western music. Furthermore, another property that cannot be observed on the figure due to plotting of only the main octave, is that it is not possible to represent pitch space of Turkish music within one octave. Depending on the ascending or descending characteristics of the melody of a *makam* type, performance of a pitch can be quite different in different octaves. Therefore it is neither straight forward to define a set of pitch-classes for Turkish music nor represent pitch histograms by 12 pitch-classes as in western music. Furthermore, although the two pieces belong to the same *makam*, the performers prefer close but different pitch intervals for the same pitches.

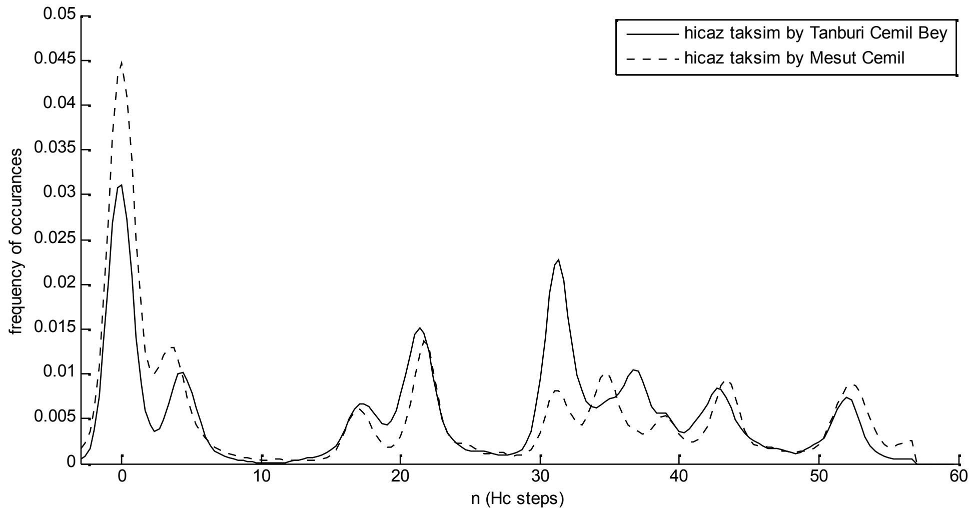


Figure 3.2. Pitch-frequency histograms of *hicaz* performances by Tanburi Cemil Bey and Mesut Cemil. [11]

The next subsection reviews MIR studies developed for western music to investigate whether any method independent from data representation can be applied to Turkish music recordings. In the same subsection, the state-of-art of relevant MIR studies on non-western musics is also reviewed.

3.1.2. Pitch Histogram based Studies for Western MIR

The current methods for tonality finding essentially diverge according to the format (symbolic (MIDI) or audio (wave)) and the content of the data (the number of parts used in musical pieces, either monophonic (single part) or polyphonic (two or more independent parts)). There is an important volume of research based on symbolic data. Audio based studies have a relatively short history (Chuan and Chew 2007). This results from the lack of reliable automatic music transcription methods. Some degree of success in polyphonic transcription has been only achieved under some restrictions (Klapuri 2006) and even the problems of monophonic transcription (especially for some signals like singing) still have not been fully solved (Klapuri 2004). As a result, most of the literature on pitch histograms consists of methods based on symbolic data, and these

[11] Histograms are smoothed by low pass filters to enable a more explicit comparison between performances.

methods also form the basis for the studies on audio data.

It has been already mentioned that tonality of a musical piece is normally found by comparing the pitch histogram of a given musical piece to major and minor tonality histogram templates. Since the representation of musical pieces as pitch-class histograms is rather a simple problem in western music, a vast amount of research is dedicated to investigation of methods for constructing the tonality templates. The tonality templates are again represented as pitch histograms consisting of 12 dimensional vectors, we refer to them as the *pitch-class histogram*. Since there are 12 major and 12 minor tonalities, the templates of other tonalities are found simply by transposing the templates to the relevant keys (Temperley 2001).

The construction of the tonality templates is mainly based on three kinds of models: music theoretical (e.g. Longuet-Higgins, and Steedman 1971), psychological (e.g. Krumhansl, 1990) and data-driven models (e.g. Temperley 2008). These models were also initially developed in the studies based on symbolic data. However, neither psychological nor data-driven models are fully independent from western music theory. In addition, two important approaches of key-finding algorithm based on music theoretical model use neither templates nor key-profiles: the rule-based approach of Lerdahl and Jackendoff (1983) and the geometrical approach of Chew (2002).

Among these models, the psychological model of Krumhansl and Kessler (1990) is the most influential one and presents one of the most frequently applied distance measures in studies based on all three models. Tonality templates are mainly derived from psychological probe-tone experiments based on human ratings, and tonality of a piece is simply found by correlating the pitch-class histogram of the piece with each of the 24 templates. Studies based on symbolic and audio data mostly apply a correlation coefficient to measure the similarity between the pitch-class distribution of a given piece and the templates as defined by Krumhansl (1990):

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 (y-\bar{y})^2}} \qquad (3.1)$$

where x and y refers to the 12 dimensional pitch-class histogram vectors for the musical piece and the template. The correlation coefficients for a musical piece are computed using Equation 3.1 with different templates (y) and the template which gives the highest coefficient is found as the tonality of the piece.

The same method is also applied in data-driven models (e.g. Temperley 2008) by simply correlating the pitch-class histogram of a given musical piece with major and minor templates derived from relevant musical databases. Even the data-driven models reflect the western music theory by the representation of musical data and templates as 12 dimensional vectors (pitch-classes).

Although studies on audio data (e.g. Zhu and Kankanhalli 2006) diverge from the ones on symbolic data by the additional signal processing steps, these studies also try to obtain a similar representation of the templates where pitch histograms are again represented by 12 dimensional pitch-class vectors. Due to the lack of a reliable automatic transcription, such studies process the spectrum of the audio data without f0 estimation to achieve tonality finding. In these studies, the signal is first pre-processed to eliminate the non-audible and irrelevant frequencies by applying single-band or multi-band frequency filters. Then, Discrete Fourier Transform (DFT) or constant Q-transform (CQT) are applied and the data in the frequency domain is mapped to pitch-class histograms (e.g. Zhu and Kankanhalli 2006; Gomez 2006). However, this approach is problematic due to the complexity of reliably separating harmonic components both for polyphonic and monophonic music which are naturally not present in symbolic data. Another problem is the determination of tuning frequency (which determines the band limits and the mapping function) in order to obtain reliable pitch-class distributions from the data in the frequency domain. Most of the studies take the standard pitch of A4=440 Hz as a ground truth for western music (e.g. Chuan and Chew 2005; Purwins et al. 2000). On the other hand, few studies estimate first a tuning frequency, considering the fact that recordings of various bands and musicians need not to be tuned exactly to 440 Hz. However, even in these studies, 440 Hz is taken as a ground truth in another fashion (Ong et al. 2006; Zhu and Kankanhalli 2006). They calculate the deviation of the tuning frequency of audio data from 440 Hz, and then take into account this deviation in constructing frequency histograms. When Turkish music is considered, no standard tuning exists (but only possible "*ahenk*"s for rather formal recordings). This is another important obstacle for applying western music MIR methods to our problem.

Although mostly the correlation coefficient presented in Equation 3.1 is used to measure the similarity between pitch-class distribution of a given piece and templates, a number of recent studies apply various machine learning methods for tonality detection such as Gomez and Herrera (2004). Chuan and Chew (2007), and, Lee and Slaney (2008) do not use templates, but their approach is based on audio data synthesized from symbolic data (MIDI). Lui et al. (2008) also do not use templates but for the first time apply unsupervised learning. Since these approaches present the same difficulties when applying them to Turkish music, they will not be reviewed here.

3.1.3. Pitch Histogram based Studies for Non-Western MIR

Although most of the current MIR studies focus on western music, a number of studies considering non-western and folk musics also exist. The most common feature of these studies is the use of audio recordings instead of symbolic data. However, most of the research is based on processing of the f0 variation in time and does not utilize pitch histograms, which is shown to be a valuable tool in analysis of large databases. There is a relatively important volume of research on the pitch space analysis of Indian music which does not utilize pitch histograms but directly the f0 variation curves in time (Krishnaswamy 2003a; 2003b; 2003c; 2004). This is also the case for the two studies on African music (Marandola 2003) and Javanese music (Carterette and Kendall 1999). There are also two MIR applications for non-western music without using pitch histograms: an automatic transcription of Aboriginal music (Nesbit et al. 2004) and the pattern recognition methods applied on South Indian classical music (Sinith and Rajeev 2007). Here, we will only review studies based on pitch histograms and refer the reader to Tzenatakis et al. (2007) for a comprehensive review of computational studies on non-western and folk musics.

The literature of non-western music studies utilizing pitch histograms for pitch space analysis is much more limited. The studies of Moelants et al. (2006; 2007) apply pitch histograms to analyze the pitch space of African music. Instead of pitch-class histograms as in western music, “pitch-frequency histograms” are preferred, and thus such continuous pitch space representation enables them to study the characteristic of the tuning system of African music. They introduce and discuss important problems related to African music based on analysis of a musical example but do no present any MIR application. Akkoc (2002) analyses pitch space characteristics of Turkish music based on the performances of two outstanding Turkish musicians again using limited data and without any MIR application. Bozkurt (2008) presented for the first time the necessary tools and methods for the pitch space analysis of Turkish music when applied to large music databases.

There is a number of MIR studies which utilize pitch histograms for aims other than analyzing the pitch space. One example is Norowi et al. (2005) who use pitch histograms as one of the features in automatic genre classification of traditional Malay music beside timbre and rhythm related features. In this study, the pitch histogram feature is automatically extracted using the software, MARSYAS, which computes pitch-class histograms as in western music. Certain points in this study are confusing and difficult to interpret, which hinders its use in our application: among other things, it is not clear how the lack of a standard pitch is solved, the effect of pitch features in classification is not evaluated, and the success rate of the classifier is not clear since only the accuracy parameter is presented.

Two MIR studies on the classification of Indian classical music by *raga* types (Chordia and Rae 2007; Chordia et al. 2008) are fairly similar to our study on classification of Turkish music by *makam* types. However, in these studies the just-intonation tuning system is used as the basis, and surprisingly 12 pitch-classes as in western music are defined for the histograms, although they mention that Indian music includes microtonal variations in contrast to western music. Chordia and Rae (2007) used pitch-class dyad histograms also as a feature which refers to the distribution of pitch transitions besides pitch-class histograms with the same basis. We find it problematic to use a specific tuning system for pitch space dimension reduction of non-western musics unless the existence of a theory well-conforming to practice is shown to exist. In addition, a database of 20 hours audio recordings manually labeled in terms of tonics is used in this study. This is a clear example showing the need for automatic tonic detection algorithms for MIR. Again the high success rates obtained for classification is subject to question for these studies due to the use of optimistic parameters for evaluation, such as accuracy. Another study (Chordia et al. 2008) presents a more detailed classification study of North Indian classical music. Three kinds of classifications are applied: classification by artist, by instrument, by *raga* and *thaat*. Each musical piece is again represented as pitch-class histograms for classification by the *raga* types. On the other hand, this time only the similarity matrix is mentioned for the *raga* classifier and the method of classification is not explained any further. Again, it is not clear how pitch histograms are represented in the classification process. The success rates for classification by raga types applied on 897 audio recordings were found to be considerably low in comparison to the previous study on *raga* classification (Chordia and Rae 2007). Finally, an important drawback of this study is again the manual adjustment of the tonic of the pieces. Again, all these problematic points hinder the application of these technologies in other non-western MIR studies: some important points related to the implementation or representations are not clear, the results are not reliable or considerable amount of manual work is needed. We believe that this is mainly due to the relatively short history of non-western MIR.

The most comprehensive study on non-western music is presented by Gomez and Herrera (2008). A new feature, harmonic pitch class profile (HPCP) proposed by Gomez (2006) which is inspired by pitch-class histograms, is applied to classify a very large music corpora of western and non-western music. Besides HPCP, other features such as tuning frequency, equal tempered deviation, non-tempered energy ratio and diatonic strength, which are closely related with tonal description of music, are used to discriminate non-western musics from western musics or vice versa. While 500 audio recordings are used to represent non-western music including musics of Africa, Java, Arabic, Japan, China, India and Central Asia, 1000 audio recordings are used to represent western music including classical, jazz, pop, rock, hiphop, country music etc.

From our point of view, an interesting point of this study is the use of pitch histograms (HPCP) without mapping the pitches into a 12 dimensional pitch-space as in western music. Instead, pitches are represented in a 120 dimensional pitch-space which thus enables to represent pitch-spaces of various non-western musics. Considering the features used, the study mainly discriminates between non-western musics from western music by computing their deviation from equal-tempered tuning system, in other words their "deviation" from western music. As a result, two kinds of classifiers, decision trees and SVM, are evaluated and success rates higher then 80% are obtained in terms of F-measure. However, the study also bears serious drawbacks as explicitly demonstrated by Lartillot et al. (2008). One of the critiques refers to the assumption of octave equivalence for non-western musics. The other criticism is related to the assumption of tempered scale for non-western musics as implemented in some features such as tuning frequency, non-tempered energy ratio, the diatonic strength etc. Finally, it is also not explained how the problem of tuning frequency is solved for non-western music collections.

Another group of study apply self-organizing maps (SOMs) based on pitch histograms to understand the non-western and folk musics by visualization. Toiviainen and Eurola (2006) apply SOM to visualize 2240 Chinese, 2323 Hungarian, 6236 German and 8613 Finnish folk melodies. Chordia and Rae (2007) also apply SOM to model tonality in North Indian Classical Music.

As a result of this review, we conclude that non-western music research is very much influenced by western music research in terms of pitch space representations and MIR methodologies. This is problematic because the properties common to many non-western musics, such as the variability in frequencies of pitches, non-standard tuning, extended octave characteristics, practice of the concept of modal versus tonal, differ highly in comparison to western music. The literature of fully automatic MIR algorithms for non-western music, taking into consideration its own pitch space characteristics without direct projection to western music, is almost non-present. The use of methodologies developed for western music is in general acceptable, but data space mappings are most of the time very problematic.

3.2. Pitch Histogram based Studies for Turkish MIR

In the literature about Turkish music, pitch-frequency histograms are successfully used for tuning research by manually labeling peaks on histograms to detect pitch frequencies (Akkoc 2006; Zeren 2003; Karaosmanoğlu and Akkoc 2003; Karaosmanoğlu 2004). As discussed in the previous sections, it is clear that representing Turkish music using a 12 dimensional pitch-class space is not appropriate. Aiming at developing fully automatic MIR algorithms, we use high resolution pitch frequency histograms, without a standard pitch or tuning system (tempered or non-tempered) taken for granted.

Following the f0 estimation, a pitch frequency histogram, $Hf_0[n]$, is computed as a mapping that corresponds to the number of f0 values that fall into various disjoint categories.

$$
\begin{aligned}
&Hf_0[n] = \sum_{k=1}^{K} m_k \\
&m_k = 1, f_n \leq f_0[k] < f_{n+1} \\
&m_k = 0, otherwise
\end{aligned}
\tag{3.2}
$$

where (f_n, f_{n+1}) are boundary values defining the f0 range for the n[th] bin.

One of the critical choices made in histogram computation is the decision of bin-width, W_b, where automatic methods are concerned. It is common practice to use logarithmic partitioning of the f0 space in musical f0 analysis which leads to uniform sampling of the log-f0 space. Given the number of bins, N, and the f0 range (f_{0max} and f_{0min}) bin-width, W_b, and the edges of the histogram, f_n, can be simply obtained by:

$$
\begin{aligned}
&W_b = \frac{\log_2(f_{0\max}) - \log_2(f_{0\min})}{N} \\
&f_n = 2^{f_{0\min} + (n-1)W_b}
\end{aligned}
\tag{3.3}
$$

For musical f0 analysis, various logarithmic units like *cents* and *commas* are used. Although the *cent* (obtained by the division of an octave into 1200 logarithmically equal partitions) is the most frequently used unit in western music analysis, it is common practice to use the *Holderian comma (Hc)* (obtained by the division of an octave into 53 logarithmically equal partitions) as the smallest intervallic unit in Turkish music theoretical parlance. To facilitate comparisons between our results and Turkish music theory, we also use the Holderian comma unit in partitioning the f0 space (as a result in our figures and tables). After empirical tests with various grid sizes, 1/3 Holderian comma (Hc) resolution is obtained by Bozkurt (2008). This resolution optimizes smoothness and precision of pitch histograms for various applications. Moreover, this resolution is the highest master tuning scheme we could find from which a subset tuning is derived for Turkish music, as specified by Yarman (2008).

In the next sections, we present the MIR methods we have developed for Turkish music based on the pitch-histogram representation.

3.2.1. Automatic Tonic Detection

In the analysis of large databases of Turkish music, the most problematic part is correlating results from multiple files. Due to diapason differences between recordings (i.e. non-standard pitches), lining up the analyzed data from various files is impossible without a reference point. Fortunately, the tonic of each *makam* serves as a viable reference point.

Theoretically and as a very common practice, a recording in a specific *makam* always ends at the tonic as the last note (Akdoğu 1989). However tracking the last note reliably is difficult especially in old recordings where the energy of background noise is comparatively high.

Bozkurt (2008) presented a robust tonic detection algorithm (shown in Figure 3.3) based on aligning the pitch histogram of a given recording to a *makam* pitch histogram template. The algorithm assumes the *makam* of the recording is known (either from the tags or track names since it is common practice to name tracks with the *makam* name as "*Hicaz taksim*").

The *makam* pitch histogram templates are constructed (and also the tonics are re-estimated for the collection of recordings) in an iterative manner: the template is initiated as a Gaussian mixture from theoretical intervals and updated recursively as recordings are synchronized. Similar to the pitch histogram computation, the Gaussian mixtures are constructed in the log-frequency domain. The widths of Gaussians are chosen to be the same in the log-frequency domain as presented in Figure 3.2 of (Bozkurt 2008). Since in the algorithm a theoretical template is matched with a real histogram, the best choice of width for optimizing the matching is to use the width values close to the ones observed in the real data histograms.

We have observed on many samples that the widths of most of the peaks in real histograms appear to be in the 1-4Hc range. As expected, smaller widths are observed on fretted instrument samples where as larger widths are observed for unfretted instruments. Several informal tests have been performed to study the effect of the width choice for the tonic detection algorithm. We have observed that for the widths in the 1.5-3.5Hc range, the algorithm converges to the same results due to the iterative approach used. Since it is an iterative process and the theoretical template is only used for initialization, the choice of the theoretical system is not very critical, nor the width of the Gaussian functions. Given any of the existing theories and a width value in the 1.5-3.5Hc range, the system quickly converges to the same point. It only serves a means for aligning histograms with respect to each other and is not used for dimension reduction. One alternative to using theoretical information is to manually choose one of the recordings to be representative as the initial template. Since it is an iterative process and the theoretical template is only used for initialization, the choice of the theoretical system is not very critical. Given any of the existing theories, the system quickly converges to the same point. It only serves a means for aligning histograms and is not used for dimension reduction.

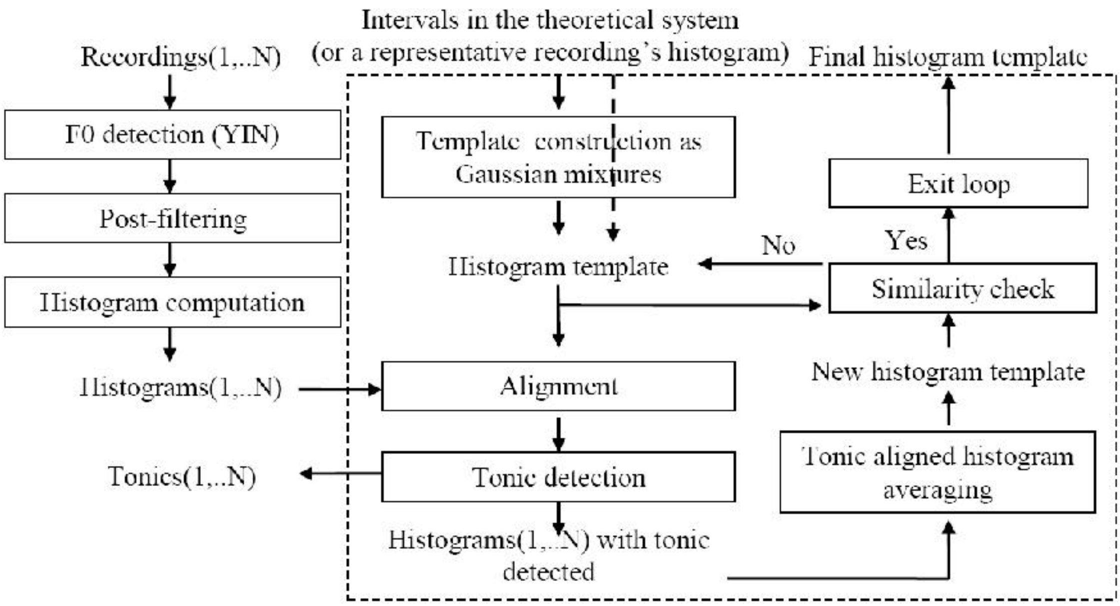
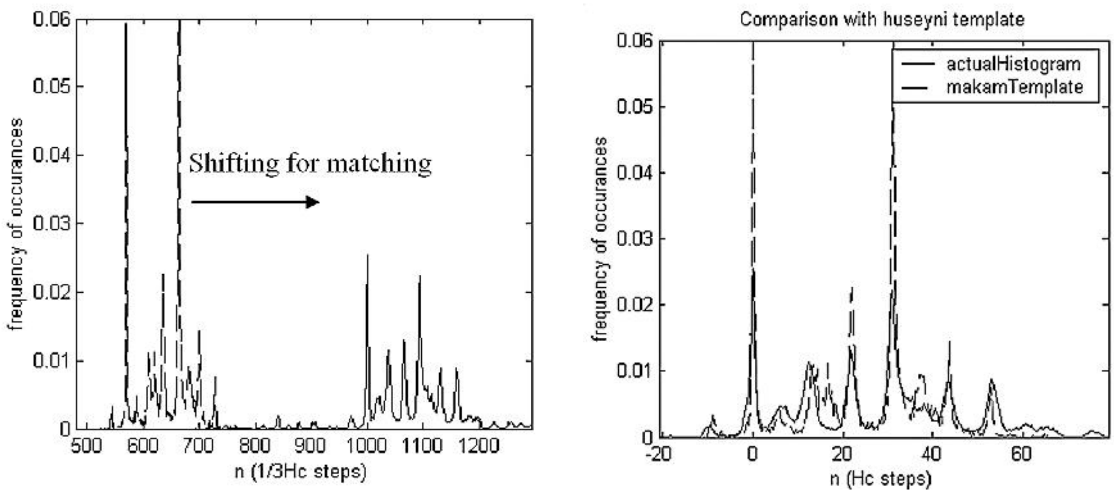


Figure 3.3. Tonic detection and histogram template construction algorithm (box indicated with dashed lines) and the overall analysis process. All recordings should be in a given *makam* which also specifies the intervals in the theoretical system. (Source: Bozkurt 2008)



(a) (b)

Figure 3.4. Tonic detection via histogram matching. a) template histogram is shifted and the distance/correlation is computed at each step, b) matching histograms at the shift value providing the smallest distance (normalized for viewing, tonic peak is labeled as the 0Hc point).

The presented algorithm is used to construct *makam* pitch histogram templates used further both in tonic detection of other recordings and for the automatic classifier explained in the next section.

Once the template of the *makam* is available, automatic tonic detection of a given recording is achieved by:

- Sliding the template over the pitch histogram of the recording in 1/3Hc steps (as shown in Figure 3.4a)
- Computing the shift amount that gives the maximum correlation or the minimum distance using one of the measures listed below
- Assigning the peak that matches the tonic peak of the template as the tonic of the recording (as shown in Figure 3.4b by indicating the tonic with 0Hc) and computing the tonic from the shift value and the template's tonic location.

These steps are represented as two blocks (Synchronization, Tonic Detection) in Figure 3.3.

Bozkurt (2008) found the best matching point between histograms by finding the maximum of the cross-correlation function, *c[n]*, computed using the following equation:

$$c[n] = \frac{1}{K}\sum_{k=0}^{K-1} h_r[k]h_t[n+k] \quad (3.4)$$

where $h_r[n]$ is the recording's pitch histogram and $h_t[n]$ is the corresponding *makam*'s pitch histogram template.

3.2.2. Automatic *Makam* Recognition

In pattern recognition literature, template matching method is a simple and robust approach when adequately applied (Cha and Srihari 2002; Brunelli and Poggio 1993; Tanaka et al. 2000; Li and Hui 2000; Santini and Jain 1999). Temperley (2001) also considers the method of tonality finding in literature on western music as template matching. We also apply template matching for finding *makam* of a given Turkish music recording. In addition, as mentioned before, a data-driven model is chosen for the construction of templates.

Similar to pitch histogram based classification studies, we also use a template matching approach to *makam* recognition using pitch frequency histograms: each recording's histogram is compared to each histogram template of the *makam* type and the *makam* type whose template is more similar is found as the *makam* type of the recording. In contrast, there is no assumption of a standard pitch (diaposon) nor a mapping to a low dimensional class space. One of the histograms is shifted (transposed in musical terms) in 1/3Hc (1/159 octaves) steps until the best matching point is found in a similar fashion to the tonic finding method described in section 3.2. The algorithm is simple and effective, and the main problem is the construction of *makam* templates.

In our design and tests, we have used nine *makam* types which represent 50% of the current Turkish music repertoire (Oztuna 2006). The list can be extended as new templates are included which can be computed in a fully automatic manner using the algorithm described in (Bozkurt 2008).

Tests: Our database consists of 172 audio recordings from nine *makam* types. The *makam* types and the number of recordings from each *makam* type are as follows: 20- *hicaz*, 19- *rast*, 21- *saba*, 20- *segah*, 16- *kürdili hicazkar*, 14- *hüzzam*, 18- *nihavend*, 20- *hüseyni* and 24- u*şşak*.

The uneven distribution of samples for each *makam* is due to the current database of recordings we have collected so far. One-hundretd-seventy-two recordings of historically most prominent musicians as well as the more actual ones were selected for classification. Recordings were not partitioned and analysed as a whole. These were monophonic (non-heterephonic) recordings of the following instruments: *ney*, *tanbur*, *kemençe*, violin, clarinet and cello. Some of the performers in the recordings were: Tanburi Cemil Bey (1873-1916), Mesut Cemil (1902-1963), Niyazi Sayın (b.1927), Necdet Yaşar (b.1930), Sadrettin Özçimi (b.1955).[12]

Due to the limited number of recordings, we apply *leave one-out cross validation* in both the construction of the templates and the evaluation of the *makam* recognition system. Therefore, when a recording is subject to comparison with *makam* templates, it does not contribute to the construction of the template of the *makam* type the recording belongs to, and the comparison is made on the basis of unknown tuning frequency of the recording. The template for each *makam* type is simply computed by averaging the pitch-frequency histograms of audio recordings from the same *makam* type after aligning all histograms with respect to their tonics ('Tonic synchronized histogram averaging' block in Figure 3.4). In other words, every time a recording is compared with the templates, the templates are reconstructed from the rest of the recordings.

Firstly, each pitch-frequency histogram is obtained and normalized to unity sum as follows:

$$Hf_0N = \frac{Hf_0}{\sum Hf_0} \tag{3.9}$$

, where Hf_0 denotes the pitch-frequency histogram and Hf_0N denotes the normalized pitch-frequency histogram. The templates for each *makam* type are obtained by summing the tonic aligned histograms and normalization:

[12] Detailed information about the recordings and Turkish music can be found at project web page: http://likya.iyte.edu.tr/eee/labs/audio/Main.html

$$T_k = \sum_{i=1}^{N} Hf_0 N_k(i) \qquad TN_k = \frac{T_k}{\sum T_k} \tag{3.10}$$

where $Hf_0N_k(i)$ denotes the normalized pitch-frequency histogram of the *i* th recording from *makam* type *k*, *N* refers to the number of recordings from *makam* type *k*, T_k refers to template for the *makam* type *k* and TN_k refers to the normalized template. As a result, templates for each *makam* type are obtained: two templates for two *makam* types are shown in Figure 3.5 as an example.

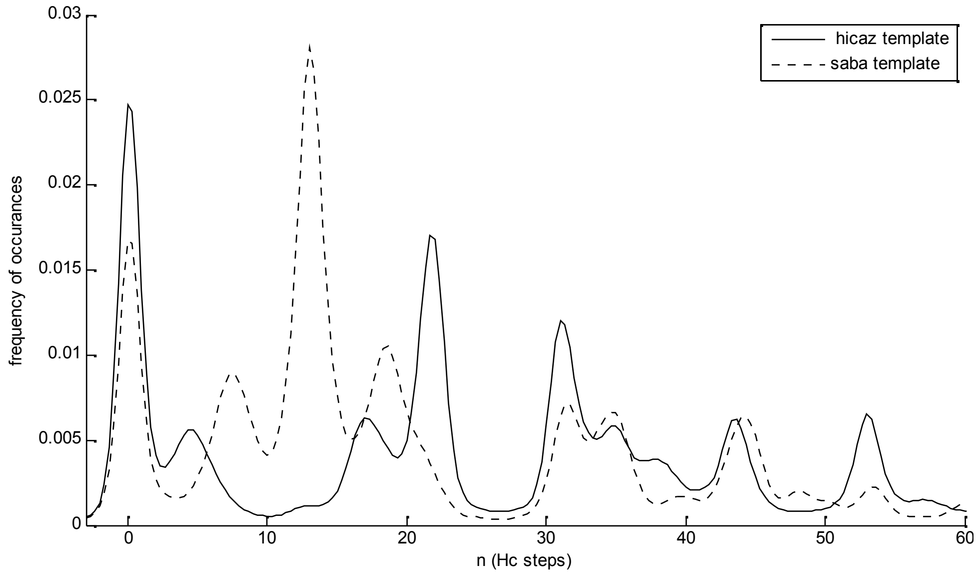


Figure 3.5. Pitch-frequency histogram templates for the two types of melodies: *hicaz makam* and *saba makam*.

Finally, when the data-driven model is finished, the similarity between templates and a recording can be measured. City-Block (L1 norm) distance measure is used in the *makam* recognition system. Given a recording's histogram and the templates, the histogram is shifted in 1/3Hc steps and the City Block distances to the templates are computed at each step. Finally, the smallest distance is obtained and the corresponding template indicates the *makam* type of the recording.

Since both makam recognition and tonic detection base on matching a histogram with a template, these two steps are indeed performed by a single histogram matching operation. Interestingly, for many cases where makam detection fails, the tonic detection can still be correctly done. This is due to the fact that makams confused often share almost the same scale structure and matching results with the same tonic.

The *makam* recognition system described above is evaluated by computing the measures and parameters presented below:

$$recall = \frac{TP}{TP+FN} \quad precision = \frac{TP}{TP+FP} \quad F-measure = \frac{2 \cdot recall \cdot precision}{(recall + precision)} \tag{3.11}$$

TP: True positive, *TN*: True negative, *FP*: False positive, *FN*: False negative

The success rates obtained are presented in Table 3.1.

Table 3.1. The evaluation results of the *makam* recognition system.

***Makam* Type**	**TP**	**TN**	**FP**	**FN**	**R**	**P**	**F-measure**
hicaz	14	150	2	6	70	88	78
rast	14	151	2	5	73	88	79
segah	17	149	3	3	85	85	85
kürdili h.	10	145	11	6	63	48	55
huzzam	10	152	6	4	71	63	67
nihavend	14	143	11	4	78	56	65
hüseyni	10	146	6	10	50	63	56
uşşak	15	138	10	9	63	60	62
saba	16	150	1	5	76	94	84
mean	**13**	**147**	**6**	**6**	**68**	**68**	**68**

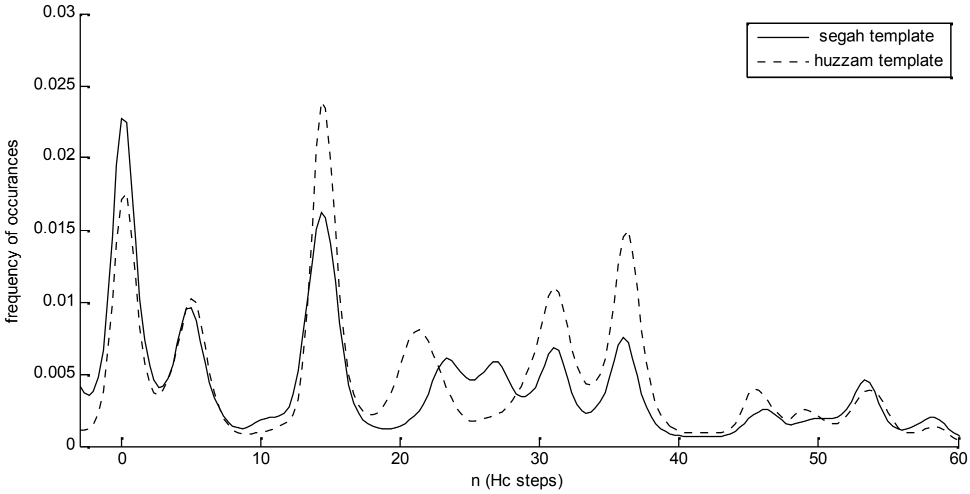
Table 3.1 shows that while the *makam* recognition system is successful for the *makam* types *hicaz*, *rast*, *segah* and *saba*, it is not very successful for the *makam* types *kürdili hicazkar*, *hüzzam*, *nihavend*, *hüseyni* and u*şşak*. Table 3.2 presents the confusion matrix of the *makam* recognition system (unsuccessfully retrieved *makam* types indicated as bold).

The highest confusion between the *makam* types are as follows: *segah* and *hüzzam* on the one side and *kürdili hicazkar*, *uşşak*, *hüseyni* and *nihavend* on the other side. Observing the templates of these two groups of *makam* types indicates the reason of this confusion. While the *segah* and *hüzzam* have similar pitch-frequency histograms on the one side, as shown in Figure 3.6a, *kürdili hicazkar*, *uşşak*, *hüseyni* and *nihavend* have similar pitch-frequency histograms on the other side, as depicted in Figure 3.6b. On the other hand, two *makam* types, *hicaz* and *saba*, are not confused due to the dissimilar pitch-frequency histograms as shown in Figure 3.5.

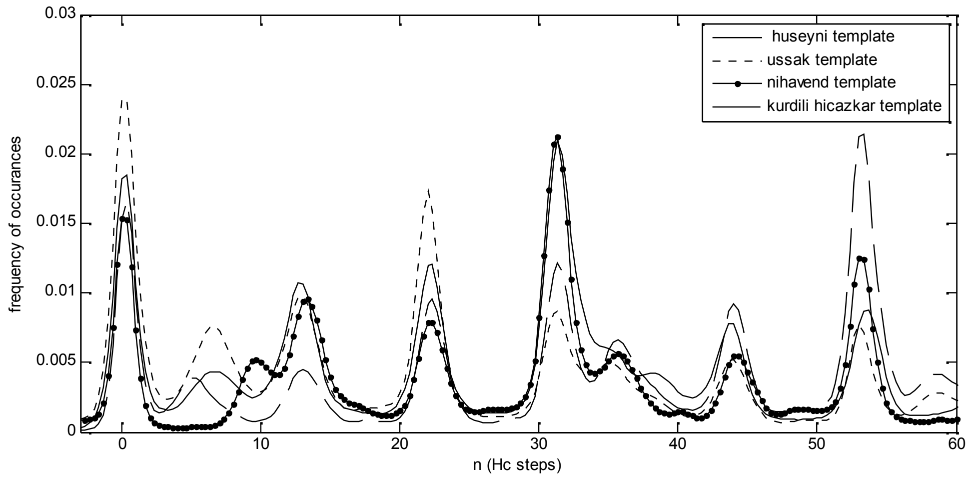
Table 3.2. Confusion matrix of the *makam* recognition system.

	hicaz	*rast*	***segah***	***kürdili hicazkar***	***hüzzam***	***nihavend***	***hüseyni***	***uşşak***	*saba*
hicaz	-	-	-	1	**-**	**2**	**2**	1	-
rast	-	-	-	-	**3**	1	-	1	-
segah	**-**	**-**	**-**	**-**	**3**	**-**	**-**	**-**	-
kürdili hicazkar	1	-	-	**-**	**-**	**2**	**-**	**3**	-
hüzzam	1	-	**-**	**1**	**-**	**-**	**-**	**-**	-
nihavend	-	1	**1**	**1**	**-**	**-**	**2**	**-**	-
hüseyni	-	-	-	**3**	**-**	**3**	**-**	**4**	-
uşşak	-	-	-	**4**	**-**	**2**	**2**	**-**	1
saba	-	1	1	1	-	1	2	1	-

The most confused *makamlar* are also evaluated from the view point of music theoretical knowledge, especially the theory founded by Arel (Gedik and Bozkurt 2008). Pitch interval values of the confused *makam* couples, *segah* - *hüzzam* and *uşşak* - *hüseyni* are very similar to the Arel theory. The theoretical pitch interval values between the *makamlar nihavend* and *kürdili hicazkar* have also certain similarities to the *makamlar*, *uşşak* and *hüseyni*.



(a)



(b)

Figure 3.6. Pitch-frequency histogram templates for the two groups of *makam* (a) *segah* and *hüzzam* (b) *kürdili hicazkar*, u*şşak*, *hüseyni* and *nihavend*.

3.3. Discussions, Conclusions and Future Work

In this chapter, the use of a high dimensional pitch-frequency histogram representation without pre-assumptions about the tuning, tonality, pitch-classes, or a specific music theory, for two MIR applications are presented: automatic tonic detection and *makam* recognition for Turkish music. In the introduction and review sections, we first discussed why such a representation is necessary by discussing similar methods in literature. We have shown that very high quality tonic detection and a fairly good *makam* recognition could be achieved using this type of representation and the simple approach of "shift and compare".

"Shift and compare" processing of pitch-frequency histograms mainly correspond to transpose-invariant scale comparison since peaks of the histograms correspond to notes in the scale. The results of the *makam* recognition system show that the scale structure is very discriminative for some of the makams such as *segah* and *saba* (F-measures: 85, 84). For other *makamlar,* such as *kürdili hicazkar* and *hüseyni*, the success rate is relatively lower (F-measures: 55, 56) though still much higher than chance (100/9 for 9 classes). The payback of using histograms instead of time-varying f0 data for analysis is the loss of the temporal dimension and therefore, the musical context of executed intervals. Referring back to Turkish music theory, we see that *makam* descriptions include ascending-descending characteristics, possible modulations and typical motives. In our future work, we will try to add new features derived from f0 curves based on this information, again using a data-driven approach.

CHAPTER 4

EVALUATION OF THE SCALE THEORY OF TURKISH MUSIC[13]

This chapter evaluates the makam scale theory of Arel. Although Arel theory gives place to other central concepts of the definition of Turkish music such as *seyir*, melodic organization and *usul*, rhythmic organization, the most disputable and discriminative dimension of the theory is the *makam* scales. In other words, the discussion of Arel theory corresponds to the discussion of *makam* scales in theoretical studies on Turkish music. Therefore, we prefer to use the term, "Arel theory" throughout the study, instead of "*makam* scale theory of Arel".

The most straightforward approach for the evaluation of the theory and its suitability to MIR-type methods is to compare the defined pitch-classes with the pitch values obtained from practice. A comprehensive computational research based on such a comparison is presented by Bozkurt et al. (2009) on five theoretical systems, including Arel theory. Although this study provides empirical results over a significantly large amount of data for the first time, the suitability of a theory for MIR applications should be evaluated within the context of MIR. As a result, our study evaluates Arel theory within the context of MIR studies.

We have presented comprehensively the obstacles against applying current MIR and tonality finding methods for Turkish music without any contribution from theory by developing a data-driven model in Chapter 3. While our study shares the conceptual framework of this study, the computational framework used in this chapter is based on a music theoretical model. Since Turkish music is based on a modal system, our study rather corresponds to modality finding, analogous to tonality finding studies on western music. In this sense, finding the *makam* of a given piece refers to finding the modality of a given piece. In this chapter, modality (*makam*) templates are constructed based on Arel theory and a given piece is compared with these modality templates. Consequently,

[13] This section is adapted from Gedik, A. C. and Bozkurt, B. (2009). Evaluation of the Makam Scale Theory of Arel for Music Information Retrieval on Traditional Turkish Art Music, *Journal of New Music Research*, 38(2): 103-116.

the modality whose template has the highest similarity is identified as the modality of the piece.

For each *makam*, a scale usually within an octave[14] and its pitch intervals are defined in Arel theory. The pitch interval types and their values in Hc defined in the Arel theory are as follows: *bakiyye*-4, *küçük müneccep*-5, *büyük müneccep*-8, *tanini*-9, *artık ikili*-12 or 13. Based on these pitch interval values and the definition of makamlar in Arel theory, we have derived a list of other pitch intervals[15] with respect to the tonic (*karar*) for the nine *makamlar* as shown in Table 4.1.

Table 4.1. Makam scale intervals of nine makamlar in Arel theory. Intervals for each makam are given in Hc with respect to tonic.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
hicaz	5	17	22	31	35	39	44	53
rast	9	17	22	31	40	48	-	53
segah	5	14	22	31	36	45	49	53
kurdili hicazkar	4	13	22	31	35	44	-	53
huzzam	5	14	19	31	36	49	-	53
nihavend	9	13	22	31	35	44	-	53
hüseyni	8	13	22	31	39	44	-	53
uşşak	8	13	22	31	35	44	-	53
saba	8	13	18	31	35	44	49	-

Finally, the automatic *makam* recognition method and the data set summarized in Subsection 1.3 are used for the evaluation.

4.1. Automatic Classification according to the *Makam* Scales[16]

As mentioned in the introduction, we consider modality finding as the aim of our study, where each *makam* corresponds to a modality, analogous to tonality finding studies on western music. However, due to the difference of pitch spaces between western music and Turkish music, both the modality templates and recordings are

[14] Only the *makam saba* among the *makamlar* used is defined in Arel theory as exceeding the range of an octave. Since we consider all *makamlar* within an octave, intervals higher than 53Hc (for example 61 Hc) of *saba* scale are omitted.

[15] 7th interval for *hicaz*, *segah* and *saba makam* scales are defined by Arel with respect to the *seyir* features of these *makamlar*. According to Arel, these *makam* scales either use 6th interval or 7th interval depending on the melodic direction (ascending or descending).

[16] All codes for automatic classification are written in MatLab 6.1

represented as pitch-frequency histograms instead of pitch-class distributions used for western music. In a similar fashion to current MIR studies, modality templates are constructed first, based on Arel theory: the pitch-frequency histograms derived from theoretical *makam* scales are used as templates. Then a given piece, represented again as pitch-frequency histogram, is compared to the modality templates.

4.1.1. Representation of Practice

The method proposed by Bozkurt (2008) for the analysis of pitch frequencies in Turkish music was used to pre-process and then to represent the recordings as pitch-frequency histograms. In this method, each recording in audio format (wav file) is analyzed by YIN (de Cheveigne A. and Kawahara, H. 2002) and the estimated fundamental frequency values are post-processed with filters. These filters are especially designed for Turkish music, based on its acoustic characteristics (Bozkurt 2008). Then the automatic tonic detection algorithm presented by Bozkurt (2008) is applied, and the results are checked and corrected manually. Pitch frequencies are converted into pitch interval values with respect to the tonic in Hc and the distributions are computed. These distributions also represent the scale structure of a *makam* performed in a recording. As a result, each recording is represented as pitch-frequency histogram (Figure 4.1).

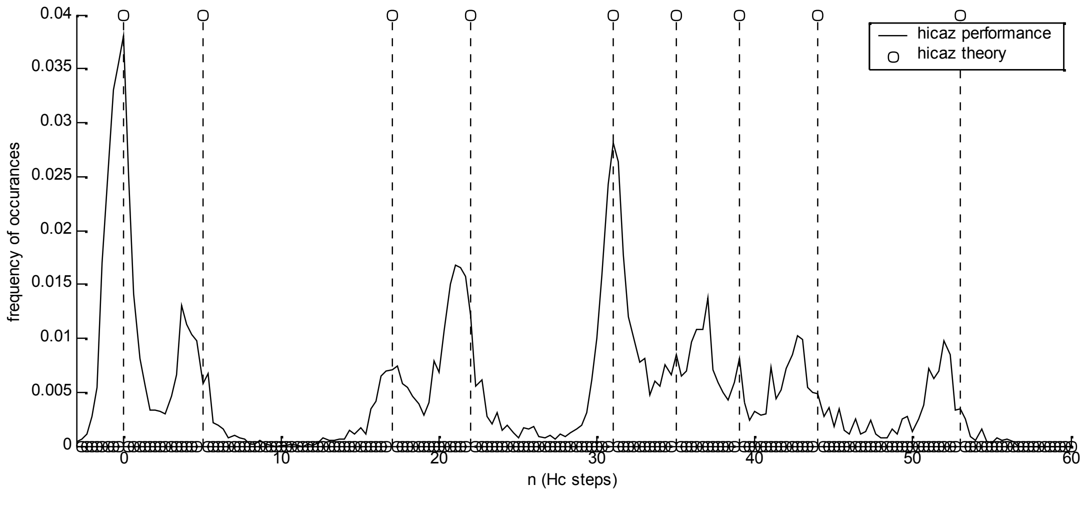


Figure 4.1. Pitch interval histogram of a *hicaz taksim* by Tanburi Cemil Bey and *hicaz* scale defined in Arel theory.

4.1.2. Representation of Theory

Although Arel defines fixed pitch intervals for each *makam* scale, none of the 172 recordings demonstrates such characteristics. All the pitch-frequency histograms we have computed in this study from recordings showed rather flexible pitch frequencies. Consequently, we have transformed the theoretical *makam* scales by converging it to the practice, and represented each fixed pitch interval value of a *makam* scale defined in theory by Gaussian distributions. The mean of each Gaussian distribution was set at the fixed pitch interval values defined in the theory for each *makam*, and their standard deviations were selected as 2 Hc, heuristically. Finally, each theoretical *makam* scale was represented as the sum of these Gaussian distributions as shown in Figure 1.5 in Chapter 1. Each of the Gaussian distributions is calculated by the equation shown below:

$$g(x,\mu)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{4.1}$$

g : Gaussion distribution
μ (mean): assigned as constant which corresponds to fixed pitch values (Hc) defined in theory for each pitch of a *makam* scale.
σ (standart deviation): constant value used as 2

The *makam* scales are then represented in terms of these Gaussian distributions as shown below:

$$s_m=\sum_{k=1}^{n} g(x,\mu_k) \tag{4.2}$$

s_m : template of *makam* m.
m : *makam* index, $1<m<9$
n : number of intervals
μ_k (mean): mean of each Gaussian distribution which are defined as fixed pitch values of each *makam* scale in theory.

As a result, each *makam* scale defined in Arel theory is represented as a template, and both the scales defined in theory and the recordings are transformed into computationally comparable formats.

4.1.3. Automatic Classifier

The same classifier defined in Chapter 3 is used. It is designed as a supervised and a non-parametric classifier where each data is labelled to its own class and no probability density is used. This means that the *makam* of each recording are known. Then the classifier is evaluated according to its ability to classify positive (P) and negative (N) samples by their true (T) and false (F) classification rates. Conventionally the design of an automatic classifier consists of four phases**:** data pre-processing, feature extraction, training and evaluation of the classifier. The first two phases are described above. However, since the templates used in the classifier are derived from the pitch interval values defined in theory, construction of the templates did not necessitate a training phase in the implementation of the classifier. Thus, we did not split the data as test and training sets. As a result, our classifier is specifically designed to evaluate the success of Arel theory for MIR studies on Turkish music. The evaluation results of the classifier is shown in Table 4.2 for the quality measures based on the parameters and measures presented in Equation 3.1 given in Chapter 3.

Table 4.2. Evaluation results of the classifier in terms of recall (R), precision (P), and F-measure

	TP	**TN**	**FP**	**FN**	**R**	**P**	**F-measure**
hicaz	20	146	0	6	77	100	87
rast	17	151	2	2	89	89	89
segah	13	140	7	12	52	65	58
kürdili hicazkar	7	143	9	13	35	44	39
huzzam	5	154	9	4	56	36	43
nihavend	16	147	2	7	70	89	78
hüseyni	6	147	14	5	55	30	39
uşşak	13	139	11	9	59	54	57
saba	17	151	4	0	100	81	89
mean	**13**	**146**	**6**	**6**	**66**	**65**	**64**

4.2. Arel Theory: A Computational Perspective

According to the success rate of automatic classification presented in Table 4.2, *makamlar* can be grouped in terms of the F-measure: while the classification results for the *makamlar segah, hüzzam, kürdili hicazkar, hüseyni* and *uşşak* demonstrated low success rates, the remaining *makamlar hicaz, rast, nihavend* and *saba* demonstrated high success rates. However, it is not straightforward to infer that Arel theory is unsuccessful for the first *makam* group from these classification results.

The confusion matrix of the automatic classification presented in Table 4.3 reveals the reason of the low classification success rates of the first *makam* group. It can be seen from the table that the most confused makamlar occur within 2 groups: *makamlar kürdili hicazkar, hüseyni* and *uşşak* are highly confused in the first confusion group (light gray) and *makamlar segah* and *hüzzam* are highly confused in the second confusion group (dark gray).

Table 4.3. The confusion matrix. Two confusion groups are marked with gray levels: segah and hüzzam (dark gray), and kürdili hicazkar, hüseyni and uşşak (light gray).

	hicaz	*rast*	*segah*	*kürdili h.*	*huzzam*	*nihavend*	*hüseyni*	*uşşak*	*saba*
hicaz	-	-	-	-	-	-	-	-	-
rast	-	-	-	-	-	1	1	-	-
segah	-	-	-	3	**4**	-	-	-	-
kürdili h.	3	-	1	-	-	-	2	**3**	-
huzzam	-	-	**9**	-	-	-	-	-	-
nihavend	-	1	1	-	-	-	-	-	-
hüseyni	3	-	-	**4**	-	3	-	**4**	-
uşşak	-	-	1	**6**	-	3	1	-	-
saba	-	1	-	-	-	-	1	2	-

This suggests that the practice of the *makamlar* from each confusion group have similar pitch-frequency histograms. Figure 3.6 presented in Chapter 3 gives visual evidence of such similarities within each confusion group by presenting mean pitch-frequency histogram of each *makam* practice in each confusion group. Mean pitch-frequency histograms of the *makamlar* from the first confusion group (*kürdili hicazkar,* *hüseyni* and *uşşak)* and the second confusion group (*segah* and *hüzzam)* are presented in Figure 3.6.a and b, in Chapter 3, respectively. As can be observed from the figures, the *makamlar* in each confusion group exhibit strong similarities. The practice of *makamlar kürdili hicazkar, hüseyni* and *uşşak* in the first confusion group are quite similar to each other as shown in Figure 3.6.a. The practice of *makamlar segah* and *hüzzam* in the second confusion group are quite similar to each other, which is shown in Figure 3.6.b.

Since the confused *makamlar* showed strong similarities in practice, it is not possible to conclude directly that Arel theory is unsuccessful for the *makamlar* with low classification success rates. Nevertheless, it can be said that the failure of the theory in reflecting the pitch intervals of practice, contributed to the low success rate. Table 4.4 presents the comparison of pitch interval values obtained from the practice and as defined in the theory for each *makam* in the confusion groups. The pitch interval values of the practice are obtained by a peak detection algorithm applied to mean pitch-frequency histogram of each *makam*. A more detailed and comprehensive comparison of the theory and the practice based on pitch interval values is presented by Bozkurt et al. (2009).

Table 4.4. Comparison of pitch interval values obtained from practice (gray) and defined in theory for each makam in the confusion groups.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1st confusion group								
kurdili hicazkar	4.7	12.7	22	31	35.3	43.7	-	53
	4	13	22	31	35	44	-	53
hüseyni	***6.3***	12.3	22	31	***38***	43.3	-	53.3
	8	13	22	31	***39***	44	-	53
uşşak	***6.7***	13	22	31.3	35	44	-	53
	8	13	22	31	35	44	-	53
2nd confusion group								
segah	5	14.3	***23.3&27***	31	36	***46***	49	53.3
	5	14	***22***	31	36	***45***	49	53
huzzam	4.7	14	***21***	30.7	36	48.7	-	53.3
	5	14	***19***	31	36	49	-	53

Table 4.4 demonstrates that the Arel theory converges considerably to practice in terms of the pitch interval values. The pitch interval values of the practice and the theory which diverge by at least 1 Hc are marked as bold italic in the table. Especially the 1st pitch interval values in *makam hüseyni* and *uşşak,* which diverge from the theory, are the intervals subject to the discussions in Turkish music. Another noticeable divergence can be observed from the 3rd pitch interval of *makam segah* as shown in Table 4.4. In

practice, the 3rd pitch interval consists of two pitch interval values: the first one 23.3 Hc diverges from the theory and the second one 27 Hc is lacking in theory. As a result, since the pitch interval values of the confused *makamlar* are also similar, Arel theory can be considered successful except for the few pitch interval values which diverges from the practice.

A similar investigation can also be made for the *makamlar* with high classification success rates, *hicaz, rast, nihavend* and *saba*. Although Arel theory seems to be successful for the automatic classification of practice for these *makamlar*, it is also possible that the theory could diverge from the practice in terms of pitch interval values which contribute to the decrease in success rates. Table 4.5 presents the comparison of pitch interval values obtained from practice (gray) and defined in theory for the *makamlar* with high classification success rates. As can be seen from the Table 4.5, Arel theory considerably converges to practice in terms of pitch interval values except the pitch interval values marked as bold italic.

Table 4.5. Comparison of pitch interval values obtained from practice (gray) and defined in theory for the makamlar with high classification success rates.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
hicaz	4.3	17	21.7	31	35	***37.7***	43.3	53
	5	17	22	31	35	***39***	44	53
rast	9	16.7	21.7	31	40.3	47.7	-	53
	9	17	22	31	40	48	-	53
nihavend	9.7	13.3	22.3	31.3	35.7	44.3	-	53
	9	13	22	31	35	44	-	53
saba	7.3	13	18.7	31.7	34.7	44.3	48	***53***
	8	13	18	31	35	44	49	***-***

Finally, despite the divergence of few pitch interval values, Arel theory seems to provide a valid framework for MIR studies on Turkish music. However, in order to obtain more robust evaluation of the Arel theory, it is necessary to measure the effect of pitch interval values which diverges from practice in automatic classification.

4.2.1. *Makam* Classification based on Pitch Intervals of Practice

In order to measure the effect of divergence of pitch intervals in automatic classification, firstly the pitch interval values obtained from practice are replaced with the theoretical pitch intervals. Therefore, Gaussian representation of *makam* templates are reconstructed by using the pitch interval values obtained from practice. Then the automatic classification of recordings is applied by using the new templates. Finally, the results of automatic classification by using the pitch interval values obtained from the practice can be compared with the automatic classification by using the pitch intervals defined in theory. This comparison would give a more robust evaluation of Arel theory within the same classification context.

The pitch interval values are obtained by applying a peak detection algorithm to the mean pitch-frequency histogram of each *makam*. Then, the templates for each *makam* are computed by using these pitch interval values as new means of the Gaussian distributions as presented in Equation 4.1 and 4.2. Finally, automatic classification is applied by new templates using the same distance measure. Table 4.6 presents the success rates of the automatic classification for each.

Table 4.6. Evaluation results of the classifier based on pitch interval values obtained from practice in terms of recall (R), precision (P), and F-measure

	TP	**TN**	**FP**	**FN**	**R**	**P**	**F-measure**
hicaz	18	149	2	3	86	90	88
rast	18	149	1	4	82	95	88
segah	19	146	1	7	76	95	84
kürdili hicazkar	7	151	9	5	58	44	50
huzzam	11	158	3	0	100	79	88
nihavend	10	150	8	4	71	56	63
hüseyni	5	145	15	7	42	25	31
uşşak	16	129	8	19	46	67	54
saba	17	148	4	3	85	81	83
mean	**13**	**147**	**6**	**6**	**72**	**70**	**70**

Consequently, the classification results obtained by using the pitch interval values based on theory and practice can be compared by looking at Table 4.2 and 4.6, respectively. The two automatic classifications can be desribed as classification based on theory and classification based on practice. First of all, it can be said that the effect of pitch intervals which diverge from practice in automatic classification results 6 % decrease in terms of mean the F-measure: while the success rate of classification based on the theory is found as 64 %, the success rate of classification based on the practice is found as 70 % as can be seen from Table 4.2 and 4.6, respectively.

Although the amount of decrease in success rate seems to be not very significant, there is a considerable amount of increase in the success rates of *makamlar segah* and *hüzzam* as 26 % and 45 % in terms of the F-measure for the classification based on practice (Table 4.6) in comparison to the classification based on theory (Table 4.2). Therefore, it can be said that the most important effect of automatic classification about the divergence between the theory and the practice occurs due to the pitch interval 27 Hc in *makam segah* which is not present in Arel theory. A simple operation of adding this pitch interval value to the theoretical definition of *makam segah* results in a 67.5 % success rate in terms of mean F-measure in automatic classification based on theory. From this it is clear that the amount of decrease, 3.5 % in success rate of the classification based on the theory in comparison to the classification based on the practice, occurs due to the lack of the 27 Hc pitch interval in the theory. A similar effect of divergence in automatic classification about the pitch interval 53 Hc for *makam saba,* which is not part of the theory, is computed in a similar way and found as 1 %. Therefore, the lack of two pitch intervals in theory results in a 4.5 % decrease in automatic classification and the remaining amount of decrease, 1.5 % in the success rate of classification based on the theory, occurs due to the other pitch intervals defined in the theory which diverge from the pitch intervals performed in practice.

On the other hand, there is a considerable amount of decrease, 15 % in the success rate of *makam nihavend* in automatic classification, based on practice in comparison to classification based on theory. Since the pitch interval values obtained from the practice provide the most reliable values, it can be argued that the high success rate of classification based on theory for *makam nihavend* does not reflect a valid success rate.

4.2.2. Arel Theory and the Pitch-Classes for Turkish Music

Our aim was to evaluate Arel theory to understand whether it can supply a basis for MIR studies on Turkish music. So the main point is to evaluate whether Arel theory is valid for its definition of pitch-classes for each *makam*. Without the existence of reliable pitch-class definitions in Turkish music, it would not be possible to apply MIR methods for Turkish music especially for the applications based on temporal information such as automatic transcription of Turkish music. The results presented so far show that Arel theory considerably converges to practice except for the few pitch-intervals (pitch-classes) marked as bold italics in Table 4.4 and 4. 5. Especially the lack of two pitch intervals 27 Hc for *makam segah* and 53 Hc for *makam saba* in theory results in a significant decrease in automatic classification.

Nevertheless, we think that the pitch intervals which are not part of the theory and diverge considerably from the practice can be corrected within the context of templates formed of Gaussian distributions for its application for MIR studies on Turkish music. Firstly, pitch interval values obtained from the practice are used as mean of each Gaussian distribution for each *makam* scale. Secondly, the weight of each pitch from the mean pitch-frequency histogram of each *makam* is applied to the Gaussian distributions. Therefore, templates are reconstructed as the sum of Gaussian distributions as defined in Equation 4.1 and 4.2, only multiplying the amplitude of the distributions by the weights obtained from the practice (Figure 4.2). Consequently, the success rate of the automatic classification based on new Gaussian distributions is found as 75 % in terms of the F-measure, which is 7 % more than the success rate obtained by the data-driven model, presented in Chapter 3.

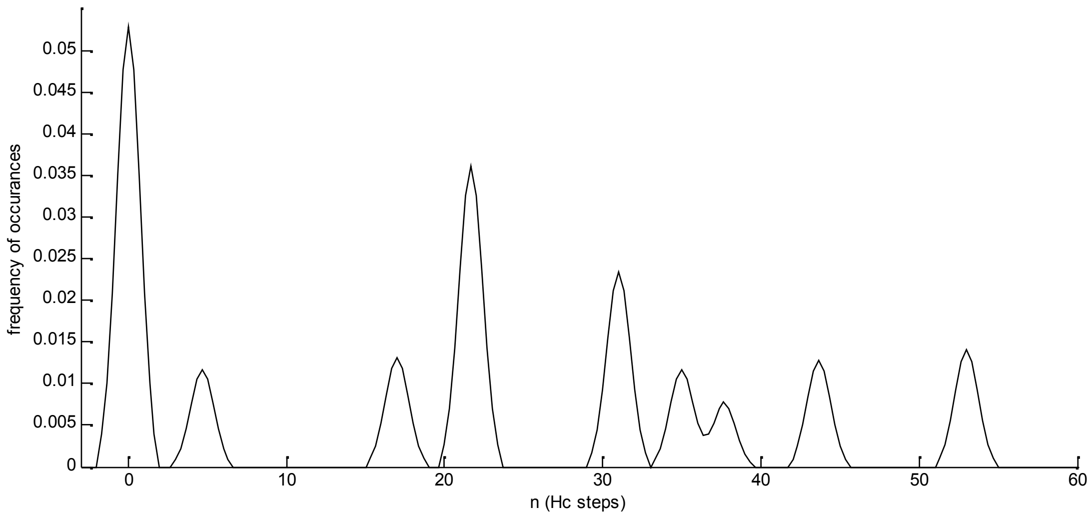


Figure 4.2. Representation of *hicaz makam* template obtained by the new Gaussian distributions where the parameters are obtained from practice.

4.3. Discussion and Conclusion

Since Arel theory is both the most influential theory and subject of discussions about its divergence from practice, we have evaluated it to understand whether it can provide a basis for MIR studies on Turkish music in a similar way western music theory provide for western music. More specifically, the main problem was to understand whether a theory of Turkish music can provide valid pitch-class definitions for MIR studies on Turkish music, as 12-pitch-classes defined in western music theory do.

Since our investigation is intended for MIR applications on Turkish music, we have evaluated the theory within the context of MIR studies. Therefore, current MIR studies on tonality finding are selected as a framework for our evaluation. Due to the significant differences between Turkish music and western music, we have adapted the current methods for Turkish music. In short, the modality (*makam*) templates are constructed based on the Arel theory and a given piece is compared with these modality templates.

It has been shown that despite the few pitch intervals defined in theory but which are not part of practice, Arel theory is found partly successful when applied in a MIR context for Turkish music for the modality finding problem. The effect of pitch intervals defined in theory which diverges from practice results only in a 6 % decrease in terms of the F-measure. However, it has been shown that these few problematic pitch intervals can be improved within Arel theory based on the pitch interval values obtained from practice. Regarding the modality finding problem, it has also been shown that when weights of the templates obtained from practice are used, the success of automatic makam recognition is found as 75%, which is 7 % more than the success rate of the data-driven model found as % 68 in terms of the F-measure. On the other hand, it is clear that without improving the pitch interval values within the theory it will not be possible to apply MIR methods based on temporal information. As a result, we conclude that Arel theory with few improvements could provide valid pitch-class definitions for MIR studies on Turkish music, similar to the 12-pitch-classes defined in western music theory.

Although Arel theory seems to be improvable by slight changes in the pitch-class definitions for computational applications, such changes mean a great change within the logic of Arel theory from the perspective of ethnomusicology. The 24 pitch-classes are the distinctive feature of Arel theory which supports the Arel discourse in terms of “westernnes” and “Turkishness” of Turkish music. However, it has been shown that two pitch intervals seem to be lacking in theory, and there are six pitch-classes defined in theory considerably diverging from practice, more than or equal to 1 Hc (Table 4.4 and 4.5), which points to serious problems for Arel theory from the ethnomusicological point of view.

CHAPTER 5

AUTOMATIC TRANSCRIPTION OF TURKISH MUSIC[17]

This chapter presents an automatic music transcription (AMT) system which accepts monophonic instrumental audio recordings of traditional Turkish art music (shortly Turkish music) as input and outputs transcriptions in the conventional staff notation format which can be used for performance and education. Our problem can be considered wtihin the context of the conventional meaning of transcription in contrast to AMT studies in the literature, since we try to obtain staff notation from recordings of Turkish music for the purposes of performance and education. Therefore, outputs of our study would also enable melodical analysis of Turkish music recordings which could lead to retrieval applications and ethnomusicological analysis. Furthermore there is a wide geographical region such as Middle-East, North Africa and Asia where the musical cultures shares close similarities with Turkish music. Thus, our study would also provide more relevant methods and techniques than the MIR literature for the study of these non-western musics.

Briefly the algorithm presented in this study consists of following steps which also reflect the organization of the paper:

i. Extraction of f0 data
ii. Automatic *makam* classification and tonic pitch detection.
iii. Segmentation and quantization of f0 curve.
iv. Determination of pitch intervals.
v. Note labelling.
vi. Rhythmic analysis and quantization of duration.
vii. Auditory/visual graphical user interface.
viii. Representation of transcription in MIDI format and staff notation.

Since the first 2 steps of the system, extraction of f0 data, and Automatic *makam* recognition and tonic pitch detection are considered within Chapter 3 and 4, we present the rest of the steps in this chapter.

---

[17] A version of this chapter was submitted to Journal of New Music Research by Ali C. Gedik and Barış Bozkurt.

Finally, we present the evaluation results of automatic transcription based on 5 monophonic instrumental recordings[18] in comparison to manual transcriptions of 2 musicians. Automatic and manual tanscriptions are evaluated with a reference to original notations. Therefore, we proposed a solution to the evaluation problem of automatic transcription by using both original notation and manual transcriptions for the first time in the literature. As a result, while automatic transcription outperforms manual transcriptions for 2 recordings, success rates of automatic transcription for the rest of 3 recordings are found close to the success rates of manual transcription. Finally we also discussed the evaluation results qualitatively.

5.1. Segmentation

Cascade approach to segmentation and labeling of f0 curve usually apply a model similar to a blackboard system proposed by Bello et al. (2000). Segmentation and labeling is usually done by applying each step based on rules, thresholds or tuning parameters. However, onset detection as one of the research domain in MIR and a robust method for segmentation takes little attention in studies based on cascade approach. Onset detection is either applied by algorithms far from the state-of-art (e.g. McNab and Smith 2000; Bruno and Nesi 2005; Antonelli and Rizzi 2008; Paiva et al. 2008) or simply not applied (e.g. Haus and Pollastri 2001; Clarisse et al. 2002; De Mulder et al. 2004).

Holzapfel, Stylianou, Gedik and Bozkurt (2010) studied problems of onset detection for Turkish music for the first time. They proposed a fusion algorithm for pitched instruments of Turkish music in comparison with western music instruments. Fusion algorithm consists of three onset detection algorithm, spectral flux (SF) and phase slope (PS) and onset detection approach based on f0 change (F0). While 57 recordings corresponding to 1829 onsets are used for evaluation, 21 recordings corresponding to 674 onsets are used for training. As a result following success rates in terms of F-measure are obtained: 74.1% for F0, 73.9% for SF, 73.7 % for PS and %82.1 for fusion algorithm. Due to the simplicity, computational costs and close successs rates, we preffered onset detection based on f0 change in our AMT system for

[18] Recordings and corresponding original notations are obtained from http://www.neyzen.com/. Reocrodings are performed by a well-known neyzen Salih Bilgin.

the segmentation block. If the difference between two successive f0 intervals is more than 2 Hc, then it is decided that there is an onset, since 4 Hc is defined as the smallest pitch interval in theory at least.

Figure 5.1 shows how segmentation block worked on an excerpt of a sample recording. The *makam* of the recording is found as *hüzzam* whose tonic pitch is A4#8 and corresponding note name is *segah,* and the tonic pitch value is found as 7294 cent. Firstly, tonic value is substracted from f0 curve and then the resulting f0 values are converted to Holder comma. Therefore pitch intervals are obtained with a resolution of 53 Hc/octave as shown in Figure 5.1.

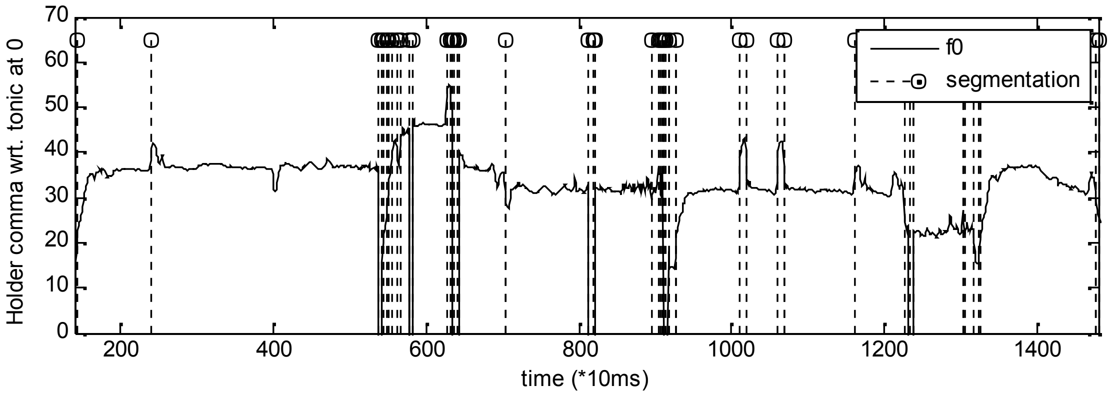


Figure 5.1. Segmentation.

5.2. Quantization of f0 Segments

Once the f0 curve is segmented it is necessary first to quantize and then to label each segment with note names. Median is the most frequent operation used for quantization of f0 segments (Haus and Pollastri 2001; Clarisse et al. 2002; Adams et al. 2006; Paiva et al. 2008; Typke 2011). Before labeling, ornamentations such as vibrato and glissando or articulations such as legato are detected in contrast to statistical approach. Haus and Pollastri (2001) detect both vibrato and legato. Vibrato is defined as a regular modulation of 4/7 Hz and legato is defined as adjacent segments 0.8 semitone apart. However quantitative results are not presented due to the challenge in detection of exact moment between two legato notes as noted in this study. Similarly Pollastri (2002) applies 0.8 semitone threshold for the detection of vibrato and legato. De Mulder et al. (2004) apply legato detection for segments longer than 300 ms by looking for multiple stable intervals having gabs in between. Paiva et al. (2008) present detection of vibrato and glissando in their automatic transcription study where threshold of 1 semitone is used for vibrato and constant increase or decrease of successive short notes for glissando.

Two studies that apply neither HMM model nor cascade approach are as follows: Adams et al. (2006) applied Kalman filter for the detection of vibrato based on statistical learning and Typke (2011) applies a clustering algorithm for note segmentation and glissandi detection.

There are also studies focused on detection of ornamentations and articulations rather than automatic transcription, for the purpose of automatic music tutor either for singing or violin performance. Studies of Mayor et al. (2006; 2009) apply expression categorization for singing by HMM to detect attack, sustain release and vibrato. Loscos et al. (2006) apply vibrato detection based on amplitude and frequency modulations for automatic violin transcription. The study of Barbancho et al. (2009) presents both articulation detection such as detache, pizicato, spicatto and ornamentation detection such as vibrato for automatic violin expressive detection. Their vibrato detection is based on the difference of average frequency of the segment and some threshold values. The study of Zhang and Wang (2009) fuse the audio and visual information for automatic transcription of violin. Although this study covers ornamentation detection such as vibrato and legato, it is not clear how they are applied.

Nevertheless, most of the studies on automatic transcription of singing or fretless instruments do not cover any ornamentation detection (e.g. McNab and Smith 2000; Hu and Dannenberg 2002; Clarisse et al. 2002; Bruno and Nessi 2005; Antonelli and Rizzi 2008; Fujihara and Goto 2011; Rao and Rao 2010).

Although not applied in an AMT study there exists also which give place or focus on ornamentations. Ryynanen (2006) in his review of singing transcription defines some ornamentations as follows: vibrato is defined as a rate of 4-7 Hz with a depth between 0.34-1.23 semitones and reported that mean of the f0 curve of a vibrato is close to the perceived pitch; glissando is defined as a musical event performed at the beginning of long notes start with a low frequency and reach to the note within 200 ms. Casey and Crawford (2004) presents automatic detection of trills and chord-spreading performed in two 17th and 18th c. lute music. Gainza and Coyle (2007) apply automatic ornamentation detection for Irish music performed with tin whistle, flute and pipe. The ornamentations of this study are peculiar to Irish music and its instruments such as cut, strike, roll, etc. The study of Duggan et al. (2008) deals again with ornamentations peculiar to Irish music but in a retrieval context.

We again applied a rule-based algorithm for quantization in our system. Each segment is searched for the existence of vibrato and glissando. According to the type of ornamentation, two different methods of quantization are applied, one for vibrato and the other for glissando. Firstly, segments are classified as vibrato or glissando segment according to the following rules: If the difference between maximum and minimum f0 values of a segment is less than or equal to 3 Hc , it is classified as a vibrato segment. Otherwise the segments are classified as glissando segment if their durations are also more than 150 ms. Therefore possible appoggiatura and acciaccatura segments with a duration less than 150ms left unclassified either as vibrato or glissando. Similarly such short ornamentations are left almost unchanged if classified as a vibrato segment. Such segments are useful for onset detections in the proceeding blocks but they are cancelled at the transcription block by a duration filter for the sake of obtaining a simple notation.

Quantization of vibrato segments: Median of the f0 values of each vibrato segment is calculated and set as the quantized value. Figure 5.2 shows some of the quantized vibrato segments marked with ellipse

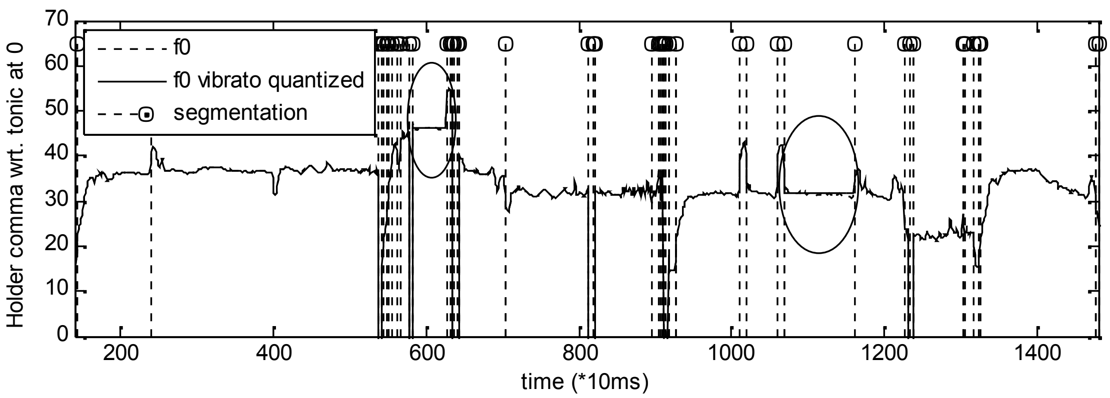


Figure 5.2. Quantization of vibrato segments.

Quantization of glissando segments: Figure 5.3 shows each glissando segment marked with 2 stems showing start and end of each glissando segment. It is clear from the figure that a glissando segment can also be a combination of long ornamentations such as glissando and vibrato and as well as short ornamentations such as appagiaturas and acciaccaturas. As a result quantization of a glissando segment corresponds to the quantization of long ornamentations and cancellation of short ornamentations.

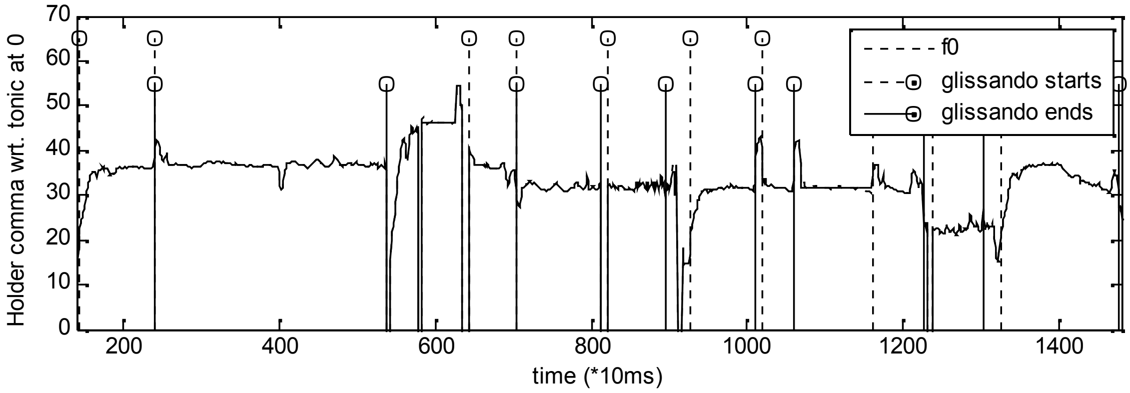


Figure 5.3. Classification of glissando segments.

Therefore it is possible to quantize each glissando segment seperately. Three level of segmentation and quantization are applied for quantization of glissando segments based on rule-based algorithms similar to the rules applied in vibrato quantization. Firstly f0 values of a glissando segment is filtered by a median filter as follows:

$$f0med(n)=median\{[f0(n-M) \dots f0(n+M)]\},\ M= 150\ ms \quad (5.1)$$

- 1st level: Each filtered glissando segment is segmented within itself by applying following rule: segmentation is applied if the difference between the two successive f0 values are more than 1/5 Hc. Then each segment is quantized if the difference between the maximum and minimum values of f0 values is less than 2 Hc. Median of the f0 values are calculated and set as the quantization value.
- 2nd level: The resulting glissando segment is segmented again according to the following rule: segmentation is applied if the difference between the two successive f0 values are more than 1 Hc. Then each segment is quantized if the difference between the maximum and minimum values of f0 values is less than 6 Hc. Median of the f0 values are calculated and used as quantization value.

- <u>3rd level:</u> A final segmentation is applied as follows: if the difference between the two successive f0 values are different than 0 Hc, a new segment is detected. Finally the segments less than 200ms are omitted by sharing their durations equally with the neighbor segments. This final operation leads to the cancellation of short ornamentation notes such as appagiaturas and acciaccaturas.

As a result, first glissando segment of the f0 curve shown in Figure 5.3 is presented as an example of glissando quantization in Figure 5.4.

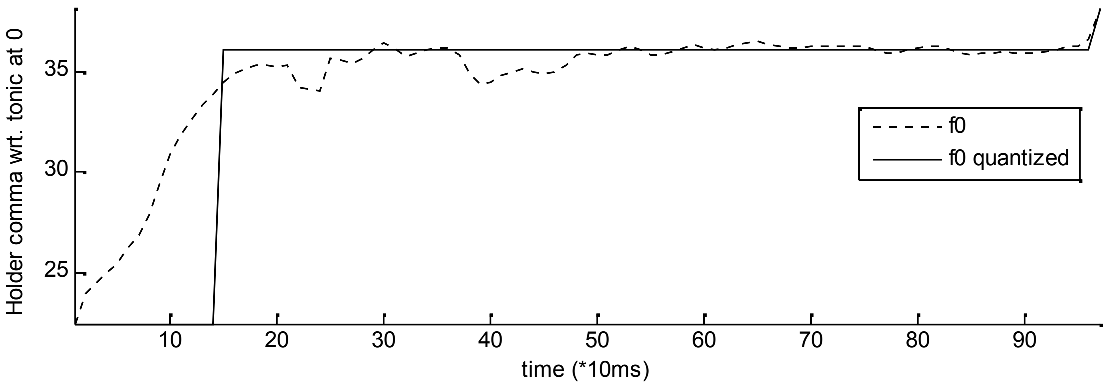


Figure 5.4. Quantization of glissando segments.

Final quantization result is shown in Figure 5.5. As can be seen from the figure there are short ornamentations, since they are either classified as vibrato segment or remained unclassified due to their durations less than 150 ms.

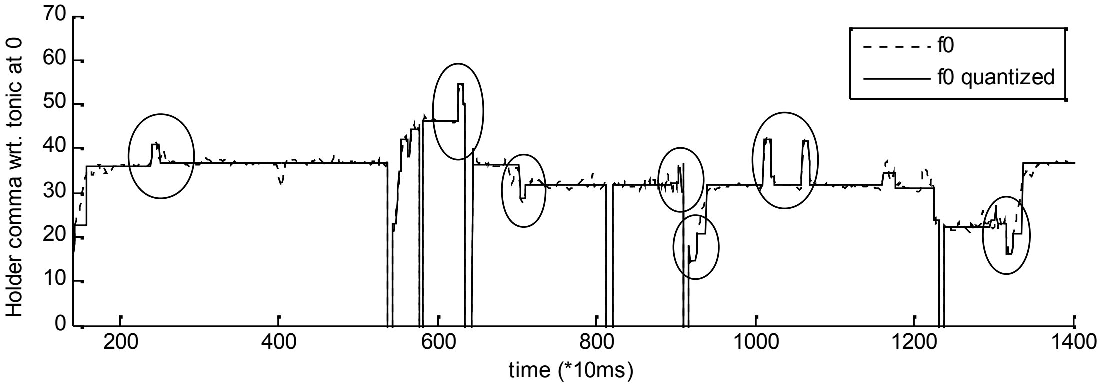


Figure 5.5. Quantization of glissando segments.

5.3. Note Labeling

In order to label quantized f0 curve with note names, a list of note names for 8 octaves is used. 53 note names for each octave is listed which corresponds to the resolution of quantized f0 curve (53 Hc/octave). Therefore successive notes are listed in the list in order to have 1 Hc difference in between, as follows: C1…C4#1 C4#2 C4#3 C4#4 C4#5 C4#6 C4#7 C4#8 D4...C8.

The distance between each pitch in the f0 curve and the tonic pitch is calculated from the note list and the corresponding note name is found. As an example, we can consider the same *hüzzam* recording. If we want to find the note name of a pitch interval of 36 Hc with respect to tonic, then we should start counting 36 note names from the note list starting from the tonic note, which is A4#8 for makam *hüzzam*. Considering a whole note is 9 Hc and a half note is 4 Hc in Turkish music the distance of 36 Hc wrt. tonic pitch can be calculated as follows: the difference between A4#8 and C5 is 5 Hc. There are 31 steps left which corresponds to the remaning distance starting from C5. It is clear that 31 Hc starting from C5 leads to G5, since 3 whole tones (C5-D5, D5-E5, F5-G5) which corresponds to 3 x 9 = 27 Hc and a half tones (E5-F5) which corresponds to 4 Hc between C5 and G5 makes 31 Hc in total. Therefore we reach to the note name G5 for the pitch interval of 36 Hc.

In order to get rid of short ornamentations in the quantized f0 curve for the sake of an easy readable notation, a final filter is applied which cancels the notes shorter than 150 ms by sharing durations of canceled notes equally between neighboring notes. Since such ornamentations mark onsets and offsets which are necessary for the note durations, the onsets of the cancelled notes are kept. Finally, Figure 5.6 shows the resulting onsets and note names above the f0 curve depending on pitch interval values of quantized f0 curve and the tonic note name found for the same excerpt of *hüzzam* recording.

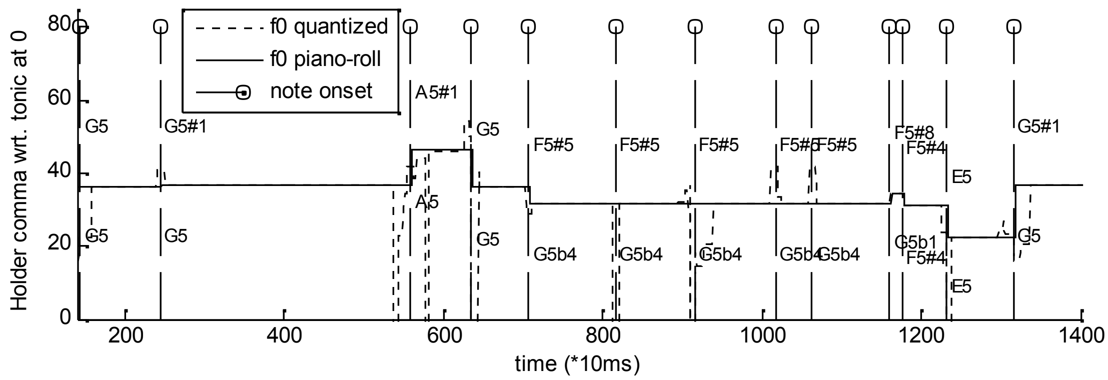


Figure 5.6. Note labeling: note names above the f0 curve found according to the resolution of 53 Hc/octave and note names below the f0 curve defined in theory.

However, the pitch space of performance in Turkish music is much richer than the pitch space defined in theory. Therfore pitches of the performance obtained according to the resolution of 53 Hc/octave is converted to the closest pitches defined in theory, since we aim to obtain conventional staff notation which is used by musicians. Following rules are applied for this conversion:

- #1 is converted to natural if the note is not F, since #1 is only defined for F in theory.
- #2 is converted to #1
- #3 is converted to #4
- #6 is converted to #5
- #7 is converted to #8

Again for the sake of easy readability of notation, sharps more than 4 are expressed in terms of flats; eg. A4#8 is written as B4b1 which is the actual representation of note *segah*. Similarly e.g. D5#5 is written as E5b4 which is the actual representation of note *hisar*. A final correction for the pitch intervals and note names is applied as follows: if the difference between two successive notes is 1 Hc, which is not a musical interval, then the most frequent one used in the whole piece is attended as the pitch interval of the less frequent one. As a result, Figure 5.6 shows the converted note names below the f0 curve. As can be seen from the figure the second note G5#1 and third note A5#1 are converted to natural G5 and A5, etc. Successive notes F5#5 are expressed as F5b4 and similarly F5#8 is expressed as G5b1. List of transcribed notes are written to a table to be used for conventional staff notation.

5.4. Quantization of Note Durations

Duration of notes also should be quantized in order to be represented as conventional note durations such as 1/16, 1/8, 1/4 etc. There are mainly three approaches for duration quantization; statistical approach, ratio approach and an approach based on tempo set by the user. Viitaniemi et al. (2003) use distribution of durations obtained from EsAC-database for the quantization of note durations of a given piece. Adams et al. (2006) applies a uniform quantization where duration levels are assumed to be uniformly distributed. Ratio approach is mainly applied in QBH systems where note durations need not to be represented conventionally. Both Haus and Pollastri (2001) and Unal et al. (2008) apply ratio of durations of consecutive notes. McNab et al. (1996) for a tempo of 120 beat/minute (bpm) set 125 ms as semiquaver (1/16) and use the resolution of semiquaver. McNab and Smith (2000) set the duration of the shortest note as semiquaver which is another expression of previous work. In other words, each note is quantized to the nearest semiquaver according to the tempo set by the user (McNab et al. 2000). Duggan et al. (2008) use a duration histogram where the highest peak is determined as the quaver note. Meek and Birmingham (2002) apply Inter-Onset-Interval (IOI) for the quantization of note durations by applying a 29 level quantization on a logarithmic scale.

However most of the studies do not study the duration quantization problem (e.g. Bello et al. 2000; Clarisse et al. 2002; De Mulder et al. 2004; Bruno and Nesi 2005; Paiva et al. 2008; Antonelli and Rizzi 2008; Rao and Rao 2010; Fujihara and Goto 2011). In this sense rhythm detection is almost out of scope within many automatic transcription studies, although there are studies focused only on rhythm detection. Hainsworth (2006) presents a detailed review for those studies. We apply the method based on duration histogram for the quantization of durations. In order to obtain a robust histogram, a rounding operation is applied to the first digit of millisecond for each note duration (e.g. 1317 ms is rounded to 1320 ms). Finally, the highest peak of the duration histogram is set as the eighth note (1/8) . Figure 5.7 shows duration histogram of the same sample of *hüzzam* recording. As can be seen from the figure eighth note is found as 0.4 sec. marked with solid ellipse and double of it, quarter note 0.8 sec. marked with dotted ellipse.

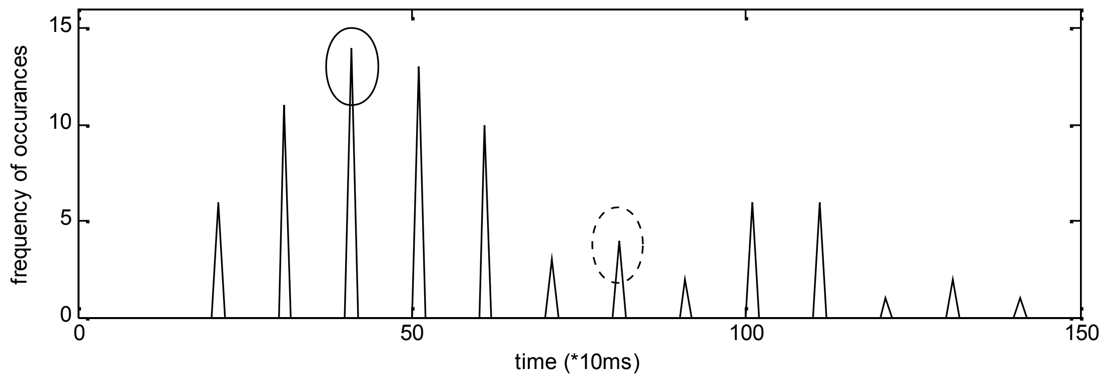


Figure 5.7. Duration histogram.

As a result, each note duration is divided to 0.4 sec. and expressed in terms of eighth note. In order to use these duration values for conventional staff notation, each value is expressed as simplest integer ratios such as 3/16, 5/8 etc., and numerator and denumerator for each duration is written to a list beside note names.

However it is not possible to proceed to rhythm analysis after the quantization of duration in Turkish music considering the state of art. One of the most important reason for this fact is the complexity of rhythmic structure in Turkish music involving compound rhythmic patterns such as 5/4, 7/8, 9/8. Furthermore when a percussion instrument is lacking, it becomes challenging to detect onsets robustly. Nevertheless, some preliminary tests are conducted for rhythm analysis by Bozkurt and Gedik (2010). Spectral flux is used for onset detection and auto-correlation function of onset signal is calculated for beat detection. While tests on 21 synthetic recordings give successful results for beat detection, tests on 4 sample audio recordings give very poor results. Therefore we leave this topic to future work.

5.5. Transcription and Graphical User Interface

As aforementioned, GUI provides a tool for the training of the system as shown in Figure 5.8.. GUI also enables users to realize the automatic transcription by her/himself and to correct any faulty information occurred at blocks of the automatic transcription as follows:

i. Open a monophonic instrumental audio recording of Turkish music.
ii. Find the *makam* of the recording and if faulty choose the correct *makam* from

the menu.

iii. Detect the tonic of the recording and if faulty set the tonic.

iv. Listen to the whole or selected part of the original recording in comparison with synthesized sound obtained from the extracted f0 data.

v. Observe visually and auditorily the final transcription in terms of note names and piano-roll and resulting MIDI in turn.

vi. Correct any faulty pitch by selecting its region.

Finally automatic transcription by the use of GUI also produces a list similar to MIDI format and this format can be opened by the software MUS2 which produce conventional Turkish staff notation. Figure 5.9 (top) shows the information used by MUS2 represented on the piano-roll: note names and durations represented as simple fractions. Figure 5.9 (middle) shows the conventional staff notation produced by MUS2. Finally Figure 5.9 (bottom) shows the original notation. In overall Figure 5.9 both demonstrates the result of automatic transcription and the divergence between the notation and performance in terms of pitch intervals and duration. Although notation dictates F5#4, the performer plays F5#5 (G5b4) as can be seen from the f0 curve. Similarly while notation dictates E5b4, the performer plays E5 natural as can be seen from the f0 curve.

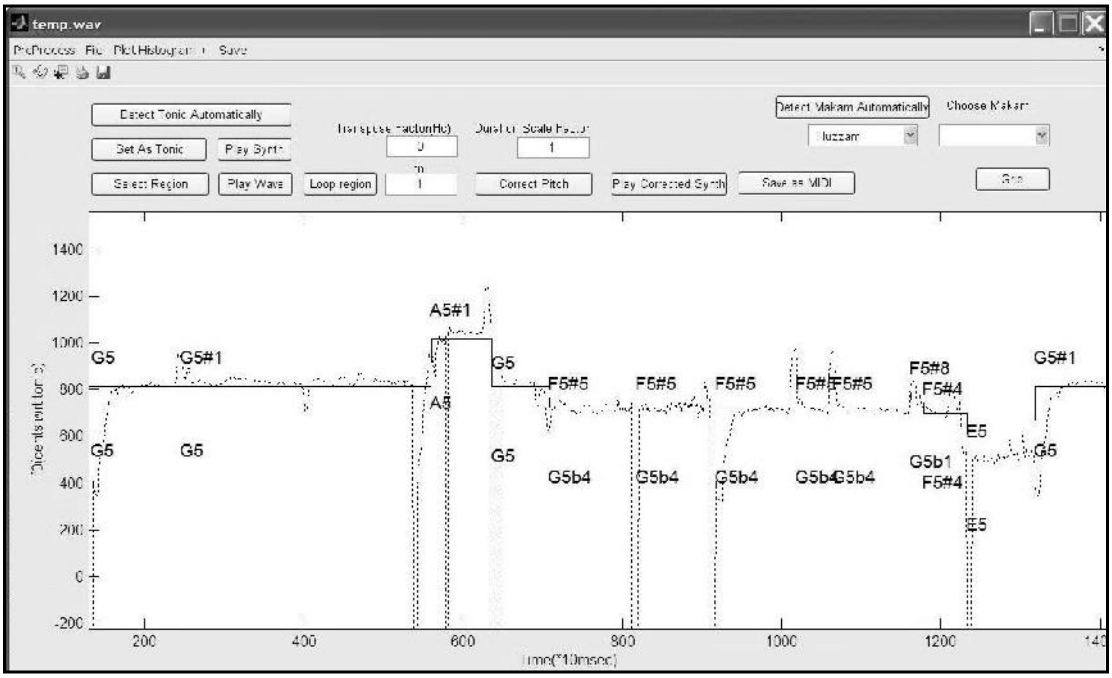


Figure 5.8. Graphical User Interface coupled with the automatic transcription system.

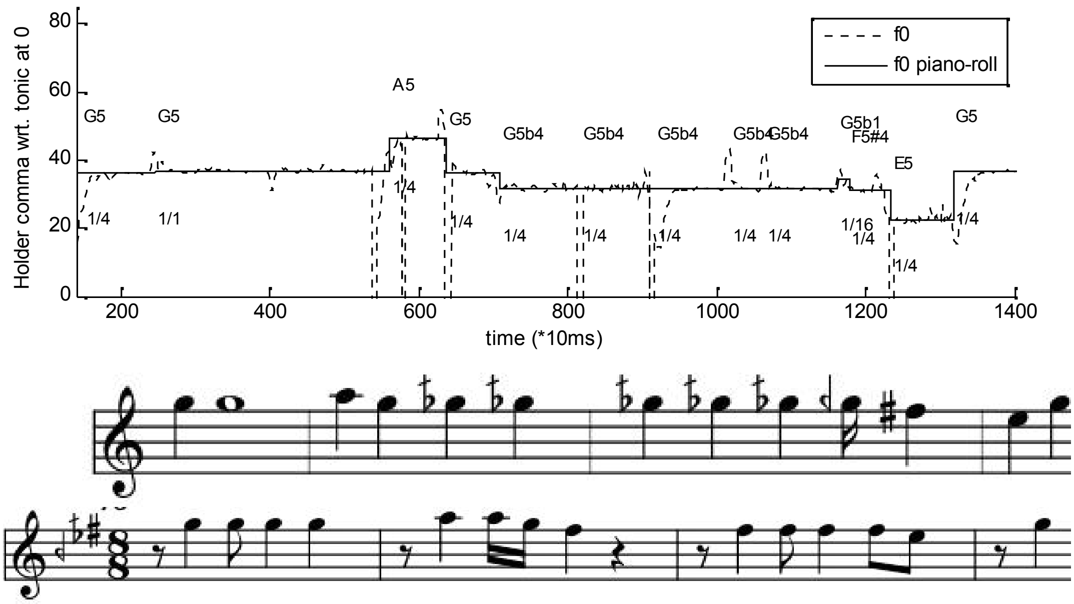


Figure 5.9. Transcription example: (top) shows the piano-roll representation; (middle) shows the conventonal staff notation produced bu MUS2; (bottom) shows the original notation.

As a result, representation of automatic transcription as a conventional staff notation and as a more detailed notation via GUI, supplies two kinds of notations as defined by ethnomusicologists: prescriptive notation and descriptive notation. While prescriptive notation can be used for performance and education, descriptive notation can be used in ethnomusicological studies on Turkish music.

5.6. Evaluation

There are mainly two approaches for the evaluation of automatic transcription in the literature: comparison of transcription and reference notation (original notation or manual transcription) and simply measuring the success of retrieval operation.. Since the latter approach to evaluation is for retrieval applications, it is out of our scope.

There are various metrics for the evaluations based on comparison of transcription and reference notation. One of them is edit distance (ED) where the transcription is compared with reference notation on the basis of number of correct, inserted and deleted notes (e.g. Martol 2003; Krige and Niesler 2006; Jiang et al. 2007; Unal et al. 2008). However the effect of duration or onset/offset times on the success rate is not clear in these studies.

Fonseca and Ferriera (2009) classify evaluation metrics as frame-based and note-based approaches. Frame-based approach is based on the comparison of two notations for every 10 ms (Dixon 2000). Note-based approach is based on classification metrics such as false negatives, false positives, recall, precision and f-measure. Transcribed notes are classified as correct if their onsets are within a certain neighbourhood of the onsets of the reference data and if difference of their pitch values are under some threshold value, usually half of a semi-tone. The neighboorhood for onset is defined in the literature as a threshold between ±25 - ±150ms (Fonseca and Ferreira 2009). Some of these studies also take duration into account by defining a overlap ratio for the definition of correctly detected notes (e.g. Ryynanen and Klapuri 2005, 2006, 2008; Antonelli and Rizzi 2008). Overlap ratio determines the tolerance for the original and transcribed notes in terms of their overlapping onset and offset times.

As a result, evaluation of AMT is mainly based on quantitative measures leaving out the questions about false transcribed notes. This quantitative approach toward evaluation makes the details of the process inaccessible. Especially when manual transcriptions are used as reference data, the procedure applied is not clear in the literature. In this sense Daniel et al. (2008) focus on the perceptual evaluations of listeners for the transcription errors and use these data for developing a perceptual-based evaluation.

However the use of original notation or manual transcription as reference data for evaluation is also problematic even for western music as our discussion on the concepts of descriptive notation and prescriptive notation showed. No doubt this approach is much more problematic for Turkish music due to the divergence of theory and performance. Two studies clearly focus on the problems of notation system in Turkish music. Ayangil (2008: 445) especially underline that although the performance styles such as melodic and rhythmic variations and ornamentations constitutes one of the most important characteristic of Turkish music, they are not represented in notation. Similarly Kaçar (2005) empirically shows this problem by comparing the notation of pieces and the performances of pieces by master musicians.

Therefore, an objective evaluation of automatic transcription system is also problematic. In order to handle this challenging problem we applied a cross-evaluation method for the first time in the literature. In short, we asked 2 locally well-known performers having a formal education on Turkish music, to manually transcribe the pieces which are selected for the evaluation of our AMT system. They have 2 alternative approaches for transcription: one is simple transcription without ornamentations in accordance with the tradition of notation in Turkish music and the other is more complicated transcription including both ornamentations and performance styles. Although both musician were familiar with both approach, we asked them to transcribe as simple as possible, similar to original notations in Turkish music.

As a result, outputs of automatic transcription and manual transcriptions are compared with original notations as a more objective measure for the success of our AMT system. Success rate of AMT system is measured by both note-based and frame-based evaluation methods.

Recall, precision and F-measures are used in note-based evaluation where TP (True Positives), FN (False Negatives) and FP (False Positives) corresponds to number of correctly transcribed notes, the number of notes not transcribed and the number of notes not present but transcribed, in turn. 150ms tolerance for onset and threshold of 3 Hc (approximately half of a semitone) condition for pitch difference for a correctly transcribed note are applied.

Overall overlap ratio is calculated by the following measure:

$$overlap\_ratio = \frac{\min\{offsets\} - \max\{onsets\}}{\max\{offsets\} - \min\{onsets\}} \tag{5.2}$$

Offsets and onsets in the overlap ratio measure correspond to offsets and onsets of a correctly transcribed note and its reference note. Therefore for each note pair ovelap ratio is calculated based on the minimum and maximum of offsets and onsets. Mean of the overlap ratios gives the overal overlap ratio. Finally a simplification proposed by Jiang et al. (2007) for note-based evaluation is applied: Silence in the transcription is deleted and adjacent notes with the same tone are merged as one note.

Frame-based evaluation is found by the following measure:

$$Acc = \frac{TP}{(FP + FN + TP)} \tag{5.3}$$

Again threshold of 3Hc is used for the classification of correctly transcribed notes for frame-based evaluation. Finally Table 5.1 presents evaluation results for both evaluation approaches. We will first consider the note-based evaluation results. The success rates of automatic transcriptions are almost equal to the success rates of manual transcriptions for piece #2 uşşak and piece #4 hüseyni. Automatic transcription outperforms manual transcriptions for piece #2 uşşak. The success rates of automatic transcriptions are lower than manual transcriptions for rest of the three recordings, piece #1 *hüzzam*, piece #3 *hicaz* and piece #5 *saba*.

Table 5.1. Evaluation results for 3 kinds of transcriptions for 5 recordings. Manual 1 and 2 corresponds to the transcriptions of two musicians.

Piece		**Transcription**	**Note-based evaluation**		**Frame-based evaluation**
#	Makam		F-measure	Overlap Ratio	Accuracy
1	*Hüzzam*	Manual 1	0.3111	0.9571	0.7593
		Automatic	*0.1263*	*0.5293*	*0.6789*
		Manual 2	0.8298	0.9382	0.9717
2	*Uşşak*	Manual 1	0.1163	0.8325	0.7110
		Automatic	*0.1887*	*0.6620*	*0.4788*
		Manual 2	0.1771	0.8869	0.3577
3	*Hicaz*	Manual 1	0.4867	0.9505	0.6547
		Automatic	*0.2300*	*0.5504*	*0.5895*
		Manual 2	0.6766	0.9319	0.8274
4	*Hüseyni*	Manual 1	0.8909	0.7381	0.8243
		Automatic	*0.6990*	*0.6624*	*0. 7444*
		Manual 2	0.5255	0.5477	0.6815
5	*Saba*	Manual 1	0.2646	0.9010	0.5629
		Automatic	*0.2097*	*0.4766*	*0.5314*
		Manual 2	0.5748	0.9384	0.7710

If we consider frame-based evaluation results, automatic transcription of piece #2 *uşşak* and piece #4 *hüseyni* do not outperform manual transcriptions, success rates of automatic transcriptions are within the confidence interval of success rates of manual transcriptions. Again the success rates of automatic transcription is more close to the success rates of manual transcriptions for the rest of other 3 recordings in comparison to the success rates found in note-based evaluation. This fact can be seen from Table 5.2. While the mean success rate of automatic transcription is much lower in note-based evaluation, it is much higher in frame-based evaluation.

Table 5.2. Overall evaluation results for 3 transcriptions.

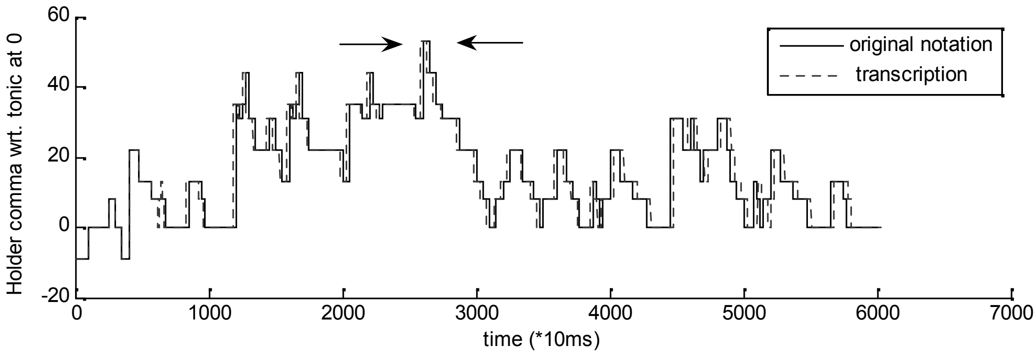
Transcription	Note-based evaluation		Frame-based evaluation
	Mean F-measure	Mean Overlap Ratio	Mean Accuracy
Manual 1	0.4139	0.8758	0.7024
Automatic	0.2907	0.5761	0.6046
Manual 2	0.6517	0.8390	0.8129

In fact, frame-based evaluation for all transcriptions gives more optimistic results in comparison to note-based evaluation. Since the same pitch interval value, 3Hc, is used as a threshold for correct transcribed notes, the main difference should result from the difference of approach to note-onsets in two evaluation metrics. In other words, while note-based evaluation has a note-onset threshold, 150 ms, frame-based evaluation does not use any condition for note-onsets.

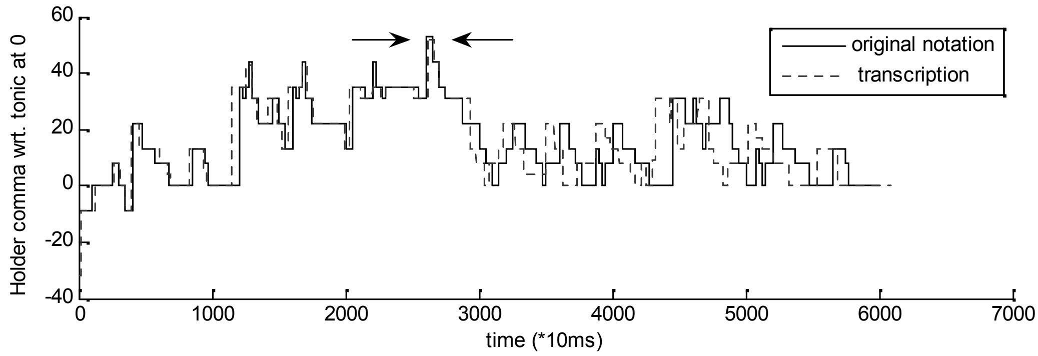
5.7. Discussions

In order to discuss the evaluation results qualitatively we firstly present piano-roll representation of 3 transcriptions of piece#2 *uşşak* in comparison to original notation as shown in Figure 5.10.

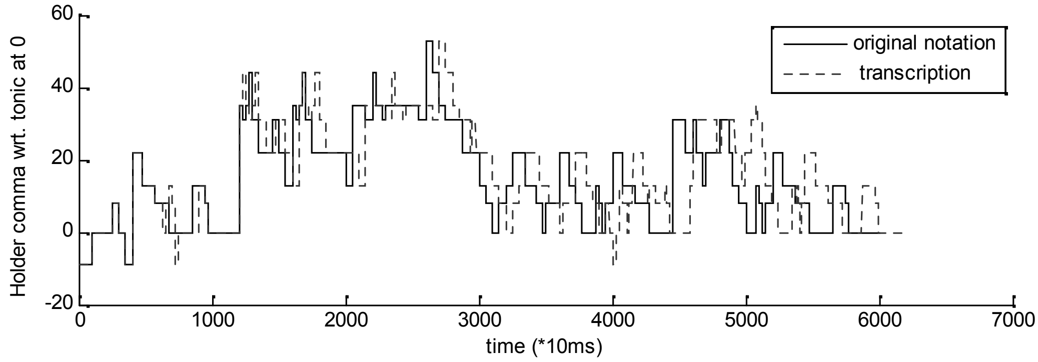
Transcriptions of this piece are the most interesting example for discussion. Although it is the worst successful example for both evaluation approaches, there is an interesting difference between the results of the note-based and frame-based approach as can be seen from Table 5.1. This fact is especially observable when the two kinds of success rates are compared for the manual transcription #1 (Figure 5.10 –top); % 12 for note-based and % 70 for frame-based.



Manual transcription #1



Automatic transcription



Manual transcription #2

Figure 5.10. Transcriptions of piece #2 *uşşak* in comparison to original notation: Manual transcription #1 (top), automatic transcription (middle) and Manual transcription #2 (bottom).

Although transcription seems to be so close to the original notation from the figure, the onsets of the notes of the transcription and original notation marked with arrows are 200 ms far away which is above the evaluation threshold 150 ms. Therefore the results of the note-based approach is not as flexible as frame-based approach and leads to dramatic decrease in the success rates even for the transcriptions fits to the original notation. The same considerations are also valid for the automatic transcription shown in Figure 5.10 (middle). Again note onsets of the transcription and notation marked with arrows seems to be very close but their distance is above the onset threshold. The transcriptions of other 4 recordings demonstrates the same fact also (see Appendix A for all transcriptions and original notations represented in piano-roll format). Therefore we can conclude that frame-based approach gives not only more optimistic results but also more realistic results for our cases. Finally Appendix B presents all transcriptions and original notations represented in conventional staff notation.

5.8. Conclusion

In this chapter, we presented an AMT system designed for Turkish music for the first time in the literature. We also proposed a new evaluation approach due to the ambiguity of the automatic transcription problem in MIR literature. The contribution of our approach is the use of 2 different manual transcriptions for the evaluation of automatic transcriptions in comparison to original notation. Our final contribution is the two kinds of transcriptions our system produce: a prescriptive notation for the use in performance and education and a descriptive notation for the use of ethnomusicological analysis. While prescriptive notation gives conventional staff notation, descriptive notation gives details of the performance in comparison to prescriptive notation via GUI designed.

As a result, AMT system consists of several block which extracts f0 data from audio recordings, automatically recognize *makam* of the recording and its tonic pitch, segment the f0 data and quantize both f0 segments and its durations and finally label them with note names. While the success rate of our system is found %60 in terms of accuracy, the success rates of 2 manual transcriptions are found % 70 and % 81.

CHAPTER 6

DISCUSSION AND CONCLUSION

This thesis presented an AMT system for Turkish music for the first time in the literature. In order to construct this system several issues and applications were discussed and applied. Firstly, automatic music transcription problem was considered from the perspective of ethnomusicology. Secondly, an automatic *makam* recognition system was developed. Thirdly, the Turkish music theory was evaluated computationally in order to understand whether it can be used for MIR studies on Turkish music.

Although each title is considered with its own discussion and conclusion section within each chapter, it is necessary to consider several issues in detail. Following subsections tries to handle these issues.

6.1. Automatic *Makam* Recognition[19]

Possible problems about our study on automatic *makam* recognition should be mentioned related with the data used. First of all, recordings consist of performances in the form of *taksim*. Arel does not give place to forms in his book, but it is known that the distinguishing feature of the form *taksim* is the modulations (i.e. short-term *makam* changes) used during a performance. Therefore, a *taksim* performance of a specific *makam* naturally shows the characteristics of other *makamlar* where it is modulated. However, the weight of these modulations changes from performance to performance which can be estimated intuitively as between 10-30 percent with respect to the whole performance. Without the existence of an automatic segmentation algorithm, it is not possible to detect the modulations in the performance. Automatic detection of

[19] This section is adapted from Gedik, A. C. and Bozkurt, B.(2009). Evaluation of the Makam Scale Theory of Arel for Music Information Retrieval on Traditional Turkish Art Music, *Journal of New Music Research*, 38(2): 103-116.

modulations is among our future goals. Therefore, this study lacks an analysis of the effect of modulations to the classifier's performance.

The other two main problems related to the data, which probably affected the classification results, is their representation value of the practice. Firstly, most of them were recorded in sound studios, far from the natural contexts of musical performance. Although we do not have enough information about the general conditions of all recordings, at least Ünlü (2004:199) reports the terrible psychological mood of Tanburi Cemil Bey during the recording sessions. Of course, the time limitations due to the recording technologies should have also affected the performances. Second, the time period of recordings spread roughly between the years 1900 and 2000. So it is hard to say that the practice is left unchanged during a century which prevents to make strict generalizations over them. It should be added that the modernization process which makes the "traditional art music" one of the most popular genre between 1950 and 1960 (Aksoy 2006:17) has effected the practice too.

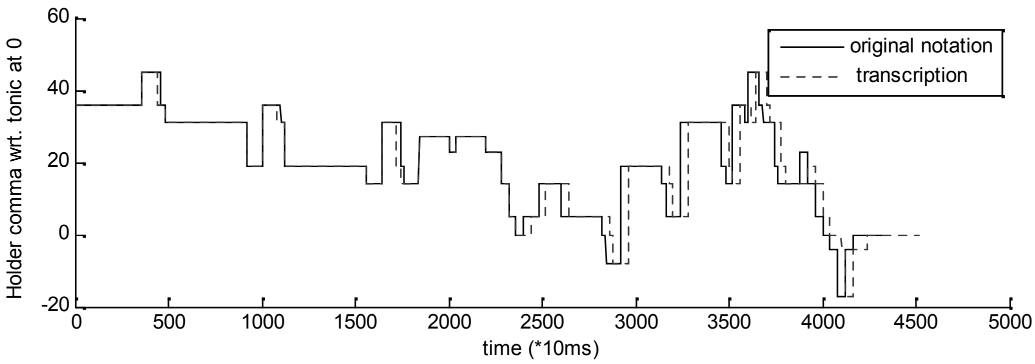
However, as mentioned before one of the most challenging issues for the pitch-class analysis of practice in Turkish music is the freedom of musicians in the performance of pitches. The performances of the same *makam* by the two most prominent performers of Turkish music, Tanburi Cemil Bey and his son Mesut Cemil, demonstrate this freedom as shown in Figure 3.2 presented in Chapter 3. Even the most characteristic pitch intervals of *makam hicaz*, 1^{st}, 2^{nd} and 3^{rd} intervals, are performed at different values. Furthermore, while all other theories including the Arel theory give almost the same interval value at least for the 3^{rd} interval as 22 Hc (Bozkurt et al. 2009), Tanburi Cemil Bey and Mesut Cemil perform this interval as 21.3 and 21.7, respectively. Consequently, such performance characteristic of pitch-classes can be considered as one of the most important divergence between theory and practice which seriously affected the success of automatic classifications presented in this study.

Finally, what makes the divergence between theory and practice more apparent in the 20^{th} century seems to be that the perception of theory has been more important than before in the circle of traditional Turkish art music. Although the practice seems to develop its own course, it is clear that the practice defines itself with reference to theory as stated by Marcus (1993:50): He quotes from Wright that "the smaller intervals of theory were then sometimes "enlarged" in practice", and adds, based on his study in Egypt, that "today, when theory dictates a large interval, musicians speak of "shrinking" these intervals". In this sense, Table 4.4 seems to demonstrate both tendencies of the performers since the theory covers both types of intervals: while the third and the sixth interval defined for *segah* seem to be “enlarged”, the first interval defined for *uşşak* and *hüseyni* seem to be “shrinked”.

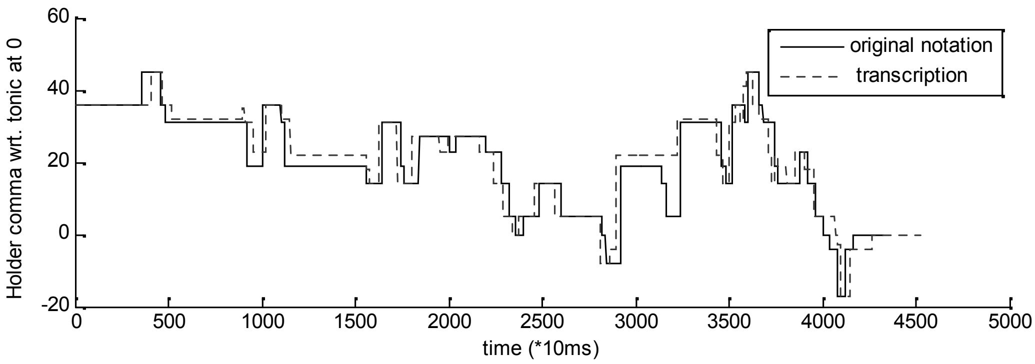
6.2. Automatic Transcription of Turkish Music

One of the problems of evaluation results about automatic transcription is already mentioned under the subsection 5.5 in Chapter 5, “Transcription and Graphical User Interface”. The real pitch values of the performance (e.g. E5) diverges both from pitch values of the original notation and transcription (e.g. E5b4) as shown in Figure 5.9. Since the example presented in Figure 5.9 is the same with the piece #1 *hüzzam*, the same consideration holds true for the the piece #1 *hüzzam* as shown in Figure 6.1 (middle). This fact can be observed from the pitch intervlas of manual and automatic transcriptions around 20 Hc in Figure 6.1 (middle). As shown also at Table 4.4 in Chapter 4, this interval is one of the pitch intervals leading to divergence of theory and practice: while the pitch interval of performance is reported as 19 Hc, the pitch interval defined in theory is 21 Hc. However we see that the performer of the piece #1 *hüzzam* preferred a higher pitch value. Naturally 4 Hc difference is evaluated as false transcription both by the note-based and frame-based approaches for automatic transcription.

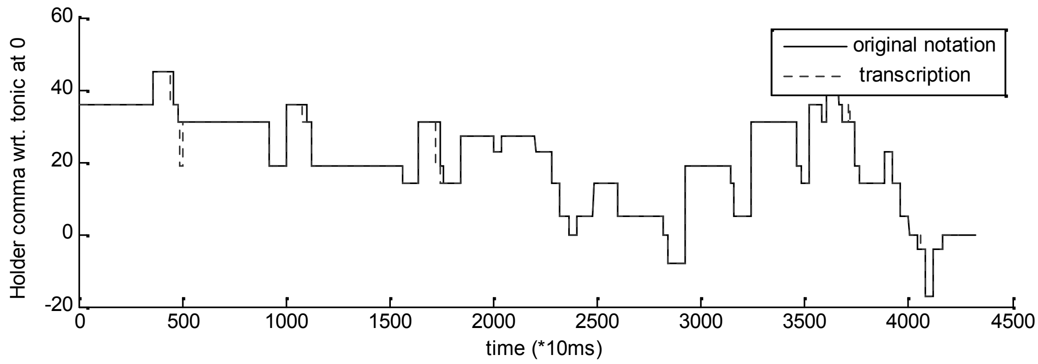
Similarly the real pitch values of the performance (e.g. D5) of the piece #3 *saba* diverges both from pitch values of the original notation and manual transcriptions (e.g. D5b4). Naturally 4 Hc difference is evaluated as false transcription both by the note-based and frame-based approaches for automatic transcription. This fact again can be observed from the pitch intervlas of manual and automatic transcriptions around 20 Hc in Figure 6.2 (middle). As shown also at Table 4.5 in Chapter 4, this interval is one of the pitch intervals leading to divergence of theory and practice: while the pitch interval of performance is reported as 18.7 Hc, the pitch interval defined in theory is 18 Hc. However we see that the performer of the piece #3 *saba* preferred a higher pitch value.



Manual transcription #1

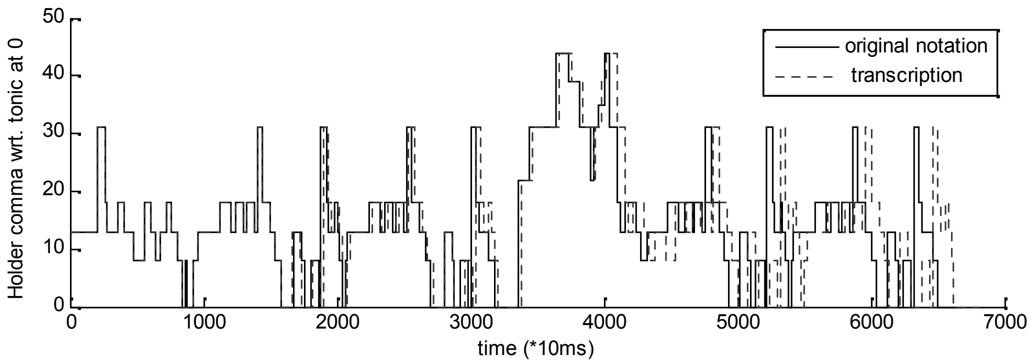


Automatic transcription

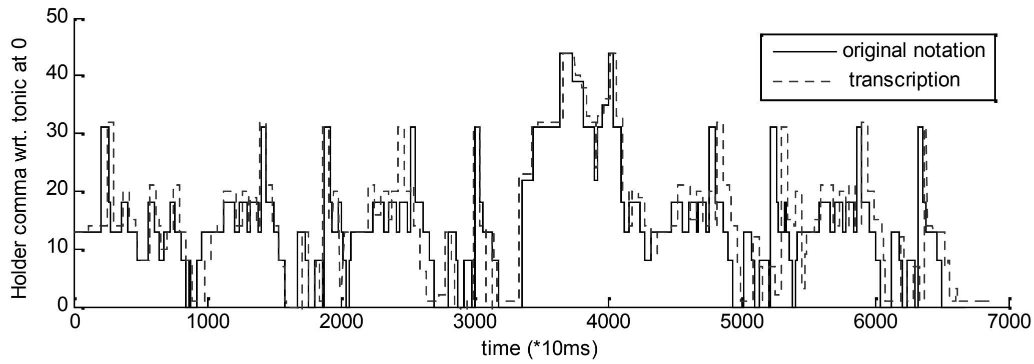


Manual transcription #2

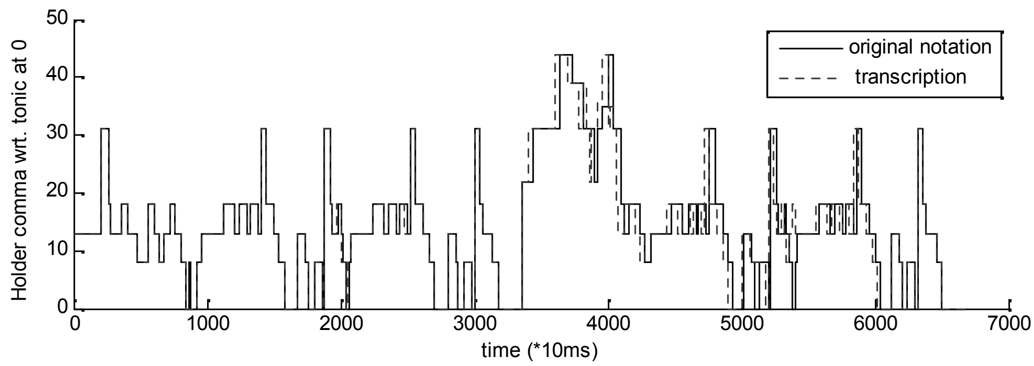
Figure 6.1. Transcriptions of piece #1 *hüzzam* in comparison to original notation: Manual transcription #1 (top), automatic transcription (middle) and Manual transcription #2 (bottom).



Manual transcription #1



Automatic transcription



Manual transcription #2

Figure 6.2. Transcriptions of piece #5 *saba* in comparison to original notation: Manual transcription #1 (top), automatic transcription (middle) and Manual transcription #2 (bottom).

This is not surprising since we already mentioned the divergence of theory and performance. This fact shows that the behavior of musicians whom transcribed the recordings was according to theory not to how they hear. Currently our transcription system can not handle such common-sense behavior. However, it is clear that in order to represent recordings with conventional staff notation, it is necessary to develop such methods. Therefore, one reason of low success rates of automatic transcription results comes from the lack of such mappings that would reflect deficiencies of the theory to represent practice (or to copy the errors of the theory to represent practice).

Another problem can be observed from Figure 5.10 which shows piano-roll representation of transcriptions similar to original notation but shifted in time. This is not peculiar only to automatic transcription and manual transcriber #2. In some other examples manual transcriber #1 shifted transcriptions in time while the manual transcriber #2 performs more close to original notation (see Appenix A). There is no doubt that automatic transcription simply follows the f0 data extracted in both time and frequency dimensions and gives the actual performed pitch interval. However this fact again leads to low success rates. Possibly, if rhythm analysis could be applied, then it would be easier to fit the automatic transcription to original notation in terms of tempo.

The results of our system are in accordance with the resuts of the emprical studies of List (1974) and Stockmann (1979) on manual transcription from the perspective of ethnomusicology. Both studies discuss the reliability of manual transcriptions and show that different participants gives out transcriptions with a certain amount of difference primarily for the durations and secondarily for the pitches of notes. It should also be mentioned that both transcribers reported a certain difficulty about fitting the note durations within the tempo of performances in their transcription experience. This fact demonstrates an additional problem about transcription; Although the performed pieces have certain tempo values, the performer prefers rather a flexible interpretation of the given tempo in the original notations.

Finally, it is clear that outputs of our system would be valuable for musicians in their performances and education. While descriptive transcription supplies all details of performance such as ornamentation and performance styles of a performance, prescriptive notation supplies staff notation of any performance within the limits of our system.

6.3. Future Work

The limitations of our study can be listed as follows:

- Accepts only monophonic instrumental audio recordings
- Rhythm analysis is missing
- Although automatically recognizes *makam* of the recordings, unable to use this knowledge for the use of accidentals in the transcription. This is not possible due to the possible modulations (*geçki*) in a given recording.
- Small number of test recordings for the evaluation of our AMT system.

Each of the items about the limitations of this thesis can be thought as a future work. On the other hand there is a more urgent issue about the transcription of Turkish music. Our system is mainly evaluated by the performances of compositions. However, improvisations also cover an important place in Turkish music. As mentioned by the participants in Chapter 2 even manual transcription of improvisations are highly challenging. Therefore, there is much less notations of improvisations in Turkish music in comparison to compositions.

This fact points a more urgent need for an AMT which could accept improvisations. Good news about our AMT system is that improvisations in Turkish music is mainly performed monophonically either by instrument or vocal, which are called as *taksim* and *gazel,* respectively. Furthermore, improvisations in Turkish music are mainly performed in free rhythm. This is another advantage of our system for *taksim* transcriptions which leaves out the rhythmic analysis. Therefore, it can be said that theoretically our system can also work for *taksim* recordings, as well. However in order to speak practically about our system's performance on *taksim* recordings, it is necessary to handle an evaluation for such recordings. No doubt any evaluation of our system for *taksim* recordings would require original notations which are seldom found and manual transcriptions which are less trustable in comparison to compositions. Nevertheless, regarding the importance of improvisation in Turkish music, our first future plan is to handle this challenging task.

REFERENCES

Adams, N.H., Bartsch, M.A. and Wakefield, G.H. (2006). Note segmentation and quantization for music information retrieval IEEE Transactions on Audio, Speech, and Language Processing, 14 (1): 131 – 141.

Akdoğu, O. (1993). Türk müziği nazariyatı tarihine genel bir bakış, in Türk musikisi Nazariyatı Dersleri, ed. Onur Akdoğu, Ankara: Kültür Bakanlğı Yay. (1989.). Taksim nedir, nasil yapilir?, Izmir.

Akkoç, C. (2002). Non-deterministic scales used in traditional Turkish music, *Journal of New Music Research*, vol. 31, no. 4. pp. 285-293.

Aksoy, B. (2006). Osmanlı musikisinin popüler kültür çevresinden çıkışı, in *20. yıl: Pan'a armağan*, İstanbul: Pan Yay.

Al-Taee, M. A., Al-Ghawanmeh, M. T., Al-Ghawanmeh, F.M. and Al-Own, B. O. A. (2009). Analysis and Pattern Recognition of Woodwind Musical Tones Applied to Query-by-Playing, *Proceedings of the World Congress on Engineering 2009* Vol I WCE 2009, July 1 - 3, 2009, London, U.K.

Antonelli, M. and Rizzi, A. (2008). A Correntropy-based voice to MIDI transcription algorithm. *MMSP'2008*. pp.978~983.

Arel, H. S. (1993). *Türk musikisi Nazariyatı Dersleri*, Onur Akdoğu(ed.), Ankara : Kültür Bakanlğı Yay.

Argenti, F., Nesi, P. and Pantaleo, G. (2011). Automatic transcription of polyphonic music based on the constant-Q bispectral analysis, *IEEE Transactions on Audio, Speech, and Language Processing*, 19 (6): 1610-1630.

Ayangil, R , (2008). Western Notation in Turkish Music, *Journal of the Royal Asiatic Society*, 2008 - Cambridge Univ Press.

Bello, J. P., Monti, G. and Sandler, M. (2000). An Implementation of Automatic Transcription of Monophonic Music with a Blackboard System, *Proceedings of the Irish Signals and Systems conference (ISSC 2000)*, Dublin, Ireland.

Bent, I. D. et al. (2011). "Notation." In *Grove Music Online*. *Oxford Music Online*, http://www.oxfordmusiconline.com/subscriber/article/grove/music/20114pg1 (accessed September 13, 2011).

Bohlman, P. V. (2008). 'Middle East, §I: Concepts of music', *Grove Music Online* ed. L. Macy (Accessed 09 February 2008), http://www.grovemusic.com

Bozkurt, B. (2008). An automatic pitch analysis method for Turkish maqam music. *Journal of New Music Research*, 37(1) : 1-13.B.

Bozkurt, B., Yarman, O., Karaosmanoğlu, M. K. and Akkoç, C. (2009). Weighing Diverse Theoretical Models On Turkish Maqam Music Against Pitch Measurements: A Comparison Of Peaks Automatically Derived From Frequency Histograms With Proposed Scale Tones, *Journal of New Music Research*, 38 (1): 45-70.

Brunelli, R. and Poggio, T. (1993). Face recognition: Features versus templates, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15 (10): 1042-1052.

Bruno, I. and Nesi, P. (2005). Automatic Music Transcription Supporting Different Instruments, *Journal of New Music Research*, 34 (2): 139-149.

Burke, P. (2009). *Popular culture in early modern Europe*, Ashgate Publishing, Ltd., 3rd Revised edition edition.

Cambouropoulos E. (2003), Pitch Spelling: A Computational Model, *Music Perception* 20 (4): 411-430.

Cambouropoulos E. (2001), Automatic Pitch Spelling: From Numbers to Sharps and Flats. In *Proceedings of the VIII Brazilian Symposium on Computer Music,* 31 July - 3 August 2001, Fortaleza, Brasil.

Cambouropoulos E. (2000), From MIDI to Traditional Musical Notation. In *Proceedings of the AAAI Workshop on Artificial Intelligence and Music: Towards Formal Models for Composition, Performance and Analysis* 30 July - 3 Aug 2000, Austin, Texas.

Carterette, E.and Kendall, R. (1999). Comparative music perception and cognition, in: D. Deutsch (Ed.), *The Psychology of Music* (2nd ed.), San Diego: Academic Press, 1999, pp. 725-792.

Casey, M.A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C. and Slaney, M. (2008). Content-based music information retrieval: Current directions and future challenges*, Proc. IEEE 96* (4) 668-696.

Casey, M. and Crawford, T., (2004). Automatic Location and Measurement of Ornaments in Audio Recordings', in C. L. Buyoli and R. Loureiro (eds.), *Fifth International Conference on Music Information Retrieval: Proceedings* (Pompeu Fabra University, Barcelona, 2004): 311-317.

Cemgil, A. T. (2004). *Bayesian Music Transcription.* Ph. D. thesis, Radboud University of Nijmegen.

Cha, S.-H. S. and Srihari, N. (2002). On measuring the distance between histograms, *Pattern Recognition* 35 (2002): 1355–1370.

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

Chew, E., (2002). The spiral array: An algorithm for determining key boundaries, *Proc. LNCS/LNAI 2445*, Scotland: Springer, (2002) pp.18–31.

Chordia, P. and Rae, A. (2007). Raag recognition using pitch-class and pitch-class dyad distributions. *Proc. International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria, 23-27 September 2007.

Chordia, P., Godfrey, M., and Rae, A. (2008). Extending content-based recommendation: The case of Indian classical music*, Proc. International Conference on Music Information Retrieval (ISMIR)*, Philadelphia, Pennsylvania, USA, 14-18 September 2008, pp. 319-324.

Chordia, P. and Rae, A. (2007). Modeling and visualizing tonality in North Indian classical music, *Neural Information Processing Systems, Music Brain Workshop (NIPS 2007)*, 2007.

Chuan, C. and Chew, E. (2005). Polyphonic audio key finding using the spiral array CEG algorithm, *IEEE Conf. on Multimedia and Expo (ICME)*, Amsterdam, Netherlands, 6-8 July 2005, pp.21-24.

Chuan, C-H. and Chew, E. (2007). Audio key finding: Considerations in system design and case studies on Chopin's 24 preludes, *EURASIP Journal on Advances in Signal Processing*, 2007, Article ID 56561, 15 pages.

Clarisse, L., Martens, J., Lesaffre, M., Baets, B., Meyer, H. and Leman, M. (2002). An auditory model based transcriber of singing sequences. In *Proceedings of the Third International Conference on Music Information Retrieval: ISMIR 2002*. 116-23.

Cornelis, O., Lesaffre, M. Dirk Moelants, Marc Leman, (2010), Access to ethnic music: Advances and perspectives in content-based music information retrieval, *Signal Processing*, 90 (4) :1008–1031.

Daniel, A., Valentin, E., David, B. (*2008)*. Perceptually-Based Evaluation of the Errors Usually Made When Automatically Transcribing Music. In *Proceedings of ISMIR'2008*. pp.550~556

De Mulder, T., Martens, J.P., Lesaffre, M., Leman, M., De Baets, B., and De Meyer, H. (2004). Recent improvements of an auditory model based front-end for the transcriptionof vocal queries. *Proceedings ICASSP 2004*.

Dixon, S., . On the computer recognition of solo piano music. *Australasian Computer Music Conf.*, 2000.

Duggan, B., O'Shea, B. and Cunningham, P. (2008) A System for Automatically Annotating Traditional Irish Music Field Recordings, *Sixth International Workshop on Content-Based Multimedia Indexing*, Queen Mary University of London, UK, Jun. 2008.

Duggan, B., O'Shea, B., Gainza, M., and Cunningham, P. (2009). The Annotation of Traditional Irish Dance Music using MATT2 and TANSEY In *8th Annual Information Technology & Telecommunication Conference* (2009).

Ellingson, T. (1992a). Transcription, in *Ethnomusicology: an introduction* (Helen Myers, ed.), Norton/Grove Handbooks in Music. New York: Norton, 1992.

Ellingson, T. (1992b). Notation, in *Ethnomusicology: an introduction* (Helen Myers, ed.), Norton/Grove Handbooks in Music. New York: Norton, 1992.

Ellingson, T. (2011). "Transcription (i)." *Grove Music Online*. *Oxford Music Online*. 13 Sep. 2011 <http://www.oxfordmusiconline.com/subscriber/article/grove/music/28268>

Erol, A. (2009). *Müzik Üzerine Düşünmek*, Bağlam Yayınları / Müzik Bilimleri Dizisi, İstanbul.

Faruqe, O. Hasan, A., Ahmad, S. and Bhuiyan, F. H. (2010). Template music transcription for different types of musical instruments, 2010 *The 2nd International Conference on Computer and Automation Engineering ICCAE (2010)*, Volume: 5, Publisher: IEEE, pp. 737-742.

Fawcett, T. (2006). An introduction to ROC analysis, *Pattern Recognition Letters*, 27 (8): 861-874.

Feldman, W. (1990). Cultural Authority and Authenticity in the Turkish Repertoire, *Asian Music*, 22 (1): 73-111.

Fonseca, N. and Ferreira, A. (2009). Measuring Music Transcription Results Based on a Hybrid Decay/Sustain Evaluation, *ESCOM 2009 - 7th Triennial Conference of European Society for the Cognitive Sciences of Music*; Finland, 2009.

Fujihara, H. and Goto, M. (2011). Concurrent estimation of singing voice F0 and phonemes by using spectral envelopes estimated from polyphonic music. *ICASSP* 2011: 365-368.

Gainza, M., and Coyle, E. (2007). Automating Ornamentation Transcription. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, ICASSP 2007, April 15-20, 2007, Honolulu, Hawaii, USA. pp. 69-72.

Gedik, A. C., Bozkurt, B. And Cirak, C. (2009). A Study of Fret Positions of Tanbur Based on Automatic Estimates From Audio Recordings, *Proc. CIM09 (Conference on Interdisciplinary Musicology)*, 26-29 Oct., Paris.

Gedik, A.C. and Bozkurt, B. (2008). Automatic classification of Turkish traditional art music recordings by Arel theory, *Proc. Conf. on Interdisciplinary Musicology*, Thessaloniki, Greece, 3-6 July 2008, web.auth.gr/cim08/cim08_papers/Gedik-Bozkurt/Gedik-Bozkurt.pdf

Gedik, A. C. and Bozkurt, B. (2009). Evaluation of the Makam Scale Theory of Arel for Music Information Retrieval on Traditional Turkish Art Music, *Journal of New Music Research*, 38 (2): 103-116.

Gedik, A. C. and Bozkurt, B. (2010). Pitch Frequency Histogram Based Music Information Retrieval for Turkish Music, *Signal Processing*, 10:1049-1063.

Gomez, E. (2006). Tonal description of polyphonic audio for music content processing, *INFORMS Journal on Computing*. Special Cluster on Music Computing, 18 (3) (2006) pp. 294-304.

Gomez, E. and Herrera, P. (2008). Comparative analysis of music recordings from western and non-western traditions by automatic tonal feature extraction, *Empirical Musicology Review*, 3(3): 140-156.

Gomez, E. and Herrera, P. (2004). Estimating the tonality of polyphonic audio files: Cognitive versus machine learning modelling strategies, *Proc. International Conference on Music Information Retrieval (ISMIR)*, Barcelona, Spain, 10-14 October 2004, pp. 92–95.

Hainsworth, S. W. (2003). *Techniques for the automated analysis of musical audio*, (Ph. D. thesis), Cambridge Univ.

Haus, G. and Pollastri, E. (2001). An Audio Front End for Query-by-Humming Systems, *2nd International Symposium on Music Information Retrieval, ISMIR2001*, Indiana, USA, Oct 2001, pp 65-72.

Heijink, H., Desain, P., and Honing, H. (2000). Make me a match: An evaluation of different ap- proaches to score-performance matching. *Computer Music Journal*, *24(1)*, 43–56.

Holzapfel, A., Stylianou, Y., Gedik, A.C. and Bozkurt, B. (2010). Three Dimensions Of Pitched Instrument Onset Detection, *IEEE Trans. on Audio, Speech and Language Procesing*, 18(6): 1517-1527.

Hu, N. and Dannenberg, R. (2002). A Comparison of Melodic Database Retrieval Techniques Using Sung Queries, in *Joint Conference on Digital Libraries*, New York: ACM Press, (2002), pp. 301-307.

Huron, D., (1996). The melodic arch in Western folksongs." *Computing in Musicology,* 10: 3-23.

Jiang, D.N., Picheny, M. and Qin, Y. (2007). Voice-melody Transcription under a Speech Recognition Framework, *Proc. of ICASSP* 2007.

Juhász, Z. and Sipos, J. (2010). A Comparative Analysis of Eurasian Folksong Corpora, using Self Organising Maps, *journal of interdisciplinary music studies*, 4 (1): 1-16.

Kaçar, G. Y., (2005). Geleneksel Türk Sanat Müziği'nde Süslemeler ve Nota Dışı İcralar [Ornamentations and Non-note Based Performances in Traditional Turkish Art Music], GÜ, *Gazi Eğitim Fakültesi Dergisi*, 25(2): 215-228.

Kranenburg, P. , Garbers, J., Volk, A., Wiering, F., Grijp, L. P. and Veltkamp, R. C. (2010). Collaboration Perspectives for Folk Song Research and Music Information Retrieval: The Indispensable Role of Computational Musicology, *journal of interdisciplinary music studies,* 4 (1): 17-43.

Kapur, A., Percival, G., Lagrange, M., and Tzanetakis, G. (2007). "Pedagogical Transcription for Multimodal Sitar Performance," *In Proceedings of the International Conference on Music Information Retrieval*, Vienna, Austria, September 2007.

Karaosmanoğlu, M.K. and Akkoc, C. (2003). Turk musikisinde icra - teori birligini saglama yolunda bir girisim. *Proceedings from 10th Müzdak Symposium*, İstanbul, Turkey, 2003.

Karaosmanoğlu, M.K., (2004) Turk musikisi perdelerini olcum, analiz ve test teknikleri. *Proceedings from Yıldız Teknik Üniversitesi Turk Muziği Geleneksel Perdelerini Çalabilen Piyano İmâli Projesi sunumu*, İstanbul, Turkey. 2004.

Karaosmanoğlu, M.K., (2007). *Turk musikisinden secmeler*, Nota Yayincilik, Istanbul, 2007.

Klapuri, A. (2006). Introduction to music transcription, in: A. Klapuri and M. Davy, (Ed.), *Signal Processing Methods for Music Transcription*, Springer-Verlag, New York, 2006, pp. 3-20

Klapuri, A. P. (2004). Automatic music transcription as we know it today, *Journal of New Music Research*, 33 (3): 269–282.

Krige, W.A and Niesler, T.R. (2006). *An HMM Based Singing Transcription System. Proceedings of the seventeenth annual symposium of the Pattern Recognition Association of South Africa (PRASA)*, Parys, South Africa, November 2006. ISBN 0-6203-7384-9.

Krishnaswamy, A. (2003a). On the twelve basic intervals in South Indian classical music, *AES 115th Convention*, New York, USA, 10-13 October 2003, paper no: 5903.

Krishnaswamy, A. (2003b). Pitch measurements versus perception of South Indian classical music, *Proc. Stockholm Music Acoustics Conference (SMAC-03)*, 6-9 August 2003, vol.2, pp. 627-630.

Krishnaswamy, A. (2003c). Application of pitch tracking to South Indian classical music, *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 19-22 Oct. 2003, pp. 49.

Krishnaswamy, A. (2004). Multi-dimensional musical atoms in South Indian classical music, *Conf. on Music Perception and Cognition, Evanston*, Illinois, USA, 3-7 August 2004, http://www-ccrma.stanford.edu/~arvindh/cmt/icmpc04.pdf

Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*, Oxford University Press, New York, 1990.

Lartillot, O., Toiviainen, P. and Eerola, T. (2008). Commentary on ‘comparative analysis of music recordings from western and non-western traditions by automatic tonal feature extraction’ by Emilia Gómez, and Perfecto Herrera, *Empirical Musicology Review*, 3(3): 157-160.

Lerdahl, F. and Jackendoff, R. (1983). *A generative theory of tonal music*, MIT Press, Cambridge, Massachusetts. 1983.

Lee, K. and Slaney, M. (2008). Acoustic chord transcription and key extraction from audio using key-dependent HMMs trained on synthesized audio, *IEEE Transactions on Audio, Speech and Language Processing*, 16 (2): 291-301.

Li, C.L. and Hui, K.C. (2000). A template-matching approach to free-form feature recognition, *Proc. IEEE International Conference on Information Visualization*, (2000) 427-433.

Lidy, T., Silla Jr., C. N., Cornelis, O., Gouyon, F., Rauber, A., Kaestner, C. A.A., Koerich, A. L. (2010) On the suitability of state-of-the-art music information retrieval methods for analyzing, categorizing and accessing non-Western and ethnic music collections, *Signal Processing* 90:1032–1048.

List, G. (1974). The reliability of transcription, *Ethnomusicology*, 18(3): 353-377.

Liu, Y. Y. Wang, A. Shenoy, W-H. Tsai, and L. Cai, (2008). Clustering music recordings by their keys, *Proc. International Conference on Music Information Retrieval (ISMIR)*, Philadelphia, Pennsylvania, USA, 14-18 September 2008, pp. 319-324.

Longuet-Higgins, H. C. and M. J. Steedman, (1971). On interpreting Bach, *Machine Intelligence*, 6 (1971) 221–241.

Loscos, A., Wang, Y., and Boo, W. (2006). Low Level Descriptors for Automatic Violin Transcription, *ISMIR*, Victoria, BC, 2006.

Marolt, M. (2004). Networks of Adaptive Oscillators for Partial Tracking and Transcription of Music Recordings, *Journal of New Music Research*, 33 (1): 49–59.

Mayor, O.**,** Bonada, J. and Loscos, A. (2006). The Singing Tutor Expression Categorization and Segmentation of the Singing Voice**.,** In *Proceedings of the 121st Audio Engineering Society Convention*.

Mayor, O., Bonada, J. and Loscos, A. (2009). Performance Analysis and Scoring of the Singing Voice. *AES 35th International Conference: Audio for Games*.

Marcus, S. (1993). The interface between theory and practice: Intonation in Arab music, *Asian Music*, 24(2): 39-56.

Marandola, F. (2003). The study of musical scales in Central Africa: The use of interactive experimental methods, *Proc. Computer Music Modeling and Retrieval*, 26-30October 2003, pp. 34-41.

McNab, R. J. and Smith**,** L. A. (2000). Evaluation of a Melody Transcription System, *IEEE International Conference on Multimedia and Expo* (II) 2000: 819-822.

Moelants, D., Cornelis, O., Leman, M., Gansemans, J., De Caluwe, R., De Tré, G., Matthé, T. and Hallez, A. (2006). Problems and Opportunities of Applying Data- & Audio-Mining Techniques to Ethnic Music. *Proc. International Conference on Music Information Retrieval (ISMIR)*, Victoria, Canada, 8 - 12 October, pp. 334-336.

Moelants, D., Cornelis, O., Leman, Gansemans, M. J., De Caluwe, R., De Tré, G., Matthé, T. and Hallez, A. (2007). The problems and opportunities of content – based analysis and description of ethnic music, *International Journal of Intangible Heritage*, 2: 58-67.

Monti, G. and Sandler, M. (2000). Monophonic Transcription with Autocorrelation, In Proc. of the COST G-6 *Conference on Digital Audio Effects (DAFX)* (December 2000), pp. 257-260.

Nesbit, A., Hollenberg, L. and Senyard, A. (2004). Towards automatic transcription of Australian Aboriginal music, *Proc. International Conference on Music Information Retrieval (ISMIR)*, Barcelona, Spain, 10-14 October 2004, pp. 326-330.

Nettl, B., (1982). *The Study of Ethnomusicology:* Thirty-one Issues and Concept, University of Illinois Press.

Norowi, M., Doraisamy, S. and Wirza, R. (2005). Factors affecting automatic genre classification: An investigation incorporating non-western musical forms, *Proc. International Conference on Music Information Retrieval (ISMIR)*, London, UK, 11 - 15 September 2005, pp. 13-20.

Ong, B. S., Ǵomez, E. and Streich, S. (2006). Automatic extraction of musical structure using pitch class distribution features, *Proc. Workshop on Learning the Semantics of Audio Signals (LSAS)*, Athens, Greece, 6 December 2006, pp. 53–65.

Orio, N. (2010). Automatic identification of audio recordings based on statistical modeling, *Signal Processing* 90: 1064–1076.

Öztuna, Y. (2006). Makam, *Türk Musikisi: Akademik Klasik Türk San'at Musikisi'nin Ansiklopedik Sözlüğü*. II. Cilt, Ankara: Orient Yay.

Öztürk, O. M. (2006a). *Zeybek Kültürü ve Müziği*, İstanbul: Pan Yay.

Öztürk, O. M. (2006b). Benzerlikler ve farklılıklar: Bütünleşik bir “geleneksel Aanadolu müziği” yaklaşımına doğru, In *20. yıl: Pan’a armağan*, İstanbul: Pan Yay. pp. 151-188.

Paiva, R. P., Mendes, Y. and Cardoso, A. (2008) From Pitches to Notes: Creation and Segmentation of Pitch Tracks for Melody Detection in Polyphonic Audio, *Journal of New Music Research*, 37(3): 185–205.

Pollastri, E. (2002).“A Pitch Tracking System Dedicated toProcess Singing Voice for Music Retrieval”, In *Pro. IEEE Int. Conf. on Multimedia and Expo*, ICME2002.

Powers, H. S. et al. (2008). "Mode." In *Grove Music Online*. Oxford Music Online, http://www.oxfordmusiconline.com/subscriber/article/grove/music/43718pg5 (accessed November 17, 2008).

Purwins, H., B. Blankertz, and K. Obermayer, (2000). A new method for tracking modulations music in audio data format, *Proc. IEEE-INNS-ENNS* , 6 (2000) pp.270-275.

Racy, A. J. (1991) "Historical Worldviews of Early Ethnomusicologists: An East-West Encounter in Cairo, 1932," In *Ethnomusicology and Modern Music History*, eds. Stephen Blum, Philip V. Bohlman, and Daniel M. Neuman (Urbana: University of Illinois Press, 1991), 68–91.

Rao, V. and Rao, P. (2010). Vocal Melody Extraction in the Presence of Pitched Accompaniment in Polyphonic Music. *IEEE Transactions on Audio, Speech & Language Processing*, 2010: 2145~2154.

Ryynänen, M. (2006). Singing Transcription, In *Signal Processing Methods for Music Transcription*, ed: Klapuri, A., Davy, M., Springer-Verlag, New York.

Ryynänen, M. and Klapuri, A. (2004). Modelling of note events for singing transcription, in *Proc. ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing*, October 2004.

Ryynänen, M. and Klapuri, A. (2006). Transcription of the Singing Melody in Polyphonic Music, in *Proc. 7th International Conference on Music Information Retrieval (ISMIR 2006)*, Victoria, Canada, October 2006.

Ryynänen, M. and Klapuri, A. (2008). Automatic Transcription of Melody, Bass Line, and Chords in Polyphonic Music, *Computer Music Journal*, 32(3): 72-86.

Santini, S. and Jain, R. 1999. Similaity Measures, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21 (9): 871 – 883.

Shiloah, A. (2008) 'Arab music, §I, 6(ii), *Grove Music Online* ed. L. Macy (Accessed 24 February 2008), http://www.grovemusic.com

Signell, K. (1976). The Modernization Process in Two Oriental Music Cultures: Turkish and Japanese, *Asian Music*, 7(2): 72-102.

Signell, K. (2006). *Makam: Türk Sanat Musikisinde Makam Uygulaması* [Makam: Modal Practice in Turkish Art Music](trans.:İlhamiGökçen), Yapı Kredi Yayınları, İstanbul.

Sinith, M.S. and K. Rajeev, (2007). Pattern recognition in South Indian classical music using a hybrid of HMM and DTW, *IEEE Computer Society, Conf. on Computational Intelligence and Multimedia Applications*, 2 (2007) 339-343.

Stockmann, D. (1979). Die Transkription in der Musikethnologie: Geschichte, Probleme, Methoden. *Acta Musicologica*, 51(2): 204-245.

Stokes, M. (1996). History, memory and nostalgia in contemporary Turkish musicology, *Music & Anthropology*, No:1. http://www.levi.provincia.venezia.it/ma/index/number1/stokes1/st1.htm

Swain, M.J. and D.H. Ballard, (1991). Color indexing, *International Journal of Computer Vision*, 7(1): 11–32.

Tanaka, K., M. Sano, S. Ohara, M. Okudaira, (2000). A parametric template method and its application to robust matching, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1 (2000) 620-627.

Tekelioğlu, O. (2001). Modernizing Reforms and Turkish Music in the 1930s, *Turkish Studies*, 2(1): 93-109.

Temperley, D. (2001). *The Cognition of Basic Musical Structures*, MIT Press, Cambridge, Massachusetts, Chapter 7, pp.167-201.

Temperley, D. (2008). Pitch-class distribution and the identification of key, *Music Perception*, 25(3): 193-212.

Theodoridis, S. and Koutroumbas, K. (1999). *Pattern Recognition*, Academic Press.

Thomas, A. E. (2007). Intervention and reform of Arab music in 1932 and beyond 2007, *Conference on Music in the world of Islam*, Assilah, (Accessed 05 February 2008)11 http://www.mcm.asso.fr/site02/music-w-islam/articles/Thomas-2007.pdf

Toiviainen, P., and Eerola, T. (2001). A method for comparative analysis of folk music based on musical feature extraction and neural networks. In H. Lappalainen (Ed.), *Proceedings of the VII International Symposium of Systematic and Comparative Musicology and the III International Conference on Cognitive Musicology (pp. 41-45)*. Jyväskylä: University of Jyväskylä.

Toiviainen, P. and Eerola, T. (2006). Visualization in comparative music research, in: A. Rizzi and M Vichi (Ed.), COMPSTAT 2006 – *Proc. in Computational Statistics*, Heidelberg: Physica-Verlag, (2006) 209-221.

Typke, R. (2011). Note recognition from monophonic audio: a clustering approach. in: M. Detyniecki, A. García-Serrano, A. Nürnberger (Eds.): *AMR 2009*, LNCS 6535, pp. 49--58. Springer, Heidelberg (2011).

Tzanetakis, G., Kapur, A., Schloss, W.A. and Wright, M. (2007). Computational ethnomusicology. *Journal of Interdisciplinary Music Studies*, 1(2): 1-24.

Unal, E., Chew, E., Georgiou, P. G. and Narayanan, S. S. (2008). Challenging Uncertainty in Query by Humming Systems: A Fingerprinting Approach. *IEEE Transactions on Audio, Speech & Language Processing*, 2008: 359~371.

Ünlü, C. (2004). *Git Zaman Gel Zaman: fonograf-gramafon-taş plak*, İstanbul: Pan Yay.

Viitaniemi, T., Klapuri, A, and Eronen, A. (2003). A probabilistic model for the transcription of single-voice melodies, *Proceedings of the 2003 Finnish Signal Processing Symposium FINSIG'03* (2003) Issue: 20, Publisher: Citeseer, Pages: 59–63.

Wang, C.-K., R.-Y. Lyu, and Y.-C. Chiang (2003). A robust singing melody tracker using adaptive round semitones (ARS). In *Proceedings of 3rd International Symposium on Image and Signal Processing and Analysis* (ISPA03), pp. 18–20.

Wright, O. (2008). Arab Music (1-5), *Grove Music Online* ed. L. Macy (Accessed 17 February 2008) http://www.grovemusic.com

Yarman, O. (2007). A comparative evaluation of pitch notations in Turkish makam music, *Journal of Interdisciplinary Music Studies*, 1(2): 43–61.

Yarman, 0. (2008). *79-tone tuning & theory for Turkish maqam music*. PhD Thesis, İstanbul Technical University, Social Sciences Inst., İstanbul.

Yavuzoğlu, N. (2008). *21. yüzyılda Türk müziği Teorisi*, İstanbul: Pan Yay.

Yekta, R. (1997a). Ziya Gökalp Bey ve Milli Musikimiz Hakkındaki Fikirleri-I, reprinted in *Musiki Mecmuası*, 50:458.

Yekta, R. (1997b). Ziya Gökalp Bey ve Milli Musikimiz Hakkındaki Fikirleri-II, reprinted in *Musiki Mecmuası*, 50:459.

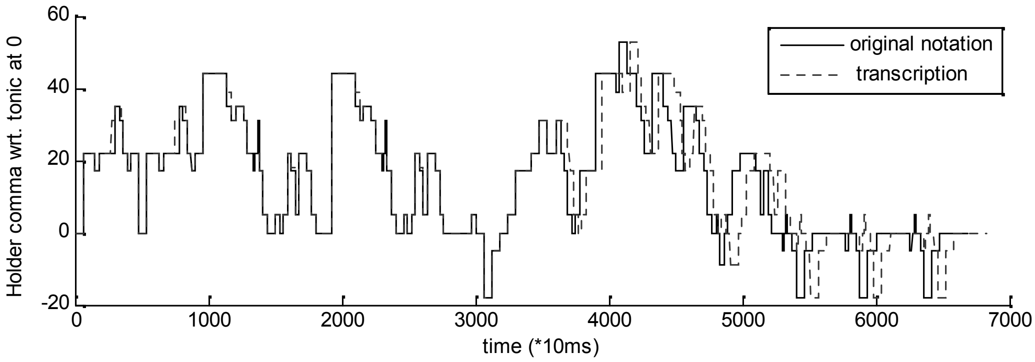
Zeren, A. (2003). *Müzik sorunlarımız üzerine araştırmalar*, İstanbul: Pan Yayıncılık. 2003.

Zhang, B. and Wang, Y. (2009). Automatic Music Transcription using Audio-Visual Fusion for Violin Practice in Home Environment, *Technical Report*, School of Computing, National University of Singapore, 2009.

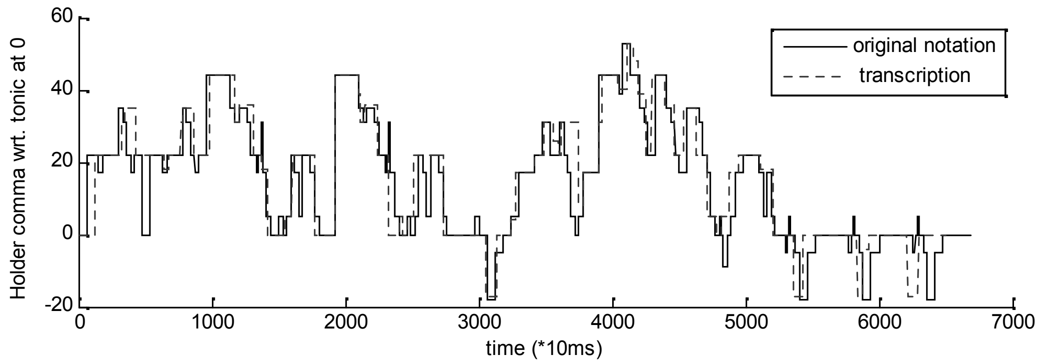
Zhu, Y. and Kankanhalli, M.S. (2006). Precise pitch profile feature extraction from musical audio for key detection, *IEEE Transactions on Multimedia*, 8 (3): 575-584.

APPENDIX A

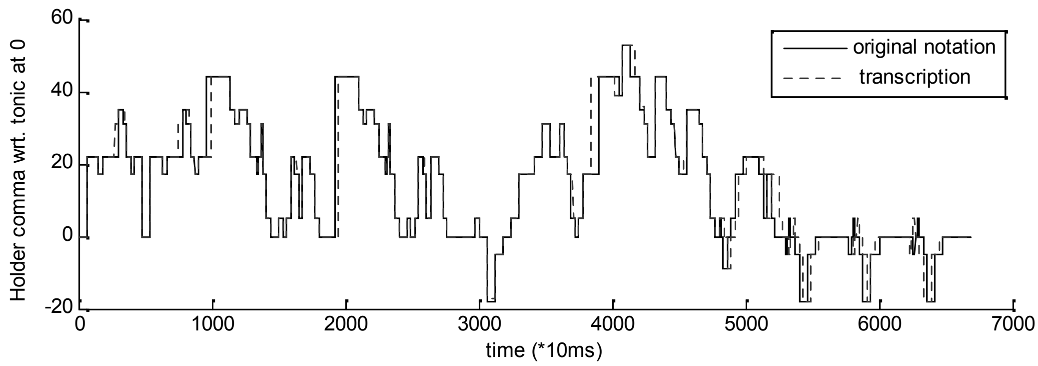
PIANO-ROLL REPRESENTATION OF TRANSCRIPTIONS



Manual transcription #1

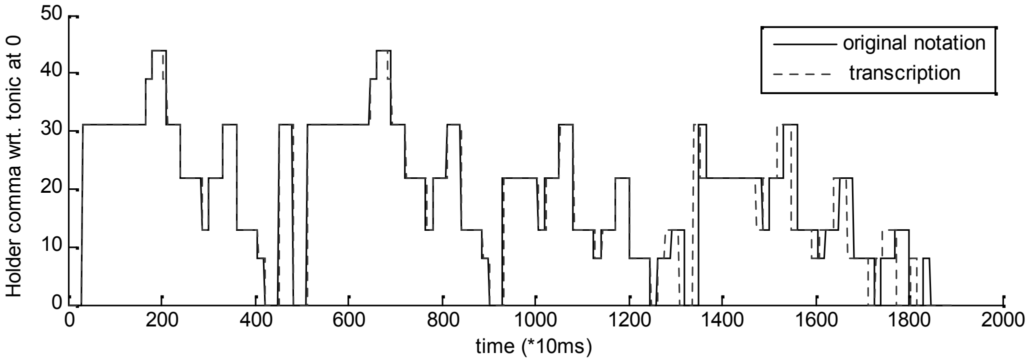


Automatic transcription

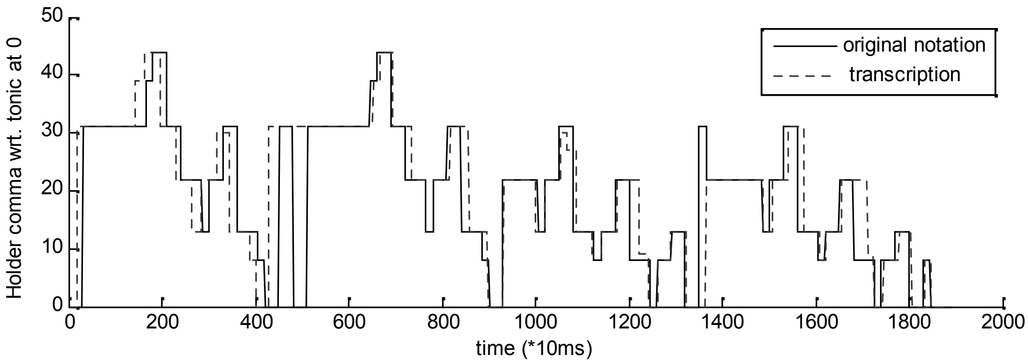


Manual transcription #2

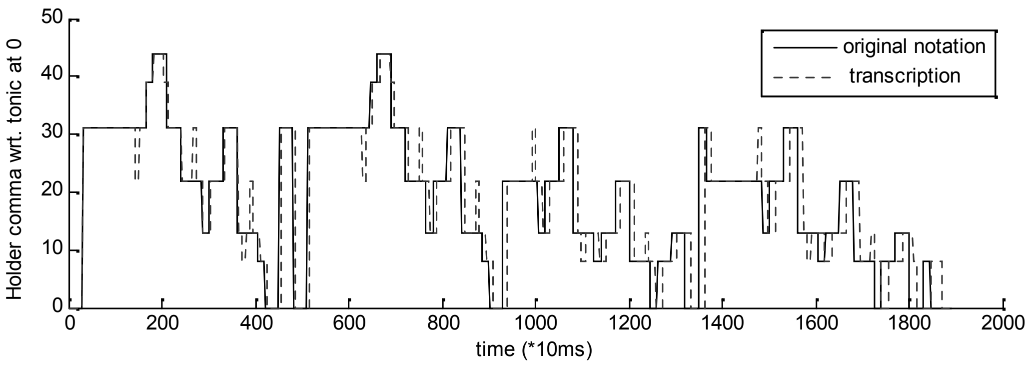
Figure A.1. Transcriptions of piece #3 *hicaz* in comparison to original notation: Manual transcription #1 (top), automatic transcription (middle) and Manual transcription #2 (bottom).



Manual transcription #1



Automatic transcription



Manual transcription #2

Figure A.2. Transcriptions of piece #4 *hüseyni* in comparison to original notation: Manual transcription #1 (top), automatic transcription (middle) and Manual transcription #2 (bottom).

APPENDIX B

STAFF NOTATION REPRESENTATION OF TRANSCRIPTIONS

B.1. Original Notations

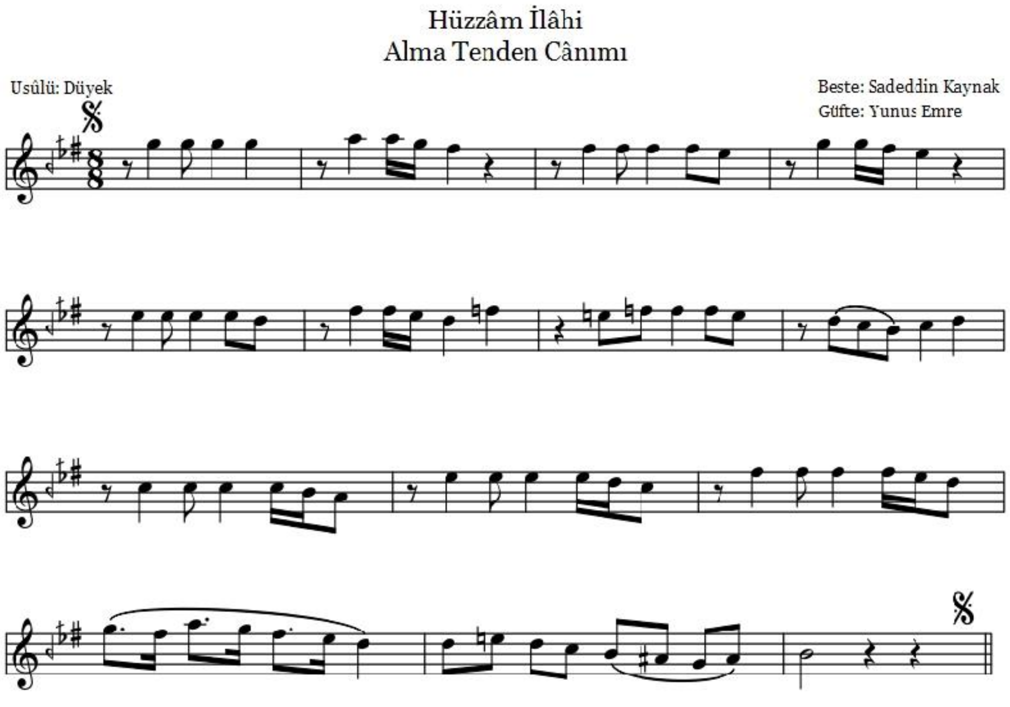


Figure B.1. Original notation of piece#1.

Uşşak İlahi

Aşkınla çak olsa bu ten

Figure B.2. Original notation of piece#2.

Figure B.3. Original notation of piece#3.

Hüseynî İlâhi
Can-ü dilde hane kıldın

Figure B.4. Original notation of piece#4.

Saba İlahi
Kıl kudumunla müşerref

Figure B.5. Original notation of piece#5.

B.2. Manual Transcriptions 1

Figure B.6. Manual Transcriptions 1 of piece#1.

Figure B.7. Manual Transcriptions 1 of piece#2.

Figure B.8. Manual Transcriptions 1 of piece#3.

Hüseyni

Sofyan

[Besteci]

Figure B.9. Manual Transcriptions 1 of piece#4.

[Başlık]

[Güfteci]

[Besteci]

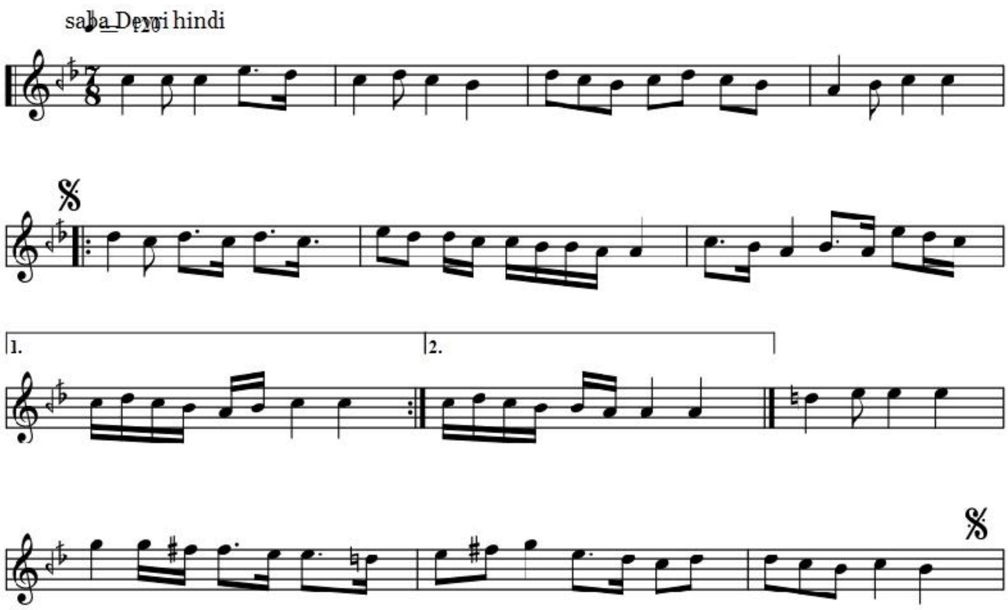


Figure B.10. Manual Transcriptions 1 of piece#5.

B.3. Manual Transcriptions 2

Figure B.11. Manual Transcriptions 2 of piece#1.

Figure B.12. Manual Transcriptions 2 of piece#2.

Hicaz

(cont. on next page)

Figure B.13. (cont.) Manual Transcriptions 2 of piece#3.

Figure B.14. Manual Transcriptions 2 of piece#4.

Saba

(e) usul: Devr-i Hindi [Besteci]

Figure B.15. Manual Transcriptions 2 of piece#5.

B.4. Automatic Transcriptions

Figure B.16. Automatic Transcription of piece#1.

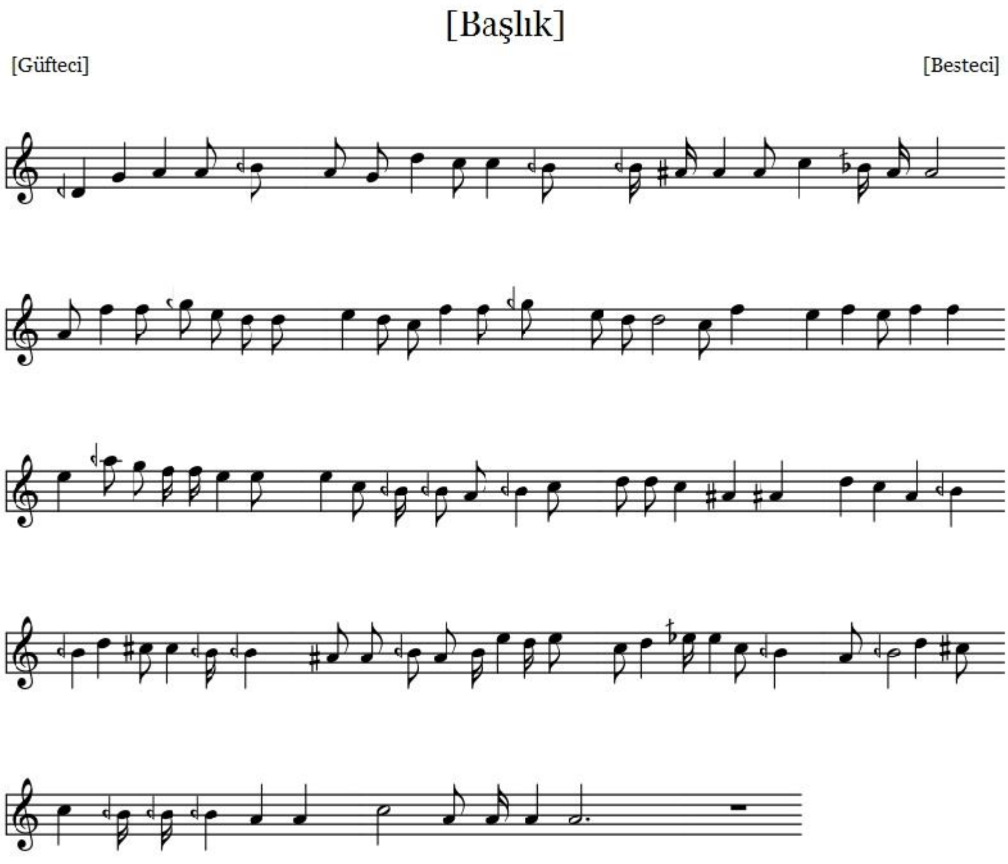


Figure B.17. Automatic Transcription of piece#2.

1

Figure B.18. Automatic Transcription of piece#3.

Figure B.19. Automatic Transcription of piece#4.

Figure B.20. Automatic Transcription of piece#5.

VITA

Education

PhD., Department of Electrical-Electronics Engineering, İzmir Institute of Technology , İzmir, 2007- .

MSc., Department of Musicology, Dokuz Eylül Üniversitesi(DEÜ), İzmir, 2007.

BSc., Department of Electrical-Electronics Engineering, Hacettepe Üniversitesi, Ankara, 1996.

Academic Employment

2011- , Lecturer, Department of Musicology, Dokuz Eylül Üniversitesi(DEÜ)

2007-2010, Scholarship from TÜBİTAK

2005-2008, Lecturer (Part-time), İzmir Ekonomi Üniversitesi, Department of Communication Sciences, İzmir.

2004-2007 Lecturer (Part-time), Dokuz Eylül Üniversitesi, Department of Musicology, İzmir.

Publications of Thesis

Refereed Journal

Gedik, A. C. and Bozkurt, B., 2010, Pitch frequency histogram based music information retrieval for Turkish music, *Signal Processing*, 90(4): 1049-1063.

Gedik, A. C. and Bozkurt, B. , 2009, Evaluation of the Makam Scale Theory of Arel for Music Information Retrieval on Traditional Turkish Art Music, *Journal of New Music Research*, 38(2): 103-116.

Holzapfel, A., Stylianou, Y., Gedik, A.C., Bozkurt, B., 2010, Three dimensions of pitched instrument onset detection, *IEEE Transactios on Audio, Speech and Language Processing*, 18(6):1517-1527

National and International Refereed Symposium and Conference

Bozkurt, B. and Gedik, A.C., 2011, Türk Müziği İçin Bir Frekans Analiz Aracı, Ulusal Türk Müziği Kurultayı, İstanbul Teknik Üniversitesi, 6-7-8 Nisan

B. Bozkurt, A. C. Gedik, M.K.Karaosmanoglu, 2011, "Klasik Türk Müziği İçin Otomatik Notaya Dökme Sistemi", Proc. SIU, Sinyal İşleme Uygulamaları, Antalya. IEEE 19. Sinyal İşleme ve İletişim Uygulamaları Kurultayı - SİU 2011

Gedik, A.C., Bozkurt, B. and Cirak, C., 2009, A study of fret positions of tanbur based on automatic estimates from audio recordings, CIM09, *5th Conference on Interdisciplinary Musicology*, Paris, France, 26-29 October. http://cim09.lam.jussieu.fr/CIM09-en/Proceedings_files/18A-GedikBozkurtCirak.pdf

Bozkurt, B. and Gedik, A.C., 2009, Turkish Music Information Retrieval: problems, proposed solutions and tools, Proc. *IEEE 17th Signal Processing and Communications Applications Conference* (SIU-2009).

Gedik, A. C. and B.Bozkurt, 2008, " Automatic Classification of Taksim Recordings in Turkish Makam Music", Conference on Interdisciplinary Musicology, 2-6 July 2008, Thessaloniki/Greece.

Personal

Date of Birth: 23.05.1972

Date of Place: Turkey