Contents lists available at ScienceDirect

Signal Processing

Pitch-frequency histogram-based music information retrieval for Turkish music

Ali C. Gedik *, Barış Bozkurt

Department of Electrical and Electronics Engineering, Izmir Institute of Technology, Gülbahçe, Urla, İzmir, Turkey

ABSTRACT

This study reviews the use of pitch histograms in music information retrieval studies for western and non-western music. The problems in applying the pitch-class histogram-based methods developed for western music to non-western music and specifically to Turkish music are discussed in detail. The main problems are the assumptions used to reduce the dimension of the pitch histogram space, such as, mapping to a low and fixed dimensional pitch-class space, the hard-coded use of western music theory, the use of the standard diapason (A4 = 440 Hz), analysis based on tonality and tempered tuning. We argue that it is more appropriate to use higher dimensional pitch-frequency histograms without such assumptions for Turkish music. We show in two applications, automatic tonic detection and *makam* recognition, that high dimensional pitch-frequency histogram representations can be successfully used in Music Information Retrieval (MIR) applications without such pre-assumptions, using the data-driven models.

1. Introduction

Traditional musics of wide geographical regions, such as Asia and Middle East, share a common musical feature, namely the modal system. In contrast to the tonal system of western music, the modal systems of these non-western musics cannot be only described by scale types such as major and minor scales. 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 [1]. In this sense, the *makam* practice in Turkey—we prefer to refer to it as Turkish music—as a modal system, is close to the melody-type,

and thus shares many similarities with *maqam* in the Middle East.

The traditional art musics of the Middle East have been practiced for hundreds of years, and their theoretical written sources date back to as early as Al-Farabi (c. 872–951). They are both influenced by and influence the folk, art and religious music of a culture, and more and more the modern musical styles: the traditional musical elements are increasingly finding their way into modern popular music idioms. For example, the representative instruments of the genre, such as the *ud* and *ney*, performing in a jazz quartet, in front of a symphony ensemble or together with popular music bands performing pop, rock, hip-hop, etc. Many western-style modern music conservatoires were founded for the education of these traditional musics by the middle of 20th century, and recently, the number of students graduating from such institutions has been increasing.

Although Music Information Retrieval (MIR) on western music has become a well-established research domain (for a comprehensive state-of-art on MIR, see

* Corresponding author. Tel./fax: +90 232 2323423.
  E-mail address: a.cenkgedik@musicstudies.org (A.C. Gedik).
  URL: http://www.musicstudies.org/ali.html (A.C. Gedik).

[2]), its applications on non-western musics are still very limited. Although there are a large number of non-western performers, listeners and a long history of non-western musics, the MIR research for non-western musics is in its early stages (a comprehensive review of "computational ethnomusicology" is presented in [3]). There is an increasing need for MIR methods for non-western musics as the databases are enlarging through the addition of both new recordings as well as re-mastered and digitized old recordings of non-western musics.

Moelants et al. [4,5] briefly introduced the main problems regarding the application of the current MIR methods developed for western music to non-western musics. In their study, these problems are presented by summarizing their research on traditional musics of Central-Africa, and these problems also hold true for Turkish music. The most important problem is the representation of data due to the different pitch spaces in African music and western music. There is no fixed tuning and relative pitch is more important than the standard pitches in African music. Pitches demonstrate a distributional characteristic and performance of pitch intervals are variable. There are also problems related to the representation of pitch space within one octave. Due to these problems, pitches are represented in a continuous pitch space in contrast to discrete pitch space representation in western music with 12 pitch-classes.

The second problem, which is not less important, is the lack of a reliable music theory for non-western musics. In this sense, the traditional art musics of the Middle East share another commonality, namely the divergence between theory and practice [8,56]. This divergence arises from the fact that the oral tradition dominated the traditional art musics of the Middle East, as well as the African music as stated [5]. This lack of music theory is not explicitly mentioned by Moelants et al. [4,5]. However, the problem is implicit in their approach to African music: their methods rely only on audio recordings in contrast to the studies on western music, where both music theory and symbolic data play crucial roles in developing MIR methods.

Turkish music theory consists of mainly descriptive information. This theory, as taught in conservatories, books, etc., is composed of a miscellany of melodic rules for composition and improvisation. These rules contain the ascending or descending characteristics of the melody, functions of the degrees of the scale(s), microtonal nuances, and possible modulations. The tuning theory has some mathematical basis but is open to discussion: there is no standardization of the pitch frequencies accepted to be true for most of the musicians. For example, in Turkish music it is still an open debate how many pitches per octave (propositions vary from 17 to 79 [7]) are necessary to conform to musical practice. It is generally accepted that the tuning system is non-tempered, consisting of unequal intervals unlike western music. For these reasons, the divergence between theory and practice is an explicit "problem" in Turkish music. There have also been attempts of westernization and/or nationalization of tuning theories in Turkey [8] which add

further complexities in music analysis. With the large diversity of tuning systems (a review can be found in [7]) and a large collection of descriptive information, the theory is rather confusing. For this reason, one is naturally inclined to 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.

For these reasons, 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.

We first show that pitch-frequency histograms can be effectively used to represent the pitch space characteristics of Turkish music and be processed to achieve the above-mentioned goals. Some of the possible MIR applications based on the tonic detection and *makam* recognition methods we present are as follows: automatic music transcription (*makam* recognition and tonic detection are crucial in transcription), information retrieval (retrieving recordings of a specific *makam* from a database), automatic transposition to a given reference frequency (which facilitates rehearsal of pieces with fixed pitched instruments such as Ney), automatic tuning analysis of large databases by aligning recordings with respect to their tonics as presented in [6].

The use of pitch histograms is not a new issue in MIR, but it needs to be reconsidered when taking into account the pitch space characteristics of Turkish music. More specifically, it is well known that pitches do not correspond to standardized single fixed frequencies. There is a dozen of possible standard pitches/diapason (called "*ahenk*"), and tuning can still be out of these standard pitches (especially for string instruments). In addition, there are many *makam* types in use with a large variety of intervals included, and musicians may have their own choices for the performance of certain pitches. Defining appropriate fret locations for fretted instruments is an open topic of research. It has been observed that the number of frets and their locations vary regionally or depend on the choice of performers.

As a result, the main contribution of the study is the presentation of a framework for MIR applications on Turkish music with a comprehensive review of the pitch histogram-based MIR literature on western and non-western musics; a framework without which possible MIR applications cannot be developed. This framework is also potentially applicable for other non-western musical traditions sharing similarities with Turkish music.

In the following sections of this paper, we present the main contribution in detail listed as: *First, a review of pitch histogram use in MIR studies both for western and non-western music in comparison with Turkish music is presented.* Then we discuss more specifically the use of pitch histograms in Turkish music analysis. Following this review part, we present the MIR methods we have

developed for the tasks mentioned above. In [6] we have presented an automatic tonic detection algorithm. In Section 3 of this paper, *as the second part of the contributions, we provide new detailed tests for evaluating the algorithm using a large corpus containing both synthetic and real audio data* (*which was lacking in* [6]). *We further investigate the possibility of using other distance measures in the actual algorithm. We show that the algorithm is improved when City-Block or Intersection distance measures are used instead of cross-correlation* (*as used in* [6]). In Section 4, *as the last part of the contributions, a simple makam classifier design is explained again using the same pitch histogram representations and methodology for matching given histograms.* The final section is dedicated to discussions and future work. The relevant database and MATLAB codes can be found at the project web site: ⟨http://likya.iyte.edu.tr/eee/labs/audio/Main.html⟩.

2. 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.

In order to clarify the review of current literature in comparison with Turkish music, a brief description of Turkish music and our study should be introduced. Turkish music is mainly characterized by modal entities called *makam* and each musical piece is identified and recognized by a *makam* type. A set of musical rules for each *makam* (*makamlar*, pl.) type are loosely defined in music theory, and these rules roughly determine the scale of a *makam* type. Although it is recorded that there were 600 *makam* types, only around 30 of them are currently used in Turkey. Each *makam* type also holds a distinct name representing the *makam* type.

We first present an appropriate representation of Turkish music. Musical data are represented by pitch histograms constructed based on fundamental frequency (f0). f0 data are extracted from monophonic audio recordings. Thus, we apply methods based on pitch histograms. Second, necessary methods to process such representation are presented. Third, automatic recognition of Turkish audio recordings by *makam* types (names) is presented. In brief, our research problem can be expressed as finding the *makam* of a given musical piece.

Pitch-class histogram versus pitch-frequency histogram: 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 Fig. 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 two tonalities, major and minor
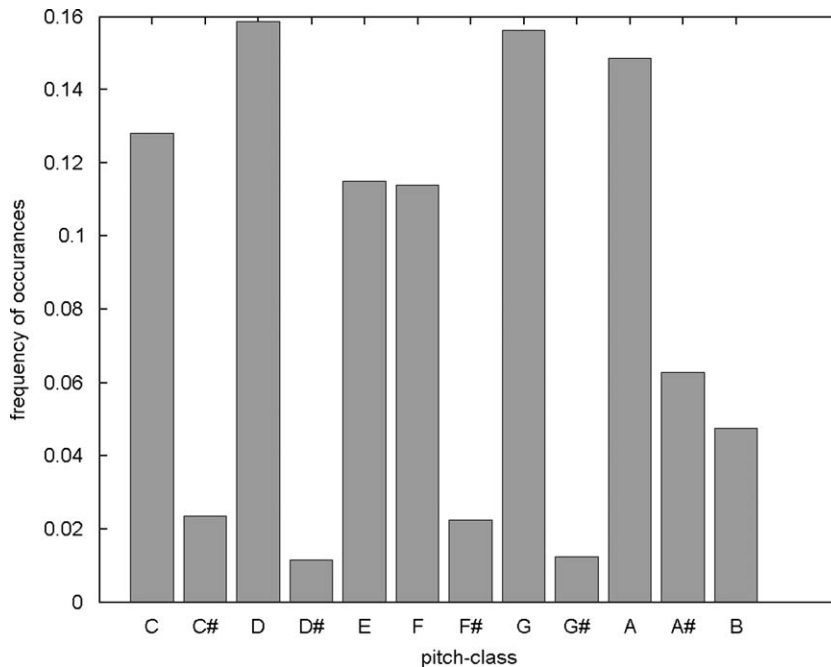

Fig. 1. Pitch histogram of J.S. Bach's C-major Prelude from *Wohltemperierte Klavier II* (BWV 870).

templates, and the tonality whose template is more similar is found as the tonality of the musical piece.

However, as mentioned in the Introduction, there are important differences in pitch spaces between western and Turkish music which can be simply observed by comparing the pitch histogram examples from western music and Turkish music as shown in Figs. 1 and 2, respectively.

Fig. 2 presents pitch histograms of two musical pieces from the same *makam* performed by two outstanding performers of Turkish music. 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 distributional quality in contrast to western music where pitches are performed in fixed frequency values. Although the two pieces belong to the same *makam*, the performers prefer close but different pitch intervals for the same pitches. The two histograms in Fig. 2 are aligned according to their tonics in order to compare the intervals visually. The tonic frequencies of the two performances are computed as 295 and 404 Hz, hence they are not in a standard pitch. This is an additional difficulty/difference 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. It is neither straightforward to define a set of pitch-classes for Turkish music nor represent pitch histograms by 12 pitch-classes as in western music.

Despite the differences in pitch spaces between western and Turkish music, 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 following subsection, the state-of-art of relevant MIR studies on non-western musics is reviewed.

2.1. 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 [9]. 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 [10] and even the problems of monophonic transcription (especially for some signals like singing) still have not been fully solved [11]. As a result, most of the literature on pitch histograms consists of methods based on symbolic data, and these 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
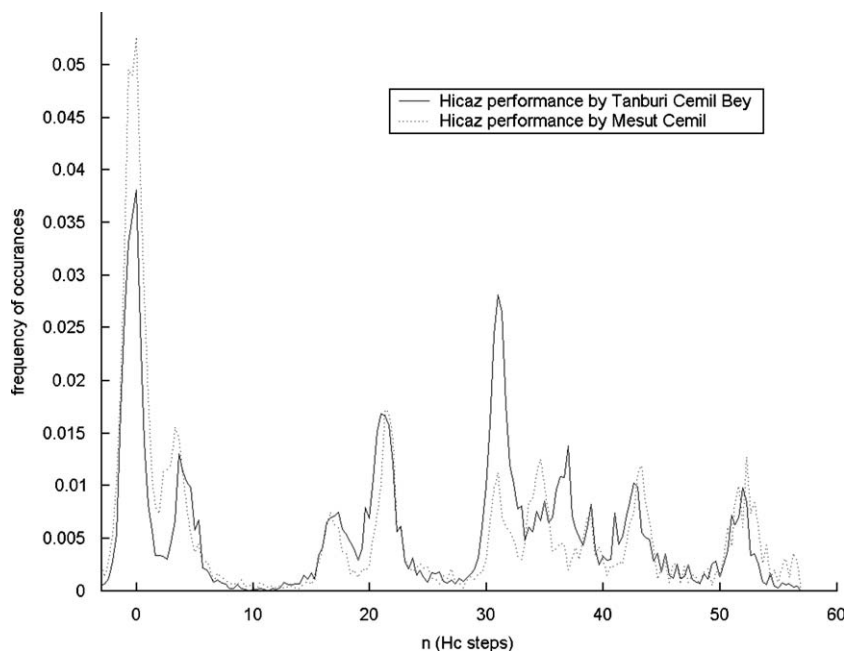

Fig. 2. Pitch-frequency histogram of *hicaz* performances by Tanburi Cemil Bey and Mesut Cemil.

tonalities, the templates of other tonalities are found simply by transposing the templates to the relevant keys [12].

The construction of the tonality templates is mainly based on three kinds of models: music theoretical (e.g. [13]), psychological (e.g. [14]) and data-driven models (e.g. [15]). 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 [16] and the geometrical approach of Chew [17].

Among these models, the psychological model of Krumhansl and Kessler [14] 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 [14]:

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2(y - \bar{y})^2}} \qquad (1)$$

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 (1) with different templates (y) and the template which gives the highest coefficient is found as corresponding to the tonality of the piece.

The same method is applied in data-driven models (e.g. [15]) also 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. [18]) 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. [18,19]). 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. [20,21]). On the other hand, few studies estimate first a tuning frequency, considering the fact that recordings of various bands and musicians need not be tuned exactly to 440 Hz. However, even in these studies, 440 Hz is taken as a ground truth in another fashion [18,22]. 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 Eq. (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 [23]. Chuan and Chew [9] and Lee and Slaney [24] do not use templates, but their approach is based on audio data synthesized from symbolic data (MIDI). Liu et al. [25] 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.

2.2. 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 [26–29]. This is also the case for the two studies on African music [30] and Javanese music [31]. There are also two MIR applications for non-western music without using pitch histograms: an automatic transcription of Aboriginal music [32] and the pattern recognition methods applied on South Indian classical music [33]. Here, we will only review studies based on pitch histograms and refer the reader to Tzenatakis et al. [3] 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. [4,5] was mentioned in the Introduction which applies 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 [34] 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. In [6], we presented for the first time the necessary tools and methods for the pitch space analysis of Turkish music when applied to large music databases. This method is further summarized and extended in Section 3.2.

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. [35] 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 [36,37] 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. In [36] pitch-class dyad histograms are also used 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 h 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 parameters for evaluation. Another study [37] 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 [36]. 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 [38]. A new feature, harmonic pitch class profile (HPCP) proposed by Gomez [19] 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, hip-hop, 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 than 80% are obtained in terms of *F*-measure. However, the study also bears serious drawbacks as explicitly demonstrated by Lartillot et al. [39]. 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 [40] apply SOM to visualize 2240 Chinese, 2323 Hungarian, 6236 German and 8613 Finnish folk melodies. Chordia and Rae [41] also apply SOM to model tonality in North Indian Classical Music.

In a recent study, Gedik and Bozkurt [8], have considered the classification of Turkish music recordings by *makam* types from audio recordings for the first time in the literature. However, this first study was not aiming at a fully automatic classification in terms of *makam* types. It applied the MIR methods to evaluate the divergence of theory and practice in Turkish music. One hundred and

eighty audio recordings, with automatically labeled and manually checked tonics, were used to classify recordings by nine *makam* types. Each recording was represented as pitch-frequency histograms, and templates for each *makam* type are constructed according to the pitch-scale definitions for each *makam* type defined by the most influential theorists, Hüseyin Saadettin Arel (1880–1955). Although the theory presents fixed frequency values for pitch intervals, each pitch is represented as a Gaussian distribution for each *makam* type in order to compare the theory with practice. As a result, it has been shown that the theory is found more successful for the definitions of some *makam* types than for others. It was also shown that pitch-frequency histograms are potentially a good representation of the pitch space and can be successfully used in MIR applications.

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. 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 note frequencies for Turkish music [34,42–44]. It is only very recent that a few studies that aim at designing automatic analysis methods based on processing of pitch-frequency histograms [6,8].

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, and without folding the data into a single octave. We present the methods developed below.

3.1. Pitch-histogram computation

For fundamental frequency (f0) analysis of the audio data, the YIN algorithm [45] is used together with some post-filters. The post-filters are explained in [6] and are mainly designed to correct octave errors and remove noise on the f0 data.

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:

$$Hf_0[n] = \sum_{k=1}^{K} m_k$$
$$m_k = 1, \quad f_n \leq f_0[k] < f_{n+1}$$
$$m_k = 0 \quad \text{otherwise} \tag{2}$$

where (f_n, f_{n+1}) are boundary values defining the f0 range for the nth 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

$$W_b = \frac{\log_2(f_{0max}) - \log_2(f_{0min})}{N}$$
$$f_n = 2^{f_{0min}+(n-1)W_b} \tag{3}$$

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, $\frac{1}{3}$ Holderian comma resolution is obtained by Bozkurt [6]. 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 [7].

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

3.2. 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 [46]. However, tracking the last note reliably is difficult especially in old recordings where the energy of background noise is comparatively high.

3.2.1. The main algorithm flow

In [6], we presented a robust tonic detection algorithm (shown in Fig. 3) based on aligning the pitch histogram of
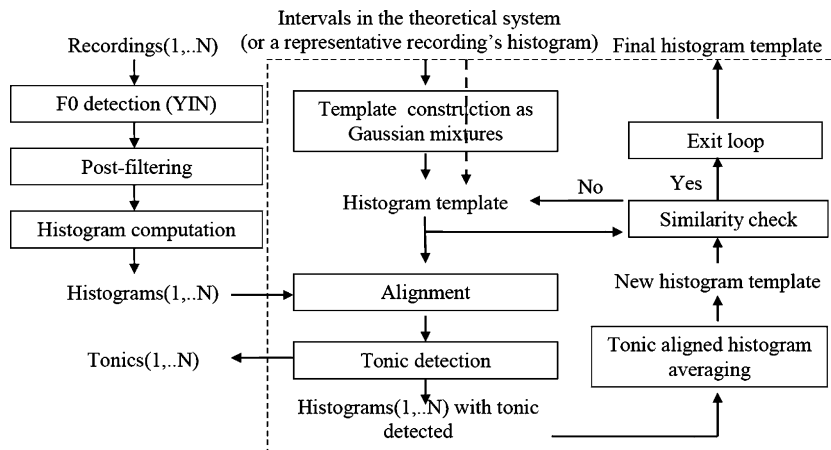
Fig. 3. Tonic detection and histogram template construction algorithm (box indicated with dashed lines) and the overall analysis process. All recordings should be in a given *makam* which also specifies the intervals in the theoretical system.

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 Fig. 4 of [6]. 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 1–4 Hc range. As expected, smaller widths are observed on fretted instrument samples, whereas 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 1.5–3.5 Hc 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 1.5–3.5 Hc 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.

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 $\frac{1}{3}$ Hc steps (as shown in Fig. 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 Fig. 4b by indicating the tonic with 0 Hc) 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 Fig. 3.

In [6], the best matching point between histograms was found by finding the maximum of the cross-correlation function, $c[n]$, computed using the equation

$$c[n] = \frac{1}{K}\sum_{k=0}^{K-1} h_r[k]h_t[n+k] \tag{4}$$

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

In this section, we have reviewed the previously presented automatic detection algorithm [6]. The results obtained in [6] were quite convincing, but based on subjective evaluation on tonic detection of 67 solo recordings of Tanburi Cemil Bey (1871–1916). However, it is an open issue whether the use of other distance measures can provide better results than the cross-correlation measure (used in [6]), which could be studied objectively using synthetic test data. As part of the

a

b
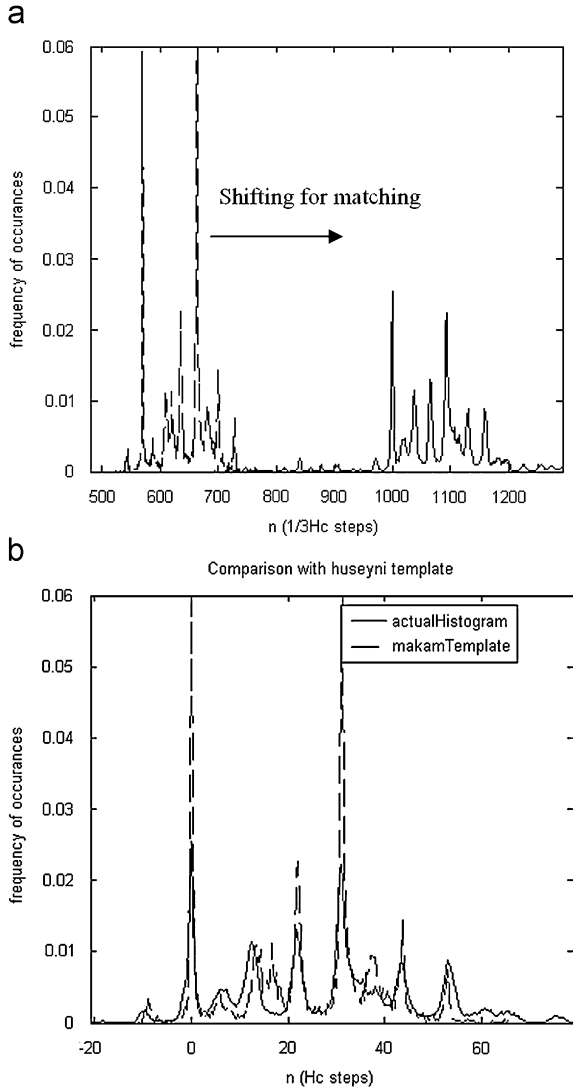
Fig. 4. Tonic detection via histogram matching. (a) Template histogram is shifted and the distance/correlation is computed at each step and (b) matching histograms at the shift value providing the smallest distance (normalized for viewing, tonic peak is labeled as the 0 Hc point).

contributions of this study, we test four other distance/similarity measures for matching histograms. We enlarge the test set to a total of 268 samples where 150 of the examples are synthetic. For the synthetic data set, quantitative results are provided. It is well known in a synthetic data set the tonic frequency for the set of recordings in a specific *makam* is the same (synthesized using the same tuning frequency). Therefore, the standard deviation of measured tonic frequencies and the maximum distance to the mean of the estimates can be used as a quantitative measure of consistency/reliability of the method. These contributions of this study to the algorithm in [6] are presented in Sections 3.2.2 and 3.2.3.

3.2.2. Additional distance measures for matching histograms

There is a large body of research on the distance measures between two histograms in the domain of image indexing and retrieval, pattern recognition, clustering, etc. (the interested reader is referred to [47] for a review). Here, we discuss a number of popular distance measures which we think are relevant to our problem.

City Block (L$_1$-norm):

$$d[n] = \frac{1}{K}\sum_{k=0}^{K-1} |h_r[k] - h_t[n + k]| \qquad (5)$$

Euclidean (L$_2$-norm):

$$d[n] = \sqrt{\sum_{k=0}^{K-1} (h_r[k] - h_t[n + k])^2} \qquad (6)$$

Intersection:

$$d[n] = \frac{1}{K}\sum_{k=0}^{K-1} \min(h_r[k], h_t[n + k]) \qquad (7)$$

These three measures are used for histogram-based image indexing and retrieval [48]. In addition, we include a distance measure from the set of popular distance measures defined for comparing probability density functions.

Bhattacharya distance:

$$d[n] = -\log \sum_{k=0}^{K-1} \sqrt{h_r[k]h_t[n + k]} \qquad (8)$$

These four additional measures are integrated in the tonic detection algorithm, and a total of five measures are compared in controlled tests as explained below.

3.2.3. Tests

The tests are performed on two groups of data: synthetic audio and real audio.

Tests on synthetic audio: Using synthetic audio, we have the chance for handling controlled tests. A total of 150 audio files are gathered from the teaching materials distributed as notation plus audio-CD format as "*makam* compilations", each track synthesized from a specific type of MIDI data for Turkish music using a string instrument, a wind instrument and a percussive instrument altogether as trio [49]. All the recordings in a given *makam* are synthesized at the same standard pitch (diaposon); therefore, the tonic frequencies are all the same, but unknown to the authors of this manuscript. This gives us the opportunity to compare tonic frequencies for each *makam* class with respect to the mean estimates. The *makam* histogram templates are computed from hundreds of real audio files using the algorithm in [6]. In Table 1, we present the results obtained by using these templates for tonic detection of the 150 synthetic audio files.

A close look at the values for a given *makam* indicates that most of the time the estimations are very consistent. For example, for *makam rast*, the mean value of the tonics in 24 files are found in the range of 109.51–110.05 Hz using five different measures with a maximum standard deviation of 0.42 Hz (for the cross-correlation methods). A visual check of the spectrum for the first harmonic peak location of the last note of the recordings reveals that the tonic is around 110 Hz.

Table 1
Automatic tonic detection results on synthetic audio files.

	CrossCorr.	City Block	Euclidean	Intersection	Bhattacharyya
RAST (24 *songs*)					
Mean tonic f0 (Hz)	110.05	109.59	109.51	109.59	109.61
Std (Hz)	0.42	0.16	0.19	0.16	0.17
MaxDist (Hz)	1.12	0.26	0.34	0.26	0.27
Std (Cents)	6.52	2.54	2.97	2.54	2.62
MaxDist (Cents)	17.49	4.03	5.29	4.03	4.32
SEGAH (22 *songs*)					
Mean tonic f0(Hz)	276.62	274.71	274.44	274.71	274.16
Std (Hz)	0.48	0.6	0.5	0.6	0.43
MaxDist (Hz)	1.18	1.27	0.95	1.27	1.23
Std (Cents)	3.01	3.8	3.12	3.8	2.72
MaxDist (Cents)	7.38	7.96	6	7.96	7.72
HUZZAM (16 *songs*)					
Mean tonic f0 (Hz)	138.17	136.48	136.37	136.48	136.37
Std (Hz)	0.69	0.19	0.21	0.19	0.2
MaxDist (Hz)	1.17	0.38	0.34	0.38	0.32
Std (Cents)	8.64	2.37	2.61	2.37	2.5
MaxDist (Cents)	14.65	4.78	4.28	4.78	4.12
SABA (23 *songs*)					
Mean tonic f0 (Hz)	123.72	123.32	123.39	123.32	123.37
Std (Hz)	0.49	0.22	0.17	0.22	0.17
MaxDist (Hz)	1.08	0.39	0.34	0.39	0.32
Std (Cents)	6.84	3.08	2.33	3.08	2.42
MaxDist (Cents)	15.05	5.41	4.78	5.41	4.46
HICAZ (19 *songs*)					
Mean tonic f0 (Hz)	132.16	123.19	131.93	123.19	123.1
Std (Hz)	20.81	0.21	20.83	0.21	0.15
MaxDist (Hz)	52.59	0.44	52.82	0.44	0.29
Std (Cents)	253.14	2.93	253.75	2.93	2.15
MaxDist (Cents)	579.9	6.22	583.01	6.22	4.13
HUSEYNI (23 *songs*)					
Mean tonic f0 (Hz)	123.54	123.54	123.52	123.54	123.56
Std (Hz)	0.32	0.12	0.13	0.12	0.14
MaxDist (Hz)	1.19	0.21	0.33	0.21	0.33
Std (Cents)	4.49	1.71	1.89	1.71	1.89
MaxDist (Cents)	16.54	3	4.56	3	4.57
USSAK (23 *songs*)					
Mean tonic f0 (Hz)	125.06	123.35	125.1	123.35	123.28
Std (Hz)	8.37	0.2	8.64	0.2	0.19
MaxDist (Hz)	38.25	0.38	39.64	0.38	0.36
Std (Cents)	112.1	2.85	115.69	2.85	2.64
MaxDist (Cents)	461.95	5.3	476.55	5.3	5.05
Mean-Stds (Cents)	56.39	2.75	54.62	2.75	2.42
Max-MaxDist (Cents)	579.9	7.96	583.01	7.96	7.72
♯ of false peak detections	4	0	4	0	0

The last row of the table lists the number of false peak detections: cross-correlation and Euclidean measures resulted in labeling wrong histogram peaks as a tonic in only four files over 150 files. This results in very large maximum-distance and standard deviation values for these two methods for *hicaz* and *ușșak makams*. The other three methods, City Block, Intersection and Bhattacharyya, all picked correct tonic peaks for all files. The results for these three methods are very consistent, and have very low variation within a recording of the same *makam* synthesized using in the same diapason (*ahenk*). Considering the fact that the templates used for alignment were constructed from real audio files, the results are surprisingly good.

The little variations in tonic estimation are mainly due to computing the tonic from the shift value and the template's tonic location. They can easily be removed by an additional step of computing the tonic as the center of gravity of the peak lobe labeled as tonic. Real audio tests are handled with this additional step included in the algorithm.

Tests on real audio: As an example of real audio data, we have chosen solo improvisation (*taksim*) recordings from musicians referred as "indisputable masters" in literature: Tanburi Cemil (tanbur, *kemençe*, violoncello), Mesut Cemil (*tanbur*, violoncello), Ercüment Batanay (*tanbur*), Fahrettin Çimenli (*tanbur*), Udi Hrant (violin), Yorgo Bacanos (*ud*), Aka Gündüz Kutbay (*ney*), Kani Karaca (vocal), Bekir Sıdkı

Sezgin (vocal), Necdet Yaşar (*tanbur*), İhsan Özgen (*kemençe*), Niyazi Sayın (*ney*). The earliest recordings are those of Tanburi Cemil dating from 1910 to 1914, and the most recent are those of Niyazi Sayın dating from 2001 (Sada: Niyazi Sayın. Mega Müzik-İstanbul: 2001).

The database is composed of 118 recordings of different types of *makams*: *rast* (15 files), *segah* (15 files), *hüzzam* (13 files), *saba* (11 files), *hicaz* (13 files), *hüseyni* (11 files), *uşşak* (11 files), *kürdili hicazkar* (17 files), *nihavend* (12 files). Again, the same *makam* histogram templates used in the synthetic audio tests are used for matching.

For each file, the tonic is found, and a figure is created with the tonic indicated, and the template histogram is finally matched to the recording's histogram to check the result visually manually. This check is quite reliable since almost all histogram matches are as clear as in Fig. 4b. For confusing figures, we referred to the recording and compared the f0 estimate of the recording's last note with the estimated tonic frequency.

Indeed, it was observed that the tonic re-computation from the corresponding peak removed the variance within the various methods (as expected). In 118 files, cross-correlation and Euclidean methods failed for only one (the same) file in *makam rast*, and City Block and Intersection methods failed for one (the same) file in *makam uşşak* and Bhattacharyya failed for two files, one in *makam uşşak*, the other in *makam rast*.

As a result, the City Block and Intersection measures were found to be extremely successful: they failed only in one file among 268 files (the synthetic data set plus the real data set) in addition to their computationally lower cost comparatively. Other measures are also quite successful: Bhattacharyya failed on two, cross-correlation and Euclidean failed on five of 268 files. These results indicate that the pitch histogram representation carries almost all of the information necessary for tonic finding and the task can be achieved via a simple "shift and compare" approach.

3.3. Automatic makam recognition

In pattern recognition literature, template matching method is a simple and robust approach when adequately applied [47,50–53]. Temperley [12] 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 $\frac{1}{3}$ Hc ($\frac{1}{159}$ 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 [54]. The list can be extended as new templates are included which can be computed in a fully automatic manner using the algorithm described in [6].
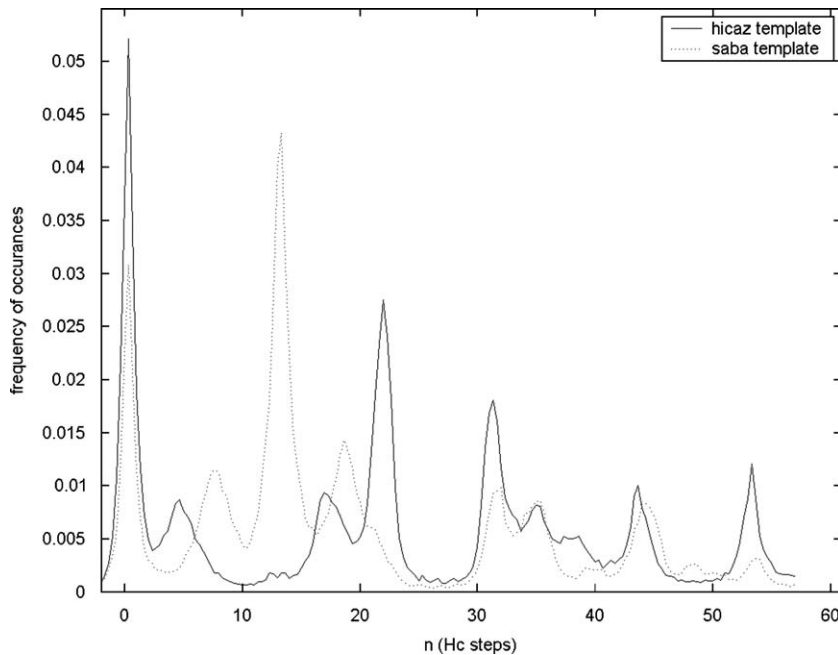

Fig. 5. Pitch-frequency histogram templates for the two types of melodies: *hicaz makam* and *saba makam*.

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—*uşşak*.

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 Fig. 3). 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 obtained by the procedure defined in Section 3.1 is normalized to unity sum as follows:

$$Hf_0N = \frac{Hf_0}{\sum Hf_0} \tag{9}$$

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:

$$T_k = \sum_{i=1}^{N} Hf_0N_k(i) \tag{10}$$

$$TN_k = \frac{T_k}{\sum T_k} \tag{11}$$

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 Fig. 5 as an example.

Finally, when the data-driven model is finished, the similarity between templates (see Fig. 5) and a recording can be measured. As the City-Block (L1 norm) distance measure performed best in the automatic tonic detection algorithm, it is also used in the *makam* recognition system. Given a recording's histogram and the templates, the histogram is shifted in $\frac{1}{3}$ Hc 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.

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

$$recall = \frac{TP}{TP + FN}, \quad precision = \frac{TP}{TP + FP}$$
$$F\text{-}measure = \frac{2 \cdot recall \cdot precision}{(recall + precision)} \tag{12}$$

TP: true positive, *TN*: true negative, *FP*: false positive, *FN*: false negative.

The success rates obtained are presented in Table 2.

Table 2 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 *uşşak*. Table 3 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.
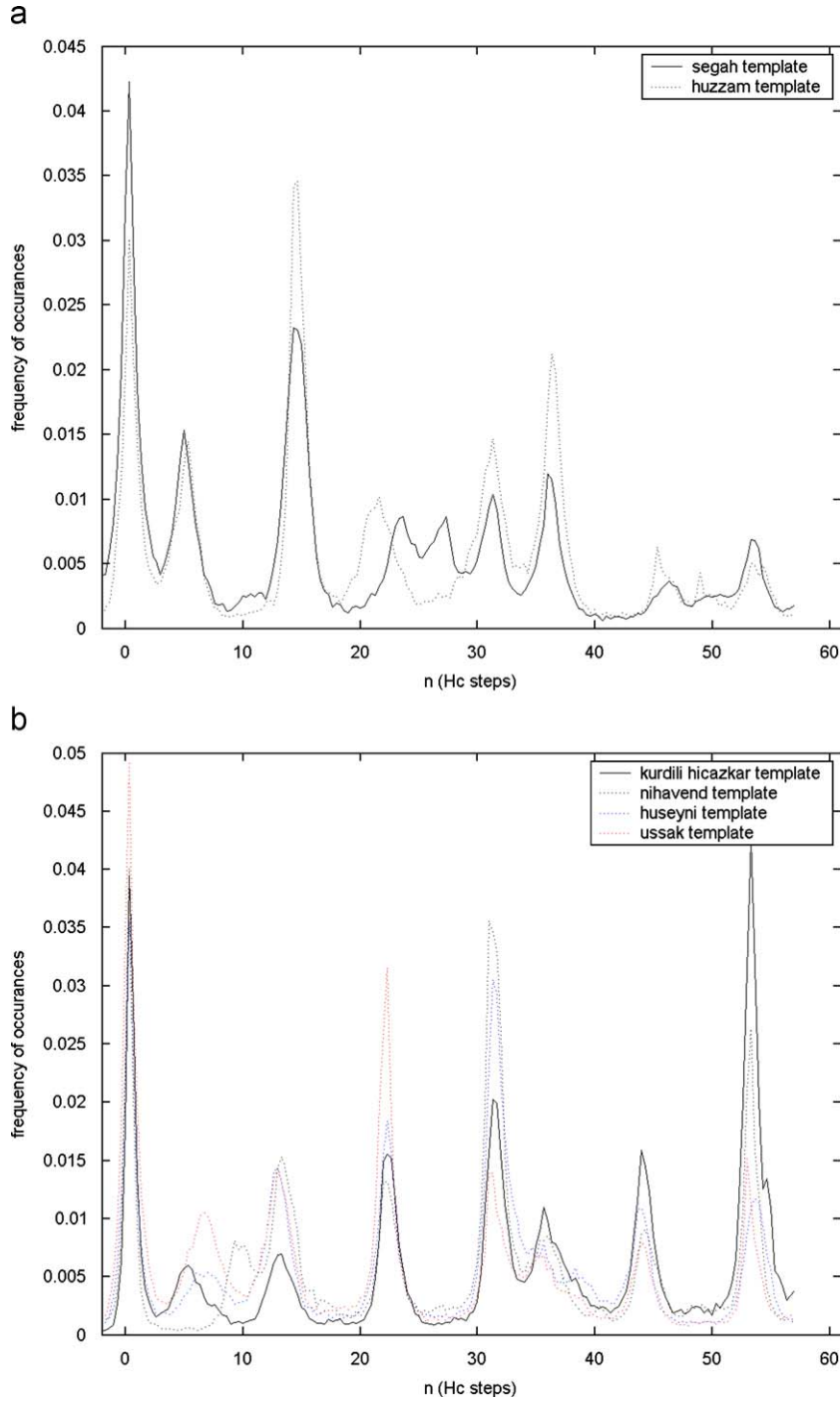
Table 2
The evaluation results of the *makam* recognition system.

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
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	–

a

b

Fig. 6. Pitch-frequency histogram templates for the two groups of *makam* (a) *segah* and *hüzzam* (b) *kürdili hicazkar*, *uşşak*, *hüseyni* and *nihavend*.

While the *segah* and *hüzzam* have similar pitch-frequency histograms on the one side, as shown in Fig. 6a, *kürdili hicazkar*, *uşşak*, *hüseyni* and *nihavend* have similar pitch-frequency histograms on the other side, as depicted in Fig. 6b. On the other hand, two *makam* types, *hicaz* and *saba*, are not confused due to the dissimilar pitch-frequency histograms as shown in Fig. 5.

The most confused *makamlar* are also evaluated from the view point of music theoretical knowledge, especially the theory founded by Arel [8]. 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*.

Lastly, in a recent study [55] focusing mainly on histogram matching techniques and similarity measures for Turkish music, the comparison of the pitch spaces of western music and Turkish music is presented empirically over only two makamlar, *saba* and *hicaz*. Instead of a continuous representation, samples are represented by the peaks of the pitch-frequency histograms. Therefore, recordings are represented similar to the pitch-classes in western music. Then, a similar data-driven model is applied for the classification of two *makamlar* using the correlation coefficient, the most frequently applied distance measure in MIR studies on western music. Consequently, a success rate of 21% is found in terms of the *F*-measure. On the other hand, when the continuous representation of pitch space, the pitch-frequency histogram, is used, a success rate of 95% is obtained in terms of the *F*-measure. This result empirically demonstrates the difference between the pitch spaces of Turkish music and western music.

4. Discussions, conclusions and future work

In this study, 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 the literature. We have shown in Section 3 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 *makams*, 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 nine 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.

Acknowledgment

References

[1] H.S. Powers, et al., Mode, in: Grove Music Online, Oxford Music Online ⟨http://www.oxfordmusiconline.com/subscriber/article/grove/music/43718pg5⟩ (accessed November 17, 2008).
[2] M.A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, M. Slaney, Content-based music information retrieval: current directions and future challenges, Proc. IEEE 96 (4) (2008) 668–696.
[3] G. Tzanetakis, A. Kapur, W.A. Schloss, M. Wright, Computational ethnomusicology, Journal of Interdisciplinary Music Studies 1 (2) (2007) 1–24.
[4] D. Moelants, O. Cornelis, M. Leman, J. Gansemans, R. De Caluwe, G. De Tré, T. Matthé, A. Hallez, Problems and opportunities of applying data- & audio-mining techniques to ethnic music, in: Proceedings of the International Conference on Music Information Retrieval (ISMIR), Victoria, Canada, 8–12 October 2006, pp. 334–336.
[5] D. Moelants, O. Cornelis, M. Leman, J. Gansemans, R. De Caluwe, G. De Tré, T. Matthé, A. Hallez, The problems and opportunities of content-based analysis and description of ethnic music, International Journal of Intangible Heritage 2 (2007) 58–67.
[6] B. Bozkurt, An automatic pitch analysis method for Turkish maqam music, Journal of New Music Research 37 (1) (2008) 1–13.
[7] O. Yarman, 79-tone tuning & theory for Turkish maqam music, Ph.D. Thesis, İstanbul Technical University, Social Sciences Institute, İstanbul, 2008.
[8] A.C. Gedik, B. Bozkurt, Automatic classification of Turkish traditional art music recordings by Arel theory, in: Proceedings of the Conference on Interdisciplinary Musicology, Thessaloniki, Greece, 3–6 July 2008 ⟨web.auth.gr/cim08/cim08_papers/Gedik-Bozkurt/Gedik-Bozkurt.pdf⟩.
[9] C.-H. Chuan, E. Chew, 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, 15pp.
[10] A. Klapuri, Introduction to music transcription, in: A. Klapuri, M. Davy (Eds.), Signal Processing Methods for Music Transcription, Springer, New York, 2006, pp. 3–20.
[11] A.P. Klapuri, Automatic music transcription as we know it today, Journal of New Music Research 33 (3) (2004) 269–282.
[12] D. Temperley, The Cognition of Basic Musical Structures, MIT Press, Cambridge, MA, 2001, pp.167–201 (Chapter 7).
[13] H.C. Longuet-Higgins, M.J. Steedman, On interpreting Bach, Machine Intelligence 6 (1971) 221–241.
[14] C.L. Krumhansl, Cognitive Foundations of Musical Pitch, Oxford University Press, New York, 1990.
[15] D. Temperley, Pitch-class distribution and the identification of key, Music Perception 25 (3) (2008) 193–212.
[16] F. Lerdahl, R. Jackendoff, A Generative Theory of Tonal Music, MIT Press, Cambridge, MA, 1983.
[17] E. Chew, The spiral array: an algorithm for determining key boundaries, in: Proceedings of the LNCS/LNAI, vol. 2445, Springer, Scotland, 2002, pp. 18–31.
[18] Y. Zhu, M.S. Kankanhalli, Precise pitch profile feature extraction from musical audio for key detection, IEEE Transactions on Multimedia 8 (3) (2006) 575–584.
[19] E. Gomez, Tonal description of polyphonic audio for music content processing, INFORMS Journal on Computing. Special Cluster on Music Computing, 18 (3) (2006) 294–304.
[20] C. Chuan, E. Chew, Polyphonic audio key finding using the spiral array CEG algorithm, in: IEEE Conference on Multimedia and Expo (ICME), Amsterdam, Netherlands, 6–8 July 2005, pp.21–24.
[21] H. Purwins, B. Blankertz, K. Obermayer, A new method for tracking modulations music in audio data format, in: Proceedings of the IEEE-INNS-ENNS, vol. 6, 2000, pp. 270–275.
[22] B.S. Ong, E. Gomez, S. Streich, Automatic extraction of musical structure using pitch class distribution features, in: Proceedings of the Workshop on Learning the Semantics of Audio Signals (LSAS), Athens, Greece, 6 December 2006, pp. 53–65.
[23] E. Gomez, P. Herrera, Estimating the tonality of polyphonic audio files: cognitive versus machine learning modelling strategies, in: Proceedings of the International Conference on Music Information Retrieval (ISMIR), Barcelona, Spain, 10–14 October 2004, pp. 92–95.
[24] K. Lee, M. Slaney, 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) (2008) 291–301.
[25] Y. Liu, Y. Wang, A. Shenoy, W.-H. Tsai, L. Cai, Clustering music recordings by their keys, in: Proceedings of the International Conference on Music Information Retrieval (ISMIR), Philadelphia, Pennsylvania, USA, 14–18 September 2008, pp. 319–324.

[26] A. Krishnaswamy, On the twelve basic intervals in South Indian classical music, in: AES 115th Convention, New York, USA, 10–13 October 2003, paper no: 5903.

[27] A. Krishnaswamy, Multi-dimensional musical atoms in South Indian classical music, in: Conference on Music Perception and Cognition, Evanston, Illinois, USA, 3–7 August 2004 ⟨http://www-ccrma.stanford.edu/~arvindh/cmt/icmpc04.pdf⟩.

[28] A. Krishnaswamy, Pitch measurements versus perception of South Indian classical music, in: Proceedings of the Stockholm Music Acoustics Conference (SMAC-03), vol. 2, 6–9 August 2003, pp. 627–630.

[29] A. Krishnaswamy, Application of pitch tracking to South Indian classical music, in: IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 19–22 October 2003, pp. 49–.

[30] F. Marandola, The study of musical scales in Central Africa: the use of interactive experimental methods, in: Proceedings of the Computer Music Modeling and Retrieval, 26–30 October 2003, pp. 34–41.

[31] E. Carterette, R. Kendall, Comparative music perception and cognition, in: D. Deutsch (Ed.), The Psychology of Music, second ed.,, Academic Press, San Diego, 1999, pp. 725–792.

[32] A. Nesbit, L. Hollenberg, A. Senyard, Towards automatic transcription of Australian Aboriginal music, in: Proceedings of the International Conference on Music Information Retrieval (ISMIR), Barcelona, Spain, 10–14 October 2004, pp. 326–330.

[33] M.S. Sinith, K. Rajeev, Pattern recognition in South Indian classical music using a hybrid of HMM and DTW, in: Conference on Computational Intelligence and Multimedia Applications, vol. 2, IEEE Computer Society, Silver Spring, MD, 2007, pp. 339–343.

[34] C. Akkoc, Non-deterministic scales used in traditional Turkish music, Journal of New Music Research 31 (4) (2002) 285–293.

[35] M. Norowi, S. Doraisamy, R. Wirza, Factors affecting automatic genre classification: an investigation incorporating non-western musical forms, in: Proceedings of the International Conference on Music Information Retrieval (ISMIR), London, UK, 11–15 September 2005, pp. 13–20.

[36] P. Chordia, A. Rae, Raag recognition using pitch-class and pitch-class dyad distributions. in: Proceedings of the International Conference on Music Information Retrieval (ISMIR), Vienna, Austria, 23–27 September 2007.

[37] P. Chordia, M. Godfrey, A. Rae, Extending content-based recommendation: the case of Indian classical music, in: Proceedings of the International Conference on Music Information Retrieval (ISMIR), Philadelphia, Pennsylvania, USA, 14–18 September 2008, pp. 319–324.

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

[39] O. Lartillot, P. Toiviainen, T. Eerola, 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) (2008) 157–160.

[40] P. Toiviainen, T. Eerola, Visualization in comparative music research, in: A. Rizzi, M. Vichi (Eds.), COMPSTAT 2006—Proceedings in Computational Statistics, Physica-Verlag, Heidelberg, 2006, pp. 209–221.

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

[42] A. Zeren, Müzik sorunlarımız üzerine araştırmalar, Pan Yayıncılık, İstanbul, 2003.

[43] M.K. Karaosmanoğlu, C. Akkoc, Turk musikisinde icra - teori birligini saglama yolunda bir girisim. in: Proceedings from 10th Müzdak Symposium, İstanbul, Turkey, 2003.

[44] M.K. Karaosmanoğlu, Turk musikisi perdelerini olcum, analiz ve test teknikleri. in: Proceedings from Yıldız Teknik Üniversitesi Turk Muziği Geleneksel Perdelerini Çalabilen Piyano İmâli Projesi sunumu, İstanbul, Turkey, 2004.

[45] A. de Cheveigne, H. Kawahara, Yin: a fundamental frequency estimator for speech and music, Journal of the Acoustic Society of America 111 (2002) 1917–1930.

[46] O. Akdoğu, Taksim nedir, nasil yapilir?, Izmir, 1989.

[47] S.-H.S. Cha, N. Srihari, On measuring the distance between histograms, Pattern Recognition 35 (2002) 1355–1370.

[48] M.J. Swain, D.H. Ballard, Color indexing, International Journal of Computer Vision 7 (1) (1991) 11–32.

[49] M.K. Karaosmanoğlu, Turk musikisinden secmeler, Nota Yayincilik, Istanbul, 2007.

[50] R. Brunelli, T. Poggio, Face recognition: features versus templates, IEEE Transactions on Pattern Analysis and Machine Intelligence 15 (10) (1993) 1042–1052.

[51] K. Tanaka, M. Sano, S. Ohara, M. Okudaira, A parametric template method and its application to robust matching, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2000, pp. 620–627.

[52] C.L. Li, K.C. Hui, A template-matching approach to free-form feature recognition, Proceedings of the IEEE International Conference on Information Visualization (2000) 427–433.

[53] S. Santini, R. Jain, Similarity measures, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (9) (1999) 871–883.

[54] Y. Oztuna, Makam, Turk Musikisi:Akademik Klasik Turk San'at Musikisi'nin Ansiklopedik Sozlugu, II. Orient Yay, Cilt, Ankara, 2006.

[55] A.C. Gedik, B. Bozkurt, Pitch histogram based similarity measures for Turkish music, 2008, submitted for publication.

[56] A.C. Gedik, B. Bozkurt, Evaluation of the Makam Scale Theory of Arel for Music Information Retrieval on Traditional Turkish Art Music, Journal of New Music Research, in press.