SCORE INFORMED TONIC IDENTIFICATION FOR MAKAM MUSIC OF TURKEY

Sertan Şentürk, Sankalp Gulati, Xavier Serra
Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
{sertan.senturk, sankalp.gulati, xavier.serra}@upf.edu

ABSTRACT

Tonic is a fundamental concept in many music traditions and its automatic identification should be relevant for establishing the reference pitch when we analyse the melodic content of the music. In this paper, we present two methodologies for the identification of the tonic in audio recordings of *makam* music of Turkey, both taking advantage of some score information. First, we compute a prominent pitch and a audio kernel-density pitch class distribution (KPCD) from the audio recording. The peaks in the KPCD are selected as tonic candidates. The first method computes a score KPCD from the monophonic melody extracted from the score. Then, the audio KPCD is circular-shifted with respect to each tonic candidate and compared with the score KPCD. The best matching shift indicates the estimated tonic. The second method extracts the monophonic melody of the most repetitive section of the score. Normalising the audio prominent pitch with respect to each tonic candidate, the method attempts to link the repetitive structural element given in the score with the respective time-intervals in the audio recording. The result producing the most confident links marks the estimated tonic. We have tested the methods on a dataset of *makam* music of Turkey, achieving a very high accuracy (94.9%) with the first method, and almost perfect identification (99.6%) with the second method. We conclude that score informed tonic identification can be a useful first step in the computational analysis (e.g. expressive analysis, intonation analysis, audio-score alignment) of music collections involving melody-dominant content.

1. INTRODUCTION

Pitch relationships in pitch space and time constitute the fundamental building block of musical melody. In many musical styles, there is the concept of "tonic," which acts as the reference tuning pitch for the melody. The interrelations between the tonic and other pitches establish a hierarchical organisation, which is highly related to the perception, cognition and anticipation of music [9]. The automatic identification of the tonic of a piece is a musically relevant step in the computational analysis of many melodic characteristics.

Nevertheless, the tonic concept encompasses different meanings and characteristics within the cultural, musicological, acoustic (and linguistic) context of different music traditions. For example, in some music traditions, the tonic of a performance is changed according to instrument characteristics, personal preferences or for historical relevance. In such cases tonic identification is necessary to establish the reference pitch to carry further computational tasks such as tuning analysis, intonation analysis and melodic structure recognition. For systems that aim to analyse the melodic content of such musics, knowledge-based tonic identification approaches may be required [7, 14, 15].

Pitch distributions (PDs) and "octave-wrapped" pitch class distributions (PCDs) are commonly used for analysis of tonic and pitch organisation. Krumhansl and Shephard [11] used 12-dimensional PCDs to study the tonal organisation of euro-genetic musics. PCDs are also used for relevant tasks such as key detection and chord recognition [8, 16] for euro-genetic musics.

For musical styles involving microtonality, the pitch space must be extended beyond 12-dimensions to model, analyze and predict the melodic properties of the studied music [3, 4, 7]. Gedik and Bozkurt [7] propose a method for tonic and *makam* recognition for *makam* music of Turkey (MMT). The method generates a histogram based, fine-grained pitch distribution (FPD) for the test audio, and for each *makam* from the annotated audio recordings. Given the *makam* of the test audio, the first bin in the template FPD is assigned to an arbitrary frequency below audio FPD such that the FPDs do not overlap initially. Then, the method compares the template FPD with the audio FPD by shifting the template FPD at each step. In the best matching shift, the frequency of the tonic of the template is labeled as the estimated tonic.

In [2], a joint tonic and raag (melodic structure) recognition methodology was presented for North Indian classical music. Instead of generating a single template for each raag, the method generates multiple PCDs for each raag from the pitch tracks extracted from the annotated audio excerpts. Given the raag of a musical excerpt, the PCD computed from the excerpt is compared with each of the template PCDs of the same raag by circularly shifting the test PCD. The shift in the closest match indicates the tonic.

The disadvantage of using audio recordings for template computation is the necessity of adequate amount of training data. Moreover, the quality of the data has to be maintained so that the intervallic properties are represented well. Even so, a test distribution can substantially differ from the corresponding template. A common confusion is the estimation of another pitch (or pitch class) when its occurrence is comparable to the occurrence of the tonic. Moreover, in cases when an audio recording includes unrelated musical content in addition to a performed piece, e.g. improvisations or performances of other pieces in different modal structures, the audio distribution would be a mixture of the distributions of these distinct musical events. This might cause substantial confusions. This problem motivates the replacement of audio recordings with a more "definitive" information source in the template training step. If available, scores can be good sources, since they provide an easily accessible symbolic description of many relevant musical components.

When score information is available, utilising the sequential note information might bring a more effective solution to tonic identification. In [4], we introduce a method to link the musically relevant structural elements (sections) given in the score and the corresponding time-intervals in a audio recording. The method estimates all possible links in an audio recording by applying Hough transform to similarity matrices computed between the prominent pitch extracted from the audio recording and the synthetic pitch extracted from the score fragments. The estimated links are searched according to the section sequence given in the score to obtain the most-likely section links.

In this paper, we present two methodologies to identify the tonic of a performed piece by comparing the melodic contents extracted from the performance and the score of the piece. We use *makam* music knowledge and the findings from previous research [2, 4, 7] to specialise both the methodologies for the melodic aspects of *makam* music of Turkey. Both methods extract prominent pitch from the audio recording. Then a fine-grained, kernel-density pitch class distribution (KPCD) is computed from the audio prominent pitch, and tonic candidates are selected. Adapted from [2, 7], Method I applies circular shifting to the audio KPCD according to tonic candidates. Each shift is then compared with a score KPCD computed from the monophonic melody in the score. Method II normalises the prominent pitch with respect to each tonic candidate. Next, it attempts to link melodic fragments in the score with the respective time intervals in the audio by using the candidate link estimation approach explained in [4]. As the first experiments in "score-informed" tonic identification, we consider inter-linked audio-score collections, where the audio recording and the score are already known to be related with the same work (composition). We use musical scores, which include the *makam* of the piece, the boundaries of the structural elements and the sequence of these elements.

The remainder of the paper is as follows: Section 2 makes a brief introduction to *makam* music of Turkey. Section 3 explains the proposed methodologies. Section 4 presents the data collection used in the experiments. Section 5 explains the experiments done to test the methodologies and provides the results. Section 6 wraps up the paper with a discussion and conclusion.

2. MAKAM MUSIC OF TURKEY

The melodic structure of most traditional music repertoires of Turkey follows the concept of *makam*s. *Makam*s are modal structures, where the melodies typically revolve around a *başlangıç* (starting, initial) tone and a *karar* (ending, final) tone [6]. *Karar* is synonymous to tonic. There are a number of different transpositions (*ahenk*), any of which might be favored over others due to instrument/vocal range or aesthetic concerns [6]. The default ahenk is called "*bolahenk*." For an extended discussion of *ahenk*s, the readers are referred to [6, Appendix F].

Currently, Arel-Ezgi-Uzdilek (AEU) theory is the mainstream theory for the *makam* music of Turkey (MMT) [6]. AEU theory argues that there are 24 equal intervals and a whole tone is divided into 9 equidistant intervals. These intervals can be approximated from 53-TET (tone equal tempered) intervals, each of which is termed as a *Holdrian comma* (1 Hc ≈ 22.6415 cents) [6]. On the other hand, the intonation of some intervals in the performance might differ from the theoretical intervals as much as a semitone [6, 17]. [1]

Since early 20^{th} century, a score representation extending the Western music notation has been used in MMT [12]. This notation typically follows the rules of AEU theory. The scores tend to notate monophonic melodic lines. However performances may differ from the score substantially due to the heterophonic characteristics of makam music and artistic decisions such as non-notated embellishment, note/ phrase insertion, repetitions and omissions. The register information given in the notation is always relative to the instrument, i.e. the lowest tone of the same pitch class that an instrument can produce is always indicated with the same symbolic note in the score. In *bolahenk*, the notes are written a perfect fourth higher than it sounds, i.e. the *rast* note represented by the $G4$ in the staff is played in the pitch class of approximately $D4 = 293.66$ Hz [6].

In the experiments, we focus on *peşrev*, *saz semaisi* (the two most common instrumental forms) and *şarkı* (the most common vocal form) forms from the classical reperoire. Both peşrev and saz semaisi typically consist of a repetitive section called *teslim*. In the *şarkı* form, there is typically a repetitive section called *nakarat*.

3. PROPOSED APPROACHES

We define tonic identification as "estimating any frequency belonging to the same pitch class of the *karar* note." Even though the octave information is important in *makam*s, we

[1] Throughout the text we will represent a note name as [note letter][octave]$\{$accidental$\}^{\{\text{Hc distance}\}}$, e.g. the note *mahur* is written as $G5\flat^4$. Note that the performed frequency might be different.

chose to generalize the term tonic to the pitch class of the *karar* note due to some performance scenarios, where it is ambiguous to define the octave of the performance tonic. [2]

Given the problem definition, we propose two methodologies to identify the tonic of a performance of MMT using score information. In this paper, we focus on *interlinked* audio-score collections, where the scores and audio recordings are already related with the respective compositions via available metadata. Both methodologies take advantage of a machine readable score, which stores the value and the duration (i.e. the $\langle$*note-name*, *duration*$\rangle$ tuple) of each note. The tuple sequence form a symbolic monophonic melody. Additionally, the score is divided into sections, some of which are repeated. The *makam* and the tempo of the piece are provided in the score. Therefore, we do not need any structural analysis to find the repetitive structural element. The audio recording might include various expressive decisions such as musical materials that are not related to the piece, phrase repetitions/omissions and pitch deviations.

From music-theory [6], we compile a dictionary consisting on $\langle$*makam*, *karar*$\rangle$ pairs, which stores the *karar* of each *makam* (e.g. if the *makam* of the piece is Hicaz, the *karar* is A4.). *Karar* note is used as the reference symbol during the generation of synthetic pitch from the score (Section 3.1.1). We also refer to theoretical intervals defined in AEU theory to generate the score features from the machine-readable score (Section 3.1.1).

Method I generates a synthetic pitch from the monophonic melody given in the score and extracts a prominent pitch from the audio recording (Section 3.1.1). Kernel density estimation (KDE) is applied to these melodic features and the 160-D kernel-density pitch class distributions (KPCDs) are obtained (Section 3.1.2). The audio KPCD is circularly shifted according to its peaks and compared to the score KPCD. The candidate used to shift the audio KPCD, which results in the minimum distance to score KPCD, is selected as the tonic (Section 3.2).

Method II (Section 3.3) uses the note sequence information given in the score. Similar to the first method, it computes a prominent pitch and audio KPCD from the audio recording. From the score, the second method only extracts the synthetic pitch from the monophonic melody of the repetitive structural element (e.g. teslim, nakarat) indicated in the score. The method then normalizes the audio prominent pitch with respect to each peak in the audio KPCD. Then, it attempts to link the repetitive structural element in the score with its respective locations in the audio recording. The candidate of the section linking result, which outputs the most confident links, is selected as the tonic.

Before presenting the methodologies, we first explain the feature extraction and tonic candidate selection steps, which are shared by both the approaches.

[2] As an example, consider an ensemble performance, where each instrument performs the same melodic contour in its own register.

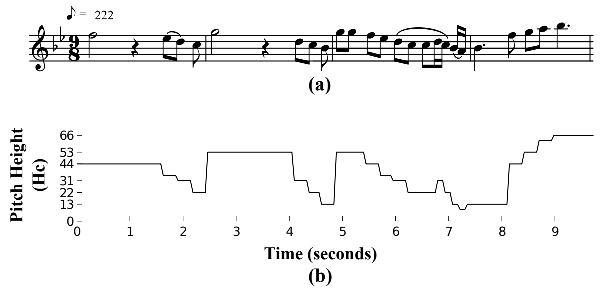

Figure 1. The first nakarat section of the composition, *Gel Güzelim*. a) Score, b) Synthetic pitch computed from the note symbols and durations.

3.1 Feature Extraction

Notation and audio recording are different representations of music. To compare these information sources, we need to extract features which adequately capture the musical content given in each representation. In [4], we found that prominent pitch is a highly effective and intuitive feature to analyze MMT due to the monophonic nature of the scores and the heterophonic practice. From the prominent pitch, we further compute a fine-grained pitch class distribution. FPCDs are shown to model the intervallic properties of the *makam*s adequately [7]. Moreover, FPCDs are able to capture the intonation information in a limited fashion, i.e. the width and shape of the peaks.

3.1.1 Synthetic Pitch and Prominent Pitch Computation

For the computation of synthetic pitch from the score and prominent pitch from the audio, we use the feature extraction step explained in [4]. Here we give a brief summary of the process.

To compute the synthetic pitch p_s for the desired score fragment (i.e. the whole score in Method I and the repetitive section in Method II), we extract the corresponding $\langle$*note-name*, *duration*$\rangle$ tuple sequence associated with the fragment. Then, we pick the *makam* of the composition, which is given in the score, and obtain the *karar*-name of the piece by checking the *makam* in the $\langle$*makam*, *karar*$\rangle$ dictionary. The note names are mapped to the Hc distances according to AEU theory with reference to the *karar* note. Finally, the synthetic pitch for the score fragment is generated at a frame rate of ~46 ms, which provides sufficient time resolution to track all changes in pitch. Figure 1 shows the repetitive section given in the score of the composition, *Gel Güzelim*, [3] and the synthetic pitch computed for this fragment.

To obtain the audio prominent pitch p_a, we use the Essentia implementation [1] of the melody extraction algorithm proposed by [13]. The approach computes the main melody after separating salient melody candidates from non-salient ones. We include all the non-salient candidates to guess the prominent pitch since non-melodic intervals are very rare in MMT. Melody extraction is done using a pitch precision of 7.5 cents ($\approx \frac{1}{3}$ Hc), which is reported

[3] *http://tinyurl.com/lfp8x83*

as a suitable pitch precision for MMT [7]. The hop size is chosen equal to the frame rate of the p_s (∼46 ms).

3.1.2 Pitch Distribution Computation and Tonic Candidate Selection

A fine-grained pitch class distribution (FPCD) is computed from the audio prominent pitch to find tonic candidates in both methods. Method I also computes a score FPCD from the synthetic pitch of the whole score and shift-compares it with the audio FPCD. For the audio FPCD computation, the unit of the prominent pitch is converted from Hz to Hc with respect to the middle C (261.63 Hz). This reference is selected as a dummy value for the unit conversion.

While shift-comparing FPCDs to every possible pitch value is effective for template matching [7], in previous research [2] we showed that picking the location of the peaks in the PCD as tonic candidates greatly reduces the computation time with minimal losses in tonic accuracy. It was also observed that, kernel-density pitch class distributions (KPCDs), which are computed using kernel density estimation (KDE), perform significantly better than histogram-based pitch class distributions, when candidates are selected as the peak locations [2]. Hence, we use kernel density estimation (KDE) to compute the FPCDs. In KDE, an observation contributes to neighbouring bins according to a kernel. When the kernel is chosen as Gaussian, the continuous kernel density $\hat{f}(x)$ is given by:

$$\hat{f}(x) = \frac{1}{nh} \sum_{j=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-p(j))^2}{2h^2}} \tag{1}$$

where h is the kernel width, $p(j)$ is the value of the j^{th} index of the prominent pitch p, and n is the total number of the prominent pitch values.

We use the function *ksdensity* in MATLAB [4] to obtain a discrete approximation of the kernel density. The discrete approximation of kernel density provides smoothness over histogram computation. KDE is especially helpful in score FPCD computation (Section 3.2), since the pitch spread provides robustness to the microtonal deviations in the tuning and intonation. We empirically set the kernel width h to 15 cents ($\approx \frac{2}{3}$ Hc) so that an observation practically contributes within an interval of 4Hc, slightly smaller than a semitone. Finally, the estimated kernel density is octave wrapped, and the KPCD is obtained. We can write a KPCD with N bins as $D = \langle d_1, \ldots, d_i, \ldots, d_N \rangle$, where d_i denotes the value of the pitch class index i. We retain the pitch precision of the prominent pitch (7.5 cents), resulting in $N = 160$ bins. In the audio KPCD (D_a), the first bin is initialised to the dummy value (261.63 Hz) used in Hc conversion. In the score KPCD (D_s), the first bin indicates the pitch class of the *karar* note, i.e. $53k$ Hc, where $k \in \mathbb{Z}$.

We select the index of all the peaks in D_a as the tonic candidates.

[4] *http://www.mathworks.com/help/stats/ksdensity.html*

3.2 Method I: Distribution Matching

In the first method, we compute a score and an audio KPCD from both score and audio recording, respectively. For the score KPCD computation, the whole note sequence is used. The locations of the peaks in the audio KPCD (D_a) are picked as the candidate tonics. Using the template matching approach [2, 7] we apply circular-shift the audio KPCD such that each candidate is carried to the first bin. Circular-shift can be simply formulated as:

$$D^i = \langle d_i, d_{i+1} ..., d_N, d_1, ..., d_{i-1} \rangle \tag{2}$$

Next, the shifted audio KPCD D_a^i is compared to D_s. We use Bhattacharyya distance, which was shown to outperform common L_n distances (e.g. City Block, Euclidean) in PD comparison [2]. Bhattacharyya distance Δ between the score KPCD and shifted audio KPCD can be written as:

$$\Delta(D_s, D_a^i) = -\ln \left(\sum_{k=1}^{N} \sqrt{D_s(k) D_a^i(k)} \right) \tag{3}$$

where k denotes an index of the D_s and D_a^i, respectively.

Bhattacharyya distance is computed between the score and each shifted audio KPCD. The index i which results in the minimum distance, indicates the estimated pitch class. The estimated tonic c_i can be represented as:

$$c_i = 261.63\, k * 2^{\frac{7.5(i-1)}{1200}},\ k \in \mathbb{Z} \tag{4}$$

3.3 Method II: Repetitive Section Linking

Using PCDs, we can only take an advantage of the interval and some limited intonation information. Nevertheless, scores also include note sequence information. In Method II, we attempt to link a melodic fragment from the score with the audio recording by using the candidate link estimation presented in [4].

We first compute the audio prominent pitch p_a. We also compute audio KPCD and obtain the candidate tonics. Using Equation 4 and setting $k = 1$, the candidate indices are converted back to Hz, i.e. $c_i(k = 1)$, where i is the index of the tonic candidate. Next we convert the prominent pitch values from Hz to Hc with respect to each *karar* candidate such that the *karar* candidate has a value of 0 Hc. The *normalised audio prominent pitch* can be expressed as:

$$p_a^i = 53 * \log_2 \left(\frac{p_a}{c_i(k=1)} \right) \tag{5}$$

Next, synthetic pitch of the repetitive section indicated in the score p_s is extracted. We compute a similarity matrix S^i between p_s and p_a^i:

$$S_{jk}^i = \begin{cases} 1, & \left(|p_s(j) - p_a^i(k) + \alpha| \bmod 53\right) - \alpha < \beta \\ 0, & \left(|p_s(j) - p_a^i(k) + \alpha| \bmod 53\right) - \alpha \geq \beta \end{cases} \tag{6}$$

where mod indicates the modulo operation. Each element in the similarity matrix indicates whether two pitch values can be deemed as the same pitch class within the binarization threshold β. α is a dummy value greater than β to ensure pitch differences between $[53k - \beta,\ 53k], k \in \mathbb{Z}$ are treated as similar. In [4], we found that 3 Hc is an optimal value for β. In the similarity matrix S^i, diagonal line segments are observed which indicate the possible performed locations of the repetitive section. We apply Hough transform to detect the diagonal line segments [5]. We observe that the tempo of a performance typically varies between between 0.55 and 1.5 times the tempo indicated in the score [4], which we use to restrict the searched angles between $-28.81°$ and $-56.31°$. We obtain a set of links $\{l_1^i, \ldots, l_m^i\}$ for each tonic candidate, where m refers the number of links found using the tonic candidate. The number of non-zero pixels forming the line segment is normalised by the length of the line segment, giving the weight $w(l_k^i)$, $1 \leq k \leq m$, of the segment. We combine the weight of each link and obtain a accumulated weight for each tonic candidate. The accumulated weight is given as:

$$w^i = \sqrt[3]{\sum_{k=1}^{m} w(l_k^i)^3} \tag{7}$$

Equation 7 ensures that (possibly erroneous) links with low weights are greatly suppressed with respect to the links with high weights. The tonic is estimated as the pitch class c_i, which has the highest accumulated weight w^i.

4. DATA COLLECTION

For our experiments, we collected 116 audio recordings of 24 peşrevs, 84 audio recordings of 19 saz semaisis, and 57 audio recordings of 14 şarkıs (257 audio recordings of 57 compositions in total). The compositions are taken from the classical repertoire, in which the makam and the karar note are clearly defined in music theory. The *makam* of each composition is included in the metadata. [5] The pieces cover 28 different *makam*s.

The scores are obtained from the symbTr database. The symbTr-score is machine-readable text format, which stores the value and the duration of the note sequences [10]. The symbolic representation also contains information about the structure of the composition, i.e. the section sequences and the indices of the initial and final note of each section are indicated.

The audio recordings are selected from the CompMusic collection, and they are either in public-domain or commercially available. Some recordings include musical events which do not belong to the composition such as improvisations and even performances of other compositions.

The ground truth is obtained by manually marking the tonic frequency using *Makam Toolbox* [7]. Figure 2a and Figure 2b shows the distribution of the annotated tonic with respect to the pitch class C and the distribution of the transpositions with respect to *bolahenk*, respectively. It can be seen that the annotated tonic are mostly distributed around the semitones with microtonal deviances. Apart from *bolahenk*, the tonic is mostly performed with a transposition around the perfect fourth, perfect fifth and minor seventh. Nevertheless a considerable number of tonic annotations reside in microtonal pitch classes.

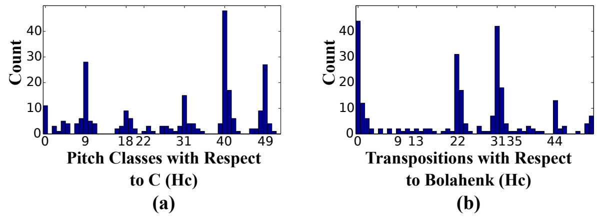

Figure 2. Distribution of the annotated tonics in the data collection. a) Pitch class histogram of the annotated tonic with respect to the pitch class C, b) Histogram of the transpositions with respect to *bolahenk*

5. EXPERIMENTS AND RESULTS

We use the methodologies explained in Section 3 to identify the tonic. We compare the estimated tonic from each algorithm with the manually annotated tonic. If the distance between the estimated and the annotated tonic are less than 1 Hc, the estimation is marked as correct. [6]

Tonic identification by repetitive section linking fails only in one piece (99.2% success rate). In this recording, [7] the vocalist sings a *gazel* (vocal improvisation) in almost three fifth of the duration of the recording with skillful vibratos extending up to ≈ 200 cents peak-to-peak. These vibratos occasionally cross *mahur* (G5$\flat^4$) and less frequently reach to *gerdaniye* (G5), which is in the pitch class of the tonic. Throughout the piece the pitch class G$\flat^4$ is visited more than G and it shows a wide spread towards G such that no peak is formed in the vicinity of the tonic pitch class. In this case the pitch class G$\flat^4$ is estimated as the tonic, having a 2.33 Hc deviation. [8]

Using distribution matching, we are able to identify the tonic of 244 performances out of 257 (94.9% success rate). Most of the errors occur in *makam*s Kürdilihicazkar (3 recordings), Muhayyer (3 recordings), Suzidilara (2 recordings), Isfahan and Mahur (1 recordings each), which have complex pitch distributions. The errors are distributed mostly to the fourth (7 recordings) and fifth (4 recordings) of the scale degree. In 4 recordings the tonic is identified as the başlangıç (initial) note, which is the other melodic center of the *makam*. The average distance between the annotated tonic and the correctly estimated tonics is 0.23 Hc with a standard deviation of 0.21 Hc for both methods.

For comparison, we also modify and test the approach in [7] using the Makam Toolbox implementation. We use the audio prominent pitch as the input to improve the f0-estimation. The *makam* of the piece is provided to the algorithm. We use a subset of the collection with 152 au-

[5] The metadata is stored in MusicBrainz: *http://tinyurl.com/mfvop6l*

[6] The results are available in *http://compmusic.upf.edu/node/164*.

[7] *http://tinyurl.com/n42g5dh*

[8] Interestingly 2 out of 3 section links produced by the erroneous tonic are correct, since the Hc distance between the annotated and estimated tonic (2.33 Hc) is less than the optimal binarization threshold $\beta = 3$ Hc. In the next step, we can align the audio and score in the note-level and correct any errors and micro-deviances in the tonic.

dio recordings. The number of failed identifications is 46, 10 and 1 for audio-based template matching, PCD matching and repetitive section linking, respectively. The results from both of our methods are substantially better than the results obtained from the Makam Toolbox.

6. DISCUSSION AND CONCLUSION

We proposed two novel methods that use score information to identify the tonic of audio recording. Assuming the most played pitch classes as tonic candidates, the first method compares pitch class distributions computed from the audio and score, and the second method searches for a repetitive score fragments in the audio. We tested the methodologies in a scenario of audio-score collections of MMT, where the audio and score are already linked with each other at the document level and the score includes the notes, as well as the structural organization, the *makam* and the tempo of the piece. The results indicate that score information greatly simplifies the tonic identification task. Moreover, the pitch deviances between the estimated tonic and the annotated tonic are mostly indiscernible. These findings point out the computational potential of knowledge-driven methodologies using multi-modal information.

While template distributions computed from audio are similar to the testing distributions with respect to the tuning and limited intonation information in *makam* level, score distributions indicate these similarities in the (more definitive) composition level. On the other hand, the distribution matching method is still susceptible to the errors seen us audio-based template matching. In the majority of the recordings where distribution matching failed, it was observed that the piece has modulations to pitches that do not belong to the scale of the *makam*. These contrastive notes and any event can be grouped into characteristic fragments, melodic progressions and structural elements; eventually building the unique the music piece. In general, the lack of such temporal information is the main problem of distribution matching.

By linking repetitive sections, we only missed the tonic of one performance. These results indicate the usefulness of the temporal information in pitch related tasks. The successful results obtained from tonic identification and previously from section linking [4] motivates adapting the "fragment linking" methodology for further computational tasks. First is to generalize the method to less "complete" scores, where structure information is unknown. Our initial tests show that tonic estimation by linking non-repetitive fragments and score fragments as short as 7 seconds is possible. In the next step, we want to work on linking audio and scores in the document level by trying to link sections in each score with corresponding audio recordings. Highly ranked links will indicate the scores and audio recordings related to the same work.

Another interesting direction is to generate predictive models from the scores of each *makam*. The models can be used to discover characteristic phrases, which could be linked with the audio to further carry relevant tasks such as *makam* recognition, melodic similarity analysis and expression analysis. Previously we found that multiple viewpoints may be highly predictive in modelling MMT [3]. We plan to take advantage of these computational methodologies and models to discover, navigate through and appreciate cultural-specific aspects of *makam* music of Turkey and other music genres/traditions involving melody-dominant content.

7. ACKNOWLEDGEMENTS

This work is partly supported by the European Research Council under the European Union's Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583).

8. REFERENCES

[1] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra. Essentia: An audio analysis library for music information retrieval. In *Proceedings of ISMIR*, 2013.

[2] P. Chordia and S. Şentürk. Joint recognition of raag and tonic in North Indian music. *Computer Music Journal*, 37(3), 2013.

[3] S. Şentürk. Computational modeling of improvisation in Turkish folk music using variable-length Markov models. Master's thesis, Georgia Institute of Technology, 2011.

[4] S. Şentürk, A. Holzapfel, and X. Serra. Linking scores and audio recordings in makam music of Turkey. *Journal of New Music Research*, (submitted).

[5] R. O. Duda and P. E. Hart. Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, 1972.

[6] E. B. Ederer. *The Theory and Praxis of Makam in Classical Turkish Music 1910-2010*. PhD thesis, University of California, Santa Barbara, September 2011.

[7] Ali Cenk Gedik and Barış Bozkurt. Pitch-frequency histogram-based music information retrieval for Turkish music. *Signal Processing*, 90(4):1049–1063, 2010.

[8] Emilia Gómez. *Tonal Description of Music Audio Signals*. PhD thesis, Universitat Pompeu Fabra, 2006.

[9] D. B. Huron. *Sweet anticipation: Music and the psychology of expectation*. MIT press, Cambridge, Massachusetts, 2006.

[10] K. Karaosmanoğlu. A Turkish makam music symbolic database for music information retrieval: SymbTr. In *Proceedings of ISMIR*, pages 223–228, 2012.

[11] C. L. Krumhansl and R. N. Shepard. Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of experimental psychology: Human Perception and Performance*, 5(4):579–594, 1979.

[12] E. Popescu-Judetz. *Meanings in Turkish Musical Culture*. Pan Yayıncılık, Istanbul, 1996.

[13] J. Salamon and E. Gómez. Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(6):1759–1770, 2012.

[14] J. Salamon, S. Gulati, and X. Serra. A multipitch approach to tonic identification in Indian classical music. In *Proceedings of ISMIR*, pages 499–504, 2012.

[15] X. Serra. A multicultural approach in music information research. In *Proceedings of ISMIR*, pages 151–156, 2011.

[16] D. Temperley and E. W. Marvin. Pitch-class distribution and the identification of key. *Music Perception: An Interdisciplinary Journal*, 25(3):193–212, 2008.

[17] Yalçın Tura. *Türk Musıkisinin Meseleleri*. Pan Yayıncılık, Istanbul, 1988.