

# A METHOD FOR TONIC FREQUENCY IDENTIFICATION OF TURKISH MAKAM MUSIC RECORDINGS

**Hasan Sercan Atlı**  
Bahçeşehir Üniversitesi  
hsercanatli  
@gmail.com

**Bariş Bozkurt**  
Koç Üniversitesi  
barisbozkurt0  
@gmail.com

**Sertan Şentürk**  
Universitat Pompeu Fabra  
sertan.senturk  
@upf.edu

## ABSTRACT

Karar is the final tone in the Turkish makam music performances. The karar frequency, hence the concert pitch varies among performances due to the existence of many diapasons in Turkish makam music, instead of a single standard one. Correct estimation of the karar frequency is critical for many computational tasks such as tuning analysis, audio-score alignment and automatic transcription. We present a new karar frequency identification method that is based on detecting the last note in the recording and estimating its frequency. The method is applied on two large datasets of Turkish makam music recordings and shown to outperform the state-of-the-art.

## 1. INTRODUCTION

In Turkish Makam Music, the melodic dimension is explained by the *makam* (melodic structure) and the *form*. *Makam* defines the melodic progression from an initial tone to a final tone. Gürmeriç describes the *makam* as "Makam, before everything else, is based on a scale. *Makam* is a progression that gives the *makam* a life by starting from somewhere of the *seyir* (melodic progression), moving towards the *güçlü* (dominant) and going towards *karar*" (Gürmeriç, 1966), (Bozkurt et al., 2014). *Karar* (final) is typically used synonymous to tonic in Turkish makam music. Theoretically and as a very common practice, Turkish makam music performances end at the karar note (Akdoğan, 1989).<sup>1</sup>

There is no commonly agreed reference frequency (such as  $A4 = 440$  Hz in euro-genetic musics) in Turkish makam music. Moreover the musicians might chose to play the music in one of the 12 *ahenk*s (diapason) due to instrument/vocal range or for musical aesthetics. The tonic note depends on the performed *makam*<sup>2</sup>. The tonic frequency is required for many computational tasks such as automatic transcription (Benetos & Holzapfel, 2013), tuning analysis (Bozkurt, 2012), makam recognition (Gedik & Bozkurt, 2010) and audio-score alignment (Şentürk et al., 2014). While manually annotating the tonic frequency is relatively straightforward (e.g. isolating a tonic note in the performance and adjusting the frequency of a tone which fits this note), it is a time-consuming task. Thus, automatic tonic identification greatly facilitates research based on melody, especially on large data collections.

<sup>1</sup> For the rest of the paper, the note names are written all in lower case and the first letter of a makam name is written in capitals for clarity.

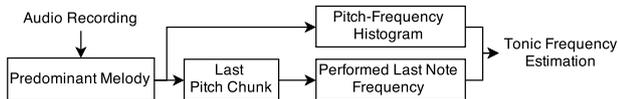
<sup>2</sup> e.g. *karar* of the Rast *makam* is the rast note (G); *karar* of the Hicaz *makam* and Saba *makam* is the düğah note (A).

Pitch distributions (such as *pitch* histograms and *pitch-class* histograms) are commonly used in automatic tonic identification (Gedik & Bozkurt, 2010; Chordia & Şentürk, 2013). In (Krumhansl & Shepard, 1979) 12-dimensional pitch frequency distribution are used to study euro-genetic musics. Pitch distributions are also used for relevant tasks such as key detection and chord recognition (Gómez, 2006) for euro-genetic musics.

For computational analysis of Turkish makam music recordings, the distributions are computed from a melody feature such as fundamental pitch (Gedik & Bozkurt, 2010) or predominant melody (Şentürk et al., 2013). Gedik & Bozkurt (2010) have proposed the so-called *makam histogram template matching* (MHTM) method for automatic tonic identification. In this method the makam of the target audio recording is known. This method computes a template pitch histogram for the *makam*. This template is computed from a set of audio recordings in the same *makam*. Similarly a pitch histogram is computed for the audio recording with unknown tonic frequency. The pitch histogram is shifted and compared against the template histograms. The tonic frequency is identified from the best matching shift. While this method is shown to provide reliable results, it requires training data and *makam* information. Moreover the method might fail in audio recordings with multiple *makams* (e.g. multiple compositions, improvisations in different *makams*) and compound makams, which show the characteristics of multiple makams.

Instead of the histogram templates, Şentürk et al. (2013) use machine readable scores to identify the tonic frequency. In this method, performed composition in the audio recording is known and the score for the composition is available. The method extracts the predominant melody from the audio recording. Then kernel-density estimation is applied to the predominant melody to obtain a pitch-class distribution. The peaks of the distribution are selected as tonic candidates. Next the method attempts to partially align the score with the audio recording. The alignment with the best score indicates the tonic frequency. This method provides almost perfect results. However it is not generalizable since it requires composition information and music scores may not be available for all performances (e.g. improvisations).

In this paper we propose a method, which uses the musical knowledge that a makam music performance ends in the tonic note. Our method detects the last note in an au-



**Figure 1:** Flow diagram of the LND method

dio recording and identifies the tonic frequency from the pitch estimation of this note. We term this methods as the last note detection (LND) method. This method provides a simpler solution than the current state-of-the-art. Moreover, it does not require any additional information such as the performed *makam* or the composition, hence it is more generalizable. We evaluate the LND method with on two different datasets and show that it outperforms the MHTM method.

The rest of the paper is as follows: Section 2 explains the steps of the MHTM and the LND methods. Section 3 represents experiments and test datasets and Section 4 wraps up the paper with a brief conclusion.

## 2. METHODOLOGY

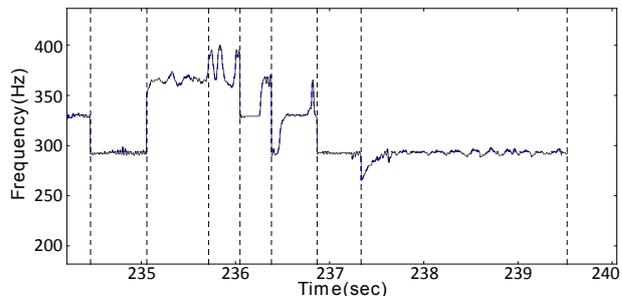
The methodology proposed in this paper assumes that Turkish makam music recordings ends on the tonic note.<sup>3</sup> Our method first extracts the predominant melody from the audio recording. The end of the predominant melody is divided into chunks according to jumps on the pitch and the last chunk is estimated as the last note. Initially we estimate the tonic frequency as the median of all the frequencies in this chunk. Next we compute a pitch histogram using the frequency values in the predominant melody, which are close to the initial estimation ( $\pm$  a semitone). Then we refine our estimation as the frequency of the closest peak in the histogram. The flow diagram of the LND method is shown in Figure 1.

Below we explain the steps of the methodology in detail.

### 2.1 Predominant Melody Extraction

Fundamental frequency estimation or predominant melody extraction of an audio recording is the first step of tonic frequency identification methodologies. In Bozkurt (2008), Bozkurt uses YIN (De Cheveigné & Kawahara, 2002) algorithm and with some post-processing steps (e.g. octave correction) to obtain fundamental frequency estimation of monophonic Turkish makam music recordings. Although the YIN algorithm works fine on monophonic recordings, it is not reliable in polyphonic (e.g. ensembles) recordings of Turkish makam music (Şentürk et al., 2014). Bozkurt (2008) also reports that YIN is not reliable in low energy parts (of the signal) with high background noise.

<sup>3</sup> This information is confirmed by the masters we have consulted, such as the kanun virtuoso Reha Sağbaş. Moreover compositions ending with a note other than the *karar* of the relevant *makam* is extremely rare. For example, there is a single example out of 2200 scores in the SymbTr (Karaosmanoğlu, 2012), the largest machine readable score database of Turkish makam music.



**Figure 2:** Last 8 pitch chunks of an audio recording. The vertical lines mark the borders of pitch chunks.

Şentürk et al. (2014) use the predominant melody extraction algorithm proposed by Salamon & Gómez (2012) with modified parameters according to the culture-specific aspects of Turkish makam music. It is reported that the algorithm provides better results for both polyphonic and noisy than YIN for polyphonic and noisy recordings. In LND method we use a simplified version of the approach explained in Atlı et al. (2014). This version is observed to provide better pitch estimations especially for the parts of the recording the musicians play comparatively softer.

### 2.2 Note Boundary Estimation

To observe the last note in LND method, the end of the extracted predominant melody is divided into pitch chunks. The boundaries are placed where the slope of the melody is higher than a specified threshold. In Figure 2, last 8 pitch chunks of a recording<sup>4</sup> is shown.

The last chunk in the predominant melody is estimated as the tonic note. The method computes the median of the frequency values in this chunk as the initial tonic frequency estimation. We prefer median instead of mean since median will be resilient to pitch transitions such as slides and vibratos.

### 2.3 Pitch Histogram Computation

In LND method, we don't need to compute an overall pitch histogram since we have already obtained an initial estimation. Instead we compute a pitch histogram only from the frequency values in the pitch chunk, which are closer than 100 cents to our initial estimation. We use  $\frac{1}{3}$  Holderian comma<sup>5</sup> bin resolution, which is the smallest audible interval in Turkish makam music (Gedik & Bozkurt, 2010).

### 2.4 Tonic Frequency Refinement

After the pitch histogram is computed for the frequencies around our initial estimation, the peak points of histogram are calculated. We use *pypeaks*<sup>6</sup> library for peak computation. We identify the tonic frequency as the frequency

<sup>4</sup> <http://musicbrainz.org/recording/26da8cac-5757-4494-a214-25ad564fc292>

<sup>5</sup> 1 Holderian comma is equal to  $\frac{1200}{53} = 22.64$  cents

<sup>6</sup> <https://github.com/gopalkoduri/pypeaks>

Makam(TD1)	#	Form(TD1)	#	Makam(TD2)	#	Form(TD2)	#
Hicaz	135	Şarkı	375	Hicaz	146	Şarkı	412
Nihavend	87	Taksim	131	Hüzzam	95	Taksim	180
K.hicazkar	74	Peşrev	125	K.hicazkar	88	Peşrev	137
Hüzzam	64	Saz Semaisi	92	Nihavend	74	Saz Semaisi	99
Uşşak	62	Ayin-i Şerif	18	Uşşak	67	Ayin-i Şerif	30
Other (25)	446	Other (34)	127	Other (67)	623	Other (38)	235
Total	868		868	Total	1093		1093

**Table 1:** The makam and form statistics of the TD1 and TD2 datasets

value corresponding to the closest peak to our initial estimation.

### 3. EXPERIMENTS

To compare against our method, we have re-implemented the MHTM method (Gedik & Bozkurt, 2010). For pitch extraction, the predominant melody estimation methodology explained in (Atlı et al., 2014) is used. *L1-norm* is used as the distance measure for the *recording - template histogram* matching step, which was reported as the best distance measure for this task (Gedik & Bozkurt, 2010).

#### 3.1 Test Datasets

We compiled two test datasets from the Turkish makam music research corpus (Uyar et al., 2014) of CompMusic project (Serra, 2011)<sup>7</sup>. The datasets consist of the audio recordings, the related metadata (such as track names, performers, *makams* and *forms*) and the tonic annotations.

The first test dataset (*TD1*) is compiled for the comparison of the MHTM method and the LND method. Because of the limitations of the MHTM (Section 1), audio recordings in a single *makam* have been selected. TD1 includes 868 audio recordings in 30 *makams* which we selected as the most common *makams* in (Gedik & Bozkurt, 2010). For the evaluation of the LND method, test dataset 2 (*TD2*) is arranged. 225 polyphonic audio recordings where more than one *makam* is performed have been added on TD1 to create TD2. Only the LND method has been tested on this test dataset. TD2 includes 1093 audio recordings, more than 7000 of metadata entries. More than 3000 of these entries are culture-specific information (e.g. *makam*, *form*).

The tonic frequencies of the audio recordings have been manually annotated by the musicians. The annotated frequencies have been matched with the nearest peak point of the pitch histogram of related audio recordings to eliminate the intonation differences. These annotations are used as the ground-truth in the experiments.

Statistics of the test datasets are shown in Table 1. While the commercial recordings cannot be shared (due to the copyright issues), pitch tracks, metadata, and tonic annotations of the audio recordings are available in the Compmusic corpora<sup>8</sup>.

Methods	TD1			TD2		
	T	F	Success(%)	T	F	Success(%)
MHTM	626	242	72.1	-	-	-
LND	775	93	89.3	964	129	88.1

**Table 2:** The results of the experiments

#### 3.2 Evaluation and Results

To evaluate the methodologies, we compare the tonic annotations with the estimated tonic frequencies as explained in (Şentürk et al., 2013). If the estimated tonic frequency is within an interval of 1 Holderian comma around the annotated tonic frequency, the estimation is marked as true. Otherwise, the estimation is marked as false. In comparison, the octave differences between the annotated frequency and the estimated frequency are ignored due to ambiguities of the tonic octave in performances, where multiple instruments play the same melody in their own register. The results of experiments are shown in Table 2.<sup>9</sup> *T* and *F* refer number of true and number of false estimations, respectively.

#### 3.3 Discussion

The distribution of incorrectly identified tonic frequencies using MHTM method on TD1 is shown in Table 3. It is observed that the tonic frequencies of compound *makams* have been identified incorrectly. This is expected since the melodic progressions of at least two *makams* are represented in compound makams. The pitch distributions of the compound *makam* performances are more complicated in comparison to the pitch distributions of the simple *makam* performances. This situation complicates the tonic frequency estimation in audio recordings with compound makams using the MHTM method.

The results obtained from TD1 show that LND method outperforms the MHTM method. Moreover the results obtained from TD2 show that LND method can also be applied to performances where multiple makams are performed without any drawbacks. Currently, if two notes are smoothly connected with glissando, the LND method cannot divide these notes properly. If this happens in the last two notes, the LND method may identify the tonic frequency as the frequency of penultimate note. As a next step we plan to apply quantisation techniques, which are commonly used in the automatic transcription (Benetos &

<sup>7</sup> <http://compmusic.upf.edu/node/235>

<sup>8</sup> <http://compmusic.upf.edu/corpora>

<sup>9</sup> The complete results are available at <http://compmusic.upf.edu/node/265>

Makam	#	F	Err%
Ferahfeza	23	22	95.6
Acemaşiran	12	11	91.6
Hicazkar	30	22	73.3
Sultan-i Yegah	22	15	71.4
Kürdilihicazkar	45	28	60.8
Bestenigar	10	6	60.0
Rast	74	14	31.1

**Table 3:** The distribution of audio recordings for which the MHTM method fails to identify the tonic correctly in the TD1 dataset

Holzappel, 2013).

#### 4. CONCLUSION

In this paper, a new methodology for automatic tonic estimation of Turkish makam music recordings is presented. The methodology is based on detecting the last note of a performance (which is also the tonic note) and estimating its frequency. The new method is compared with a previous, more-complicated method. Experiments were performed on two test datasets, which include 1093 audio recordings in total. Our method estimates the tonic frequencies with 89.3% and 88.1% success rate on these datasets. The experiments show that our method outperforms the more-complicated method.

#### 5. ACKNOWLEDGMENTS

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

#### 6. REFERENCES

- Akdoğdu, O. (1989). *Taksim nedir?, Nasıl yapılır?* İhlas A.Ş.
- Atlı, H. S., Uyar, B., Şentürk, S., Bozkurt, B., & Serra, X. (2014). Audio feature extraction for exploring Turkish makam music. In *3rd International Conference on Audio Technologies for Music and Media*, Ankara, Turkey. Bilkent University. Faculty of Art, Design and Architecture. Department of Communication and Design.
- Benetos, E. & Holzappel, A. (2013). Automatic transcription of Turkish makam music. In *Proceedings of ISMIR - International Conference on Music Information Retrieval*, (pp. 355–360)., Curitiba, Brazil.
- Bozkurt, B. (2008). An automatic pitch analysis method for Turkish maqam music. *Journal of New Music Research*, 37(1), 1–13.
- Bozkurt, B. (2012). A system for tuning instruments using recorded music instead of theory-based frequency presets. *Computer Music Journal*, 36(3), 43–56.
- Bozkurt, B., Ayangil, R., & Holzappel, A. (2014). Computational analysis of Turkish makam music: Review of state-of-the-art and challenges. *Journal of New Music Research*, 43(1), 3–23.
- Chordia, P. & Şentürk, S. (2013). Joint recognition of raag and tonic in North Indian music. *Computer Music Journal*, 37(3).
- De Cheveigné, A. & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *Journal of Acoustical Society of America*, 111(4), 1917–1930.
- Gedik, A. C. & Bozkurt, B. (2010). Pitch-frequency histogram-based music information retrieval for Turkish music. *Signal Processing*, 90(4), 1049–1063.
- Gómez, E. (2006). *Tonal Description of Music Audio Signals*. PhD thesis, Universitat Pompeu Fabra.
- Gürmeriç, A. (1966). Lecture notes. İBK Türk Müziği Bölümü Talebe Cemiyeti Neşriyatı.
- Karaosmanoğlu, K. (2012). A Turkish makam music symbolic database for music information retrieval: SymbTr. In *Proceedings of 13th International Society for Music Information Retrieval Conference (ISMIR)*, (pp. 223–228)., Porto, Portugal.
- Krumhansl, C. L. & Shepard, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of experimental psychology: Human Perception and Performance*, 5(4), 579–594.
- Salamon, J. & Gómez, E. (2012). Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(6), 1759–1770.
- Şentürk, S., Gulati, S., & Serra, X. (2013). Score informed tonic identification for makam music of Turkey. In *Proceedings of 14th International Society for Music Information Retrieval Conference (ISMIR)*, (pp. 175–180)., Curitiba, Brazil.
- Şentürk, S., Holzappel, A., & Serra, X. (2014). Linking scores and audio recordings in makam music of Turkey. *Journal of New Music Research*, 43(1), 34–52.
- Serra, X. (2011). A multicultural approach in music information research. In *Proceedings of 12th International Society for Music Information Retrieval Conference (ISMIR)*, (pp. 151–156)., Miami, Florida (USA).
- Uyar, B., Atlı, H. S., Şentürk, S., Bozkurt, B., & Serra, X. (2014). A corpus for computational research of Turkish makam music. In *1st International Digital Libraries for Musicology Workshop*, (pp. 1–7)., London, UK.