AUTOMATIC TRANSCRIPTION OF TURKISH MAKAM MUSIC

Emmanouil Benetos
City University London
emmanouil.benetos.1@city.ac.uk

Andre Holzapfel
Boğaziçi University
andre.holzapfel@boun.edu.tr

ABSTRACT

In this paper we propose an automatic system for transcribing makam music of Turkey. We document the specific traits of this music that deviate from properties that were targeted by transcription tools so far and we compile a dataset of makam recordings along with aligned microtonal ground-truth. An existing multi-pitch detection algorithm is adapted for transcribing music in 20 cent resolution, and the final transcription is centered around the tonic frequency of the recording. Evaluation metrics for transcribing microtonal music are utilized and results show that transcription of Turkish makam music in *e.g.* an interactive transcription software is feasible using the current state-of-the-art.

1. INTRODUCTION

The process of deriving a description for music in the form of a graphical representation is referred to as transcription in musicology. The practical use of this process lies usually in simplifying the analysis of mainly melodic, rhythmic and harmonic properties of a piece. The discussion of a proper graphical representation is almost as old as (comparative) musicology [2]. It can be stated as common sense that the goal of the analysis and traits of the music should guide the choices towards obtaining a transcription.

Automatic music transcription (AMT) is the process of automatically converting an acoustic music signal into some form of musical notation and is considered to be an open problem in the music information retrieval (MIR) literature, especially for transcribing multiple-instrument polyphonic music [5, 10]. The vast majority of AMT systems are targeted for transcribing well-tempered Eurogenetic [1] music and typically convert a recording into a piano-roll or a MIDI file (cf. [8] for a recent review of AMT systems). Evaluation of AMT systems is also typically made using a quarter tone tolerance, as in the MIREX Multiple-F0 Estimation and Note Tracking Tasks [1].

---

[1] Term used to avoid the misleading dichotomy of Western and non-Western music. It was proposed by Prof. Robert Reigle (MIAM, Istanbul) in personal communication.

In Turkish makam music, it is an open discussion which amount of steps should be used to divide an octave in order to be able to adequately describe pieces in all existent modes (*makams*). Academic theory adopted the division of the octave into 53 equal tempered intervals, which results in an interval of 22.642 cent, referred to a Holderian comma (see [6] for a concise description). A Western staff notation enhanced to represent such finer resolution (see Section 2) is used by musicians to practice and memorize pieces. Therefore, it appears to be reasonable to target a similar representation for a transcription system for Turkish makam music.

In this work, a system for transcribing Turkish makam music is proposed; we compile a dataset containing microtonal ground-truth and modify a state-of-the-art multi-pitch detection algorithm [4] to the specific challenges of Turkish makam music. This way we can obtain a first insight into the difficulty of the task, and gain insight into necessary future development steps. This includes the question if different music demands the design of a radically different approach, or if enhancing existing software can be an appropriate choice.

We will start by explaining the basic challenges that this music poses for AMT systems, and give our motivations to pursue this task for Turkish makam music in Section 2. We then describe the music collection that was compiled and used for evaluation, along with evaluation measures in Section 3. Section 4 describes the automatic transcription system, and experimental results are presented in Section 5. In the final Section we discuss our results and propose a strategy for an improved transcription system.

2. CHALLENGES AND MOTIVATION

The makam in Turkish music is a modal framework for melodic development. It includes the notion of a scale, proposes certain modulations to other makams, and describes characteristics of melodic progression. While a comprehensive explanation of these concepts is out of the focus of this paper, we are going to emphasize certain characteristics of Turkish makam music that make it a challenging repertoire for automatic transcription. Regarding the relation between music and notation, four traits deserve closer attention:

1. The notation of Turkish makam music uses a system that is based on Western staff notation, which introduces additional accidentals that signify a subset of the intervals obtained by dividing a whole tone into
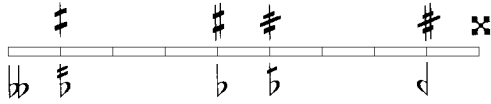
Figure 1. Visualization of the accidentals used in Turkish music. Only four of the possible eight intermediate steps that divide a whole tone are used.

nine equal steps (Holderian comma). This system comes with four additional flat and sharp accidentals respectively, see Fig. 1. However, as intensively discussed in music practice as well as in literature [6], this coarse grid does not properly reflect the intonation that is applied in practice. The knowledge about this correct intonation is part of the oral tradition.

2. Apart from the intonation of intervals, the fundamental frequency of a note depends on the transposition which a musician chooses. Only in one particular transposition the note a' is equal to $440Hz$, while there are a total of 12 transpositions to choose from.

3. A third particularity of Turkish music is that once a transposition is chosen, musicians in an ensemble choose an octave which suits the tonal range of their instrument. This can lead to up to three parallel melodies in octave intervals, while they are still considered to represent the (single) melody in the notation.

4. As fourth and final aspect, we have to emphasize the importance of augmenting the notes written in the score by various types of ornamentations and note additions. This is valid for pieces interpreted by a single musician, but also in the presence of several instruments. In the latter case, all musicians interpret a single melodic line, but are given a wide range of freedom to deviate from the simple representation in the score. This musical interaction is commonly referred to as *heterophony*.

These four traits of the relation between music and notation in Turkish makam music clarify that it poses some novel challenges to existing AMT systems. Current systems are usually focused on either polyphonic and monophonic sounds, and aim at a transcription into a representation with only 12 intervals per octave. The ability to extend systems beyond these specifications, and to tackle the above mentioned traits with an algorithmic approach can provide us with the means to transcribe performances of Turkish makam music. Such transcription is of great value for studying *e.g.* improvised performances of this music, or to study performance differences between different musicians.

It is worth to note that Turkish makam music is related to modal practice in many other cultures, in the sense that it shares the traits of microtonality and heterophonic performance. Its advantage for our studies is the availability of a large corpus of notation from more than 400 years. It

therefore represents an ideal point of entry into the development of automatic transcription approaches for modal music throughout the world.

3. MUSIC COLLECTION AND EVALUATION METRICS

3.1 Music Collection

The main instrumental forms of Turkish makam music, besides the improvised *taksim*, are the *Peşrev* and *Saz Semaisi*. Many notations of such pieces are publicly available in collections such as the one presented in [9], which contains notation with microtonal information in a machine-readable format (referred to as SymbTr format).

As our goal is to transcribe performances that use the most popular Turkish instruments, we decided to organize our collection according to that. As can be seen in Table 1, we chose five performances that use *tanbur*, a plucked string instrument, and five performances that contain *ney*, a rim-blown reed flute. Apart from solo performances, we also picked six ensemble performances that include various instruments. The horizontal lines in Table 1 divide between groups that represent different recordings of the same composition. We restricted ourselves to pieces that are available in the SymbTr collection [9], in order to use those notations as a starting point for the note-to-note alignment between notation and performance, which is a necessary element of the evaluation procedure for a transcription system (see Section 3.2).
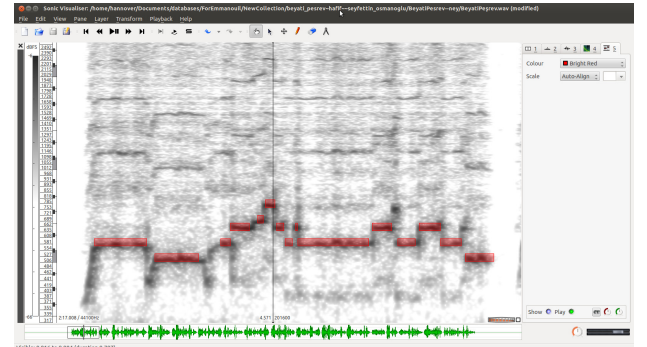

Figure 2. Screenshot of the manual alignment: the spectrogram for the first melodic phrase of the Beyati peşrev, with the aligned MIDI note events overlaid as red rectangles.

In order to obtain a note-to note alignment, we followed a semi-automatic approach. First, we determined if a performance deviates from the sequence of sections as found in the notation, and edited the SymbTr notation accordingly. This was necessary, as it is common to omit some units of a composition, or not to follow the repetitions as given in the notation. Then we converted the SymbTr representation to MIDI, and applied the approach presented in [11] in order to get a first estimate of the temporal alignment between the recording and the notes in the MIDI representation. The obtained aligned MIDI file was then loaded to the Sonic Visualiser [2] software as a notation layer

---

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

on top of the spectrogram of the recording, and the alignment was corrected manually; see Fig. 2 for an illustration. It is important to point out that the notations used in the Turkish tradition only outline the basic structure of the melody. The pieces are meant to be livened up by the performers using embellishments and adding notes in a way that still respects the notes in the score. That means that in most cases, the notes found in the score are present in the performance, but possibly with changed durations due to the additions of the performer. Therefore, we restricted the manual correction mainly to aligning the notes in the MIDI layer with the performance, and did not annotate the added performance features. This means that we did not attempt to obtain a descriptive transcription of the performance [13], but rather target at the correct recognition of the outline given in the score in our evaluations.

The alignment results in a list of MIDI notes with the correct temporal alignment to the performance. As a final step, we obtain the microtonal information for each MIDI note from our edited SymbTr notation, which includes a resolution equal to the Holderian comma (53 equal tempered steps per octave). All note values are given in *cent*, normalized such that the tonic of the piece obtains a value of $0 cents$. Which note represents the tonic in the notation is a given information when we assume the melodic mode (*makam*) to be known. This is a realistic assumption, as for most recordings the makam is an available metadata.

	Form	Makam	Instr.	Notes	Tonic/Hz
1	Peşrev	Beyati	Ensemble	906	125
2	Peşrev	Beyati	Ney	233	438
3	Saz S.	Hicazkar	Tanbur	706	147
4	Peşrev	Hüseyni	Ensemble	302	445
5	Peşrev	Hüseyni	Ensemble	614	124
6	Saz S.	Muhayyer	Ney	560	495
7	Saz S.	Muhayyer	Ensemble	837	294
8	Peşrev.	Rast	Tanbur	658	148
9	Peşrev	Rast	Ney	673	392
10	Peşrev.	Segah	Ney	379	541
11	Peşrev.	Segah	Ensemble	743	246
12	Saz S.	Segah	Ensemble	339	311
13	Saz S.	Segah	Tanbur	364	186
14	Saz S.	Uşşak	Tanbur	943	165
15	Saz S.	Uşşak	Tanbur	784	162
16	Saz S.	Uşşak	Ney	566	499

Table 1. Collection of recordings used for transcription.

Summing up, our ground truth annotation consists of a list of time instances and a note value in cent assigned to each time instance. In order to compare the output of our transcription system with such a ground truth, we need to estimate the tonic frequency of the recording and then normalize the estimated note values accordingly. We manually determined the tonic frequencies for all recordings, and list them along with other information of our collection

in Table 1. We also include the automatic tonic detection approach proposed by Bozkurt [6] into our system, and we monitor how its errors affect the transcription. The listed tonic frequencies illustrate the wide range between tanbur and ney registers, which causes the playing of the melody in different octaves by these instruments (see the trait list in Section 2). Our ground truth annotation consists of a total of 9607 notes, distributed into 2411 for ney, 3455 for tanbur and 3741 for ensemble pieces.

3.2 Evaluation Metrics

For assessing the performance of Turkish music transcription systems, we propose a set of transcription metrics which are based on the metrics used for the MIREX Note Tracking evaluations [3]. In onset-based transcription evaluation of Eurogenetic music, a note event is assumed to be correct if its F0 is within 50 cent of the ground-truth pitch and its onset within a +/-50 or +/-100ms tolerance. Likewise for onset-offset evaluation, a note is considered correct if its offset time is also within 20% of the ground-truth note's duration around its offset value, or within a tolerance of the ground-truth notes offset.

For the proposed evaluations, we consider a note to be correct if its F0 is within a +/-20 cent tolerance around the ground-truth pitch and its onset is within a 100ms tolerance. The +/-20 cent and +/-100ms tolerance levels are considered to be "fair margins for an accurate transcription" according to Charles Seeger [13]. Thus, we define the following onset-based Precision, Recall, and F-measure:

$$\mathcal{P}_{ons} = \frac{N_{tp}}{N_{sys}}, \quad \mathcal{R}_{ons} = \frac{N_{tp}}{N_{ref}}, \quad \mathcal{F}_{ons} = \frac{2\mathcal{R}\mathcal{P}}{\mathcal{R} + \mathcal{P}} \quad (1)$$

where N_{tp} is the number of correctly detected notes, N_{sys} the number of notes detected by the transcription system, and N_{ref} the number of reference notes. It should be noted that duplicate notes are considered as false alarms.

Even though onset-based evaluation is sufficient when transcribing recordings from pitched percussive instruments like the tanbur (where defining note offsets is generally an ill-defined problem), it can be useful performing onset-offset evaluation for pitched non-percussive instruments like ney. We thus consider a note event to be correct if, apart from the onset condition, its offset time is also within 20% of the ground-truth note's duration around its offset value, or within 100ms offset tolerance. We thus propose onset-offset Precision, Recall, and F-measure, denoted $\mathcal{P}_{off}$, $\mathcal{R}_{off}$, and $\mathcal{F}_{off}$, respectively, in the same way as in (1).

4. SYSTEM

The proposed transcription system takes as input a recording and information about the makam. Multi-pitch detection is performed using a shift-invariant model which was proposed in [4], modified for using ney and tanbur templates. Note tracking is performed as a post-processing step, followed by tonic detection. The final transcription output is a list of note events in cent scale centered around the tonic. A diagram of the proposed transcription system can be seen in Fig. 3.
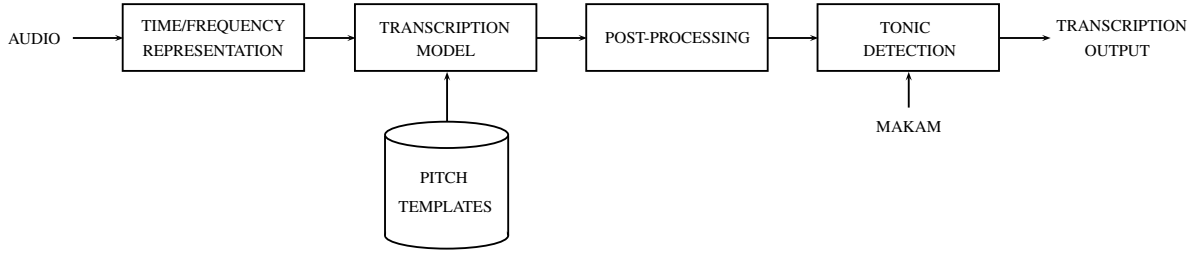
Figure 3. Proposed transcription system diagram.

4.1 Pitch Template Extraction

In systems for transcribing Eurogenetic music, pitch templates are typically extracted from isolated note samples (e.g. [7]). Since to the authors' knowledge such a database of isolated note samples for Turkish instruments does not exist, we extracted pitch templates using a dataset of 3 solo ney and 4 solo tanbur recordings.

Each note segment is identified and manually labeled, and the probabilistic latent component analysis (PLCA) method [14] with one component is employed per segment in order to extract a single spectral template per pitch. The time/frequency representation used is the constant-Q transform (CQT) with a spectral resolution of 60 bins/octave, with 27.5Hz as the lowest bin [12]. Since in log-frequency representations like the CQT the inter-harmonic spacings are constant for all pitches, templates for missing notes in the set were created by shifting the CQT spectra of neighboring notes. In the final collection of pitch templates, the range (in MIDI scale) for ney is 60-88 and the range for tanbur is 39-72.

4.2 Transcription Model

For performing multi-pitch detection, we employ the model of [4], which was originally developed for transcribing Eurogenetic music. The model expands PLCA techniques by supporting multiple templates per pitch and instrument source, as well as shift invariance over log-frequency; the latter is useful for performing multi-pitch detection in frequency resolution higher than MIDI scale.

The model takes as input a log-frequency spectrogram $V_{\omega,t}$ and approximates it as a joint distribution over time and log-frequency $P(\omega, t)$ (ω is the log-frequency index and t the time index). $P(\omega, t)$ is factored into $P(t)$ (spectrogram energy, which is a known quantity) and $P(\omega|t)$, which is modeled as:

$$P(\omega|t) = \sum_{p,s,f} P(\omega - f|s,p)P(f|p,t)P(s|p,t)P(p|t)$$
(2)

In (2), p is the pitch index in MIDI scale, s is the instrument source index, and f the pitch shifting factor (which accounts for frequency modulations or tuning deviations). Thus, $P(\omega|s,p)$ denotes the spectral template for instrument source s and pitch p, $P(f|p,t)$ is the time-varying log-frequency shift per pitch, $P(s|p,t)$ is the time-varying source contribution per pitch, and $P(p|t)$ the pitch activation over time. The shifting factor f is constrained to one
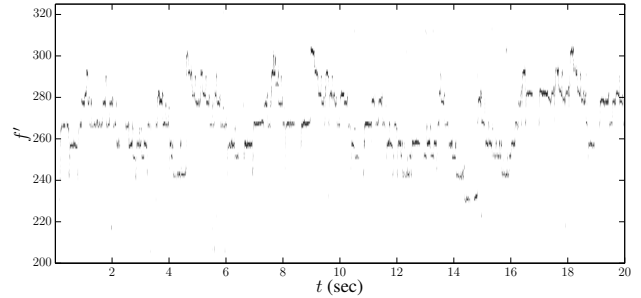

Figure 4. The time-pitch representation $P(f',t)$ for the 'Beyati Peşrev' piece performed by ney.

semitone range, thus with a CQT resolution of 60 bins/octave f has a length of 5.

The unknown model parameters $P(f|p,t)$, $P(s|p,t)$, and $P(p|t)$ are estimated using iterative update rules based on the Expectation-Maximization algorithm, and can be found in [4]. The spectral templates $P(\omega|s,p)$ are kept fixed using the pre-extracted pitch templates from Section 4.1 and are not updated. The number of iterations is set to 15. The output of the transcription model is a pitch activation matrix and a shifting tensor which are respectively given by:

$$P(p,t) = P(t)P(p|t) \tag{3}$$
$$P(f,p,t) = P(t)P(p|t)P(f|p,t) \tag{4}$$

By stacking slices of $P(f,p,t)$ for all pitch values, a time-pitch representation with 20 cent resolution, useful for pitch content visualization, can be created:

$$P(f',t) = [P(f,p_{low},t)\cdots P(f,p_{high},t)] \tag{5}$$

where f' denotes pitch in 20 cent resolution, $p_{low} = 39$ is the lowest pitch value, and $p_{high} = 88$ the highest pitch value considered. In Fig. 4 the time-pitch representation for a ney piece can be seen.

4.3 Post-processing

The output of the transcription model of Section 4.2 is a non-binary pitch activation matrix which needs to be converted into a list of note events, listing onset, offset, and pitch. In order to achieve that, we first perform median filtering on $P(p,t)$ and then perform thresholding on the activation, followed by minimum duration pruning (with a minimum note event duration of 130ms), in the same way as in [7].

Since a significant portion of the dataset consists of ensemble pieces where the tanbur and ney are performing in octave unison, we need to convert that heterophonic output of the multi-pitch detection algorithm into a monophonic output which will be usable as a final transcription. Thus, a simple 'ensemble detector' is created by measuring the percentage of octave intervals in the detected transcription. If the percentage is above 20%, the piece is considered to be an ensemble one. Then, for each ensemble piece each octave interval is processed by merging the note event of the higher note with that of the lowest one.

In order to convert a detected note event into the cent scale, information from the pitch shifting factor f is used. For each detected event with pitch p and for each time frame, we find the value of f that maximizes $P(f, p, t)$:

$$\hat{f}_{p,t} = \arg\max_f P(f, p, t) \qquad (6)$$

Then, the median of $\hat{f}_{p,t}$ for all time frames belonging to that note event is selected as the pitch shift that best represents that note event. Given the CQT resolution (60 bins/octave), the value in cent scale wrt the lowest frequency bin of the detected pitch is simply $20(\hat{f}-1)$, where $\hat{f}$ is the pitch index (in 20 cent scale) of the detected note.

4.4 Tonic detection

In order to detect the tonic frequency of the recording we apply the procedure described in [6]. It computes a histogram of the detected pitch values, and aligns it with a template histogram for a given makam using the cross-correlation function. Then the peak in the pitch histogram is assigned to the tonic which is closest to the peak of the tonic in the template, and all detected pitches are centered around this value. Finally, after centering the detected note events by the tonic, we eliminate note events that occur more than 1700 cents or less than -500 cents apart from the tonic, since such note ranges are rarely encountered in Turkish makam music.

5. RESULTS

Using the evaluation metrics defined in Section 3.2, transcription results using the proposed system with manually annotated tonic can be seen in Table 2, where the reported F-measure reaches 51.24% using 20 cent tolerance. It can be seen that the worst performance of $\mathcal{F}_{ons} = 47.35\%$ is reported for the subset of ensemble recordings, which is to be expected compared to the performance of the monophonic recordings. Also, results using automatically estimated tonic can be seen in Table 3, where there is a performance drop of about 10.5% in terms of F-measure. It should be noted that since the output of the transcription system is centered around the tonic and the F0 tolerance for the evaluation is 20 cent, even a slight tonic miscalculation might lead to a performance decrease. Using the dataset of Section 3, major tonic mis-estimations were observed for recordings 1, 5, and 10 (described in Table 1), leading to an F-measure for these recordings which is close to zero.

	$\mathcal{P}_{ons}$	$\mathcal{R}_{ons}$	$\mathcal{F}_{ons}$
Ney recordings	51.58%	52.67%	51.51%
Tanbur recordings	61.69%	49.66%	54.82%
Ensemble recordings	41.48%	56.28%	47.35%
All recordings	51.58%	52.85%	51.24%

Table 2. Transcription onset-based results using manually annotated tonic.

	$\mathcal{P}_{ons}$	$\mathcal{R}_{ons}$	$\mathcal{F}_{ons}$
Ney recordings	44.67%	43.98%	43.90%
Tanbur recordings	36.27%	45.13%	40.04%
Ensemble recordings	45.56%	33.21%	38.13%
All recordings	41.23%	42.07%	40.89%

Table 3. Transcription onset-based results using automatically detected tonic.

Regarding the impact of the F0 tolerance in the evaluation, Fig. 5 shows transcription results in F-measure using different F0 tolerance values, ranging from 10 cent to 50 cent. It can be seen that using 50 cent tolerance (*i.e.* the standard for evaluating Eurogenetic music transcription systems), the performance of the proposed system reaches $\mathcal{F}_{ons} = 61.58\%$. As far as the impact of the onset tolerance is concerned, Fig. 6 shows transcription results in F-measure using different onset tolerance values, ranging from 50ms to 150ms.

As far as onset-offset evaluation results using the ney recordings, the reported F-measure is 25.04%, while $\mathcal{P}_{off} = 25.12\%$ and $\mathcal{R}_{off} = 25.54\%$. Even though the onset-offset detection results might seem low, it should be stressed that estimating offsets (both manually and automatically) is an ill-defined problem and that similar results are reported in the MIREX Note Tracking evaluations [1].

The impact of several sub-components of the system can also be seen by disabling the 'ensemble detection' procedure, which leads to an F-measure of 45.08% for the ensemble pieces, which is more than a 2% decrease in performance. By removing the minimum duration pruning process (which was applied for detected note events less than 130ms), the reported F-measure with manually annotated tonic is 49.35%, which is a performance decrease of about 2%. Finally, by disabling the system sub-component which deletes note events that appear more than 1700 cents or less than -500 cents away from the tonic, the system performance drops to 47.42%; system decrease is more apparent in the ensemble pieces (which are performed in an octave unison, spanning a wider note range), leading to an F-measure of 40.60%.

6. DISCUSSION

Overall, it was shown that although the performance of automatic music transcription systems is still below that of a human expert, it is possible to produce fairly accurate transcriptions from Turkish music, which can be used as a basis for manually corrected scores. This can be verified
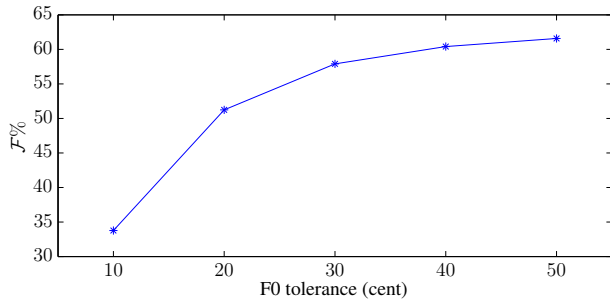
Figure 5. Onset-based transcription results in $\mathcal{F}$ with manually annotated tonic using different F0 tolerance.
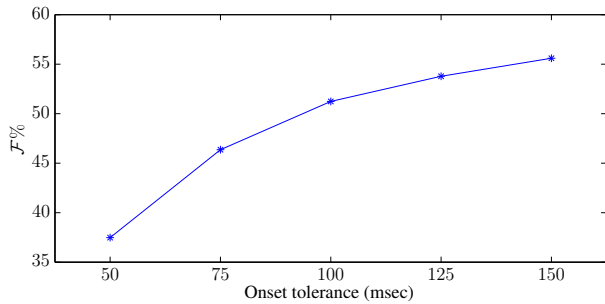

Figure 6. Onset-based transcription results in $\mathcal{F}$ with manually annotated tonic using different onset tolerance.

by comparing original recordings and MIDI transcriptions (without pitch bend), which can be found online [3] .

It was also shown that transcribing heterophonic music is an additional challenge for creating a system for transcribing Turkish makam music. On the one hand, one major drawback from automatic transcription systems is the fact that they produce octave errors, while on the other hand the multi-pitch output needs to be converted into a single melodic line for a makam score, which poses a different set of challenges.

One major issue with creating a spectrogram factorization-based transcription system for Turkish music is the lack of isolated note recordings for creating a training set; a set was created by segmenting some relatively clean solo recordings but we believe that an isolated sounds database would considerably improve the performance of the proposed system, especially in the tanbur case, where there are strong transient elements in tanbur tones. Another issue is the presence of percussion in makam pieces, which in certain cases resulted in false alarms in the transcriptions; again, a set of isolated percussive sounds could improve the performance of the proposed system. Apart from these template-related improvements, we intend to evaluate the system on vocal recordings and aim at including an improved tonic detection into the algorithm.

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] Music Information Retrieval Evaluation eXchange (MIREX). http://music-ir.org/mirexwiki/.

[2] O. Abraham and E. M. von Hornbostel. Propositions for the transcription of exotic melodies, (in german language). In *Sonderdruck aus "Sammelbände der Internationalen Musikgesellschaft" XI, 1*. Leipzig, Internationale Musikgesellschaft, n.d., 1909.

[3] M. Bay, A. F. Ehmann, and J. S. Downie. Evaluation of multiple-F0 estimation and tracking systems. In *ISMIR*, pages 315–320, 2009.

[4] E. Benetos and S. Dixon. A shift-invariant latent variable model for automatic music transcription. *Computer Music Journal*, 36(4):81–94, 2012.

[5] E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff, and A. Klapuri. Automatic music transcription: breaking the glass ceiling. In *ISMIR*, pages 379–384, 2012.

[6] B. Bozkurt. An automatic pitch analysis method for turkish maqam music. *Journal of New Music Research*, 37(1):1–13, 2008.

[7] A. Dessein, A. Cont, and G. Lemaitre. Real-time polyphonic music transcription with non-negative matrix factorization and beta-divergence. In *ISMIR*, pages 489–494, 2010.

[8] P. Grosche, B. Schuller, M. Müller, and G. Rigoll. Automatic transcription of recorded music. *Acta Acustica united with Acustica*, 98(2):199–215, March 2012.

[9] K. Karaosmanoğlu. A turkish makam music symbolic database for music information retrieval: Symbtr. In *ISMIR*, 2012.

[10] A. Klapuri and M. Davy, editors. *Signal Processing Methods for Music Transcription*. New York, 2006.

[11] R. Macrae and S. Dixon. Accurate real-time windowed time warping. In *ISMIR*, pages 423–428, 2010.

[12] C. Schörkhuber and A. Klapuri. Constant-Q transform toolbox for music processing. In *7th Sound and Music Computing Conference*, Barcelona, Spain, July 2010.

[13] C. Seeger. Prescriptive and descriptive music-writing. *Music Quarterly*, 64(2):184–195, 1958.

[14] P. Smaragdis, B. Raj, and M. Shashanka. A probabilistic latent variable model for acoustic modeling. In *Neural Information Processing Systems Workshop*, Whistler, Canada, December 2006.

---

[3] http://www.soi.city.ac.uk/~sbbj660/MakamTranscriptionSamples.zip