NAWBA RECOGNITION FOR ARAB-ANDALUSIAN MUSIC USING TEMPLATES FROM MUSIC SCORES

Niccolò Pretto
University of Padova, Padova, Italy
niccolo.pretto@dei.unipd.it

Bariş Bozkurt, Rafael Caro Repetto, Xavier Serra
Universitat Pompeu Fabra, Barcelona, Spain
baris.bozkurt@upf.edu
rafael.caro@upf.edu
xavier.serra@upf.edu

ABSTRACT

The Arab-Andalusian music is performed through *nawabāt* (plural of *nawba*), suites of instrumental and vocal pieces ordered according to their metrical pattern in a sequence of increasing tempo. This study presents for the first time in literature a system for automatic recognition of *nawba* for audio recordings of the Moroccan tradition of Arab-Andalusian music. The proposed approach relies on template matching applied to pitch distributions computed from audio recordings. The templates have been created using a data-driven approach, utilizing a score collection categorized into *nawabāt*. This methodology has been tested on a dataset of 58 hours of music: a set of 77 recordings in eleven *nawabāt* from the Arab-Andalusian corpus collected within the CompMusic project and stored in Dunya platform. An accuracy of 75% on the *nawba* recognition task is reported using Euclidean distance (L2) as distance metric in the template matching.

1. INTRODUCTION

This study targets automatic *nawba* [1] recognition for Arab-Andalusian music recordings, a task that has not been considered in any previous study. Our approach relies on template matching applied to pitch distributions computed from audio recordings to match with templates learned from a score collection categorized into *nawabāt* (plural of *nawba*).

Template matching is a widely used technique in the computer vision field where shape templates or color/brightness templates are used to match images or sub images [2]. It has also been used for various content-based music retrieval tasks since early days of music information retrieval. For example, in [3], template matching is proposed for audio retrieval using representations in the form of distributions computed from quantized MFCC vectors. Another example is the well-known N-gram approach for melody retrieval applied on symbolic data which also basically relies on matching distribution templates [4].

Following the highly influential study of Krumhansl and Kessler [5] on tonality perception, distribution template matching has also been largely used for automatic tonality or modality detection tasks. Note distributions [6] or pitch class distributions (using a twelve-tone equal temperament — 12-TET — pitch representation) are the features that are commonly used for matching [7, 8]. For many non-Western music cultures, 12-TET representation fails to represent the note/pitch space especially those explicitly characterized to be microtonal. For example, for Turkish *makam* music, the most widely used theory [9] proposes a non-tempered 24 tone system which is also found to be insufficient in representing all pitches by the master musicians of that culture. In [10], the author proposes the use of high resolution/dimensional pitch distributions for automatic tonic detection via matching transposed versions of distribution templates with recording pitch distribution. Later, in [11], the same approach was applied in an automatic *makam* recognition task. In [12], Chordia and Şentürk compared the use of 12-dimensional pitch class distribution versus high dimensional distributions and confirmed significant improvement with high dimensional distributions for tonic detection and *raag* recognition for North Indian classical music. In [13], Heydarian tested the same approach using various distribution resolutions for *dastgah* recognition for Iranian music and reported that 24-TET resolution is preferable.

This study targets testing the distribution template matching for *nawba* detection in Arab-Andalusian music context for the first time in literature. The study makes use of a database [14] collected within the CompMusic project [15] that contains audio recordings and corresponding scores, manually transcribed by an expert. Our approach differs from previous work (in high-resolution pitch distribution template matching) by the way templates are constructed. Data-driven approaches (like [11]) rely on a supervised classification approach where templates are learned from the training audio data. One basic difficulty in such approaches is the lack of tonic frequency information which is known to vary largely among different recordings in the same mode. To align distributions, tonic detection is performed which is not free of errors. One other way of template construction is the use of theoretical information. In [10], the author uses theoretical scale descriptions to synthesize pitch distribution templates for the automatic tonic detection task. Such a representation does not reflect dis-

[1] All the arabic terms in this paper have been transcribed according to the standard proposed in [1]

tributional characteristics of the actual mode but simply the scale with all scale degrees/notes assigned to have the same distribution. Here, we make use of available scores accompanying the audio collection to create templates for categories. Pitch class distributions computed from scores of the training data are used to create high resolution pitch distribution templates which are used in a template matching methodology for automatic *nawba* recognition. We have tested this approach on a dataset containing 77 recordings of the eleven *nawabāt* using cross-validation. An accuracy of 0.75 is reported using Euclidean distance (L2) as distance measure in the template matching.

In the following section, the Arab-Andalusian music is presented in order to delineate its main peculiarities. In section 3, the methodology is explained in depth. The experiment and the dataset used to assess this methodology are described in section 4. In section 5 and section 6 respectively, the results are discussed and further works are proposed.

2. ARAB-ANDALUSIAN MUSIC

Arab-Andalusian music is the term [2] given to the music tradition formed around the 12th Century in the Islamic territories of the medieval Iberian Peninsula known as Al-Andalus and that has been preserved to this day as a classical repertoire in several North African countries. Born as a result of the combination of Iberian local traditions with Arab elements, it acquired different personalities in each of the countries were it survived. In this paper we will focus on the Moroccan tradition, known as *al-Āla* [17]. The Arab-Andalusian music is performed through *nawabāt* (plural of *nawba*), suites of instrumental and vocal pieces ordered according to their metrical pattern in a sequence of increasing tempo. All the pieces contained in one *nawba* share the same *ṭāb'*, mode. Each *ṭāb'* is defined by an ascending and descending diatonic scale — no microtones are used [16,18] — built upon a fundamental degree, within a specific range, with several stressed degrees and a "persistent degree" (similar in function to a reciting tone), and a stock of characteristic melodic figures [18]. In the 18th Century, the scholar al-Haiek fixed the number of *nawabāt* for the Moroccan tradition to eleven (Table 1). Fragmentary pieces from other 15 *ṭubū'* were attached to eight of these *nawabāt*, according to the similarity of their modal character [17]. Consequently, all eleven *nawabāt* are defined by a main *ṭāb'*, which gives the name to the *nawba*, and eight of them also have a different number of related secondary or neighbor *ṭubū'* [16], giving a total of 26 *ṭubū'*. Arab-Andalusian music is performed by a mixed choir and an instrumental ensemble, including solo performances by either a vocalist or an instrumentalist. Some pieces are composed, while others are improvised, and all are performed in a monodic or heterophonic texture. Even though contemporary ensembles differ in their composition, they are mostly formed by string instruments such as as 'ūd, rabāb, qānūn, violin, viola, cello, double bass and piano, percussion instruments such as tar and darbuka, and occasionally also a clarinet.

Table 1. List of *nawabāt*.

Dunya ID	*Nawba* transliterated name
1	*al-istihlāl*
2	*al-isbahān*
3	*al-ḥiŷāz al-kabīr*
4	*al-ḥiŷāz al-māšriqī*
5	*al-rasd*
6	*al-'uššāq*
7	*al-māya*
10	*rasd al-ḏāyl*
11	*raml al-māya*
12	*'irāq al-'aŷam*
13	*garībat al-ḥusayn*

In this paper we propose a method for automatic *nawba* recognition based on their modal profile. Although some *nawabāt* contain more than one *ṭāb'*, all the *ṭubū'* in a *nawba* share common modal characteristics, each *nawba* presents a unique set of *ṭubū'*, no *ṭāb'* is performed in more than one *nawba*, and in any case, the performance of each *nawba* is dominated by the main *ṭāb'*. Therefore, we argue that even for those *nawabāt* containing secondary *ṭubū'* a unique modal template can be defined.

3. METHODOLOGY

In this work, we propose a novel approach to *nawba* recognition by using templates obtained from the music scores. The core idea of this work is to compare pitch distribution of an Arab-Andalusian recording with several templates in order to discover the *nawba* to which this recording belongs. The initial data necessary to use this methodology are several scores for every *nawba* in order to build the templates and audio recordings for which the *nawba* is unknown. From each score, a pitch class distribution in total duration is computed. These distributions are also folded by considering an interval of twelve semitones (one octave). The resulting twelve bar distributions for recordings of a same *nawba* are averaged and normalized to a total sum of 1.

In the next step, the templates are synthesized from the pitch class distribution using a Gaussian curve for each value of a distribution. To obtain a normalized distribution comparable to the pitch distribution of a single recording, the values of a bar p is considered as the area of each Gaussian curve $g(x)$ and the following formula is used in order to calculate the area under the curve:

$$p = \int_{-\infty}^{\infty} g(x)dx = \int_{-\infty}^{\infty} ae^{\frac{-(x-b)^2}{2c^2}}dx = a\sqrt{2\pi c^2}$$

From this formula, the variable a is obtained as follows:

$$a = p\sqrt{2\pi c^2}$$

where c is the standard deviation (in the experiments considered in a range between 20 and 40 cents). The variable

[2] For a discussion about the terminology for referring to this music tradition, please refer to [16]

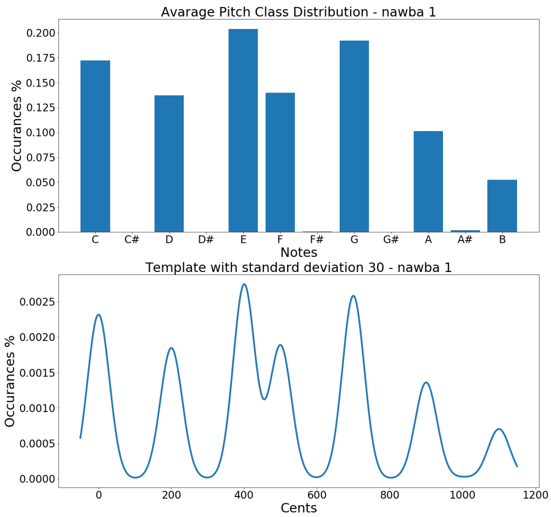

Figure 1. An example of average folded pitch class distribution for *nawba al-istihlāl* and the corresponding template synthesized by using a value of standard deviation equal to 30 cents.

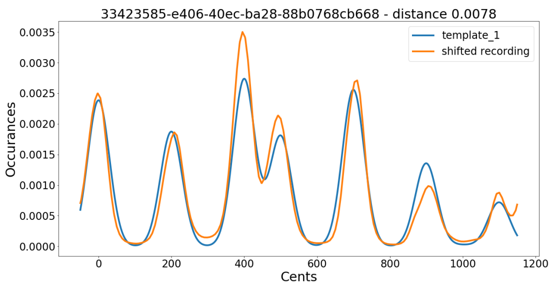

Figure 2. Example of comparison between a recording pitch distribution and the template of *nawba al-istihlāl*.

b, the average value and center of the curve, is positioned in relation to the disposition of the distribution with intervals of 100 cents. Figure 1 shows an example of resulting template for *nawba al-istihlāl*. This template is normalized to 1 and is compared with the pitch distribution of a single track.

The pitch distribution of a recording is computed from the fundamental frequency series extracted using an algorithm originally developed to analyze Turkish *makam* music [19]. To study the quality of the pitch estimation, we have visually inspected for various sound samples plots of pitch series together with spectrograms. The recordings involve four categories of signals: vocal-only improvisation, single instrument improvisation, heterophonic multi-instrumental performance with or without vocal. We observed that the pitch estimation quality is high in most parts of the audio with occasional octave errors mainly during the low-pitch instrumental improvisation (e.g. 'ūd). From the pitch profile, the pitch distribution is extracted. The pitch distributions of the recordings are computed using a 7.5 cents resolution and smoothed using a Gaussian kernel with standard deviation of 7.5 cents. As a result, the distributions computed for these long recordings (average duration of 45 minutes) are highly smooth. Furthermore, the pitch distribution is also folded in an interval of one octave. This folding operation requires a reference value used as origin and usually the tonic frequency is the most commonly used for this kind of operation. The tonic frequency is unknown, so the frequency of the maximum peak of the distribution is used as origin. Nevertheless, the choice of the origin doesn't affect the algorithm since template matching involves rolling the pitch distribution completely in one octave.

With the aim to find the best match with a template, the pitch distribution is shifted and the distance is calculated at each shift. The minimum distance refers to best match. In Figure 2, an example of good matching between a *nawba* template and the pitch distribution of a recording is shown.

4. EXPERIMENTS AND DATASET

The experiment to assess the methodology delineated in the previous section is based on a dataset of 77 long recordings corresponding to more than 58 hours of music. This is a subset of the Arab-Andalusian Corpus [14] collected in Dunya [20]. Dunya platform comprises the music corpora gathered in the CompMusic project [15] for five music traditions, and offers access to their data, metadata and annotations. Furthermore, this platform provides a web-based graphical user interface and an API to access to the contents. With the last one, we retrieved all the data and the metadata of the Arab-Andalusian corpus. For every recording, this corpus contains the mp3, the related metadata and also the complete transcription in the XML format, essential to compute the pitch class distribution from scores. As reported in [14], all the transcriptions and the labelling of the *nawba* are provided by an expert musicologist specialized in this genre of music. This complete set of data and metadata is essential to obtain a reliable ground truth for the experiment. In order to maintain the relation with the Dunya platform, we preferred to use the reference identifiers for referring *nawba* provided by Dunya API, as can be seen in Table 1.

The dataset is equally distributed across *nawabāt*. For each of the eleven *nawabāt* there are seven representative recordings of an average duration of 45 minutes. We consider this number of recordings for each *nawba* in order to obtain a balanced dataset: for some *nawabāt*, the corpus contains only seven tracks. The experiment divides the dataset in two stratified random subsets composed as follows: for all the *nawabāt*, the scores of six recordings are selected to train the templates, the remaining track is part of the test set. In this way the templates are completely independent form the test recordings. The experiment was repeated seven times: each time a different recording for every *nawba* was chosen. With this method, each recording was tested once.

As explained in the previous section, the standard deviation value, that characterizes the width of the "bell" of the Gaussians functions, strongly affects the performance. The

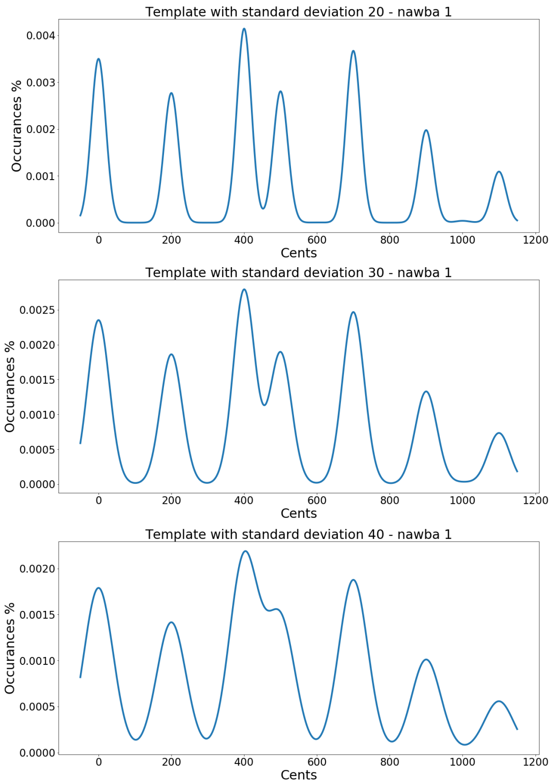

Figure 3. Three templates for *nawba al-istihlāl* synthetized using respectively 20, 30 and 40 cents as standard deviation value.

Distance measure	Standard Deviation		
	20 cents	30 cents	40 cents
City-Block (L1)	0.727	0.740	0.688
Euclidean (L2)	0.740	0.753	0.688
Correlation	0.377	0.390	0.390
Canberra	0.377	0.701	0.571

Table 2. *Nawba* recognition performance in the experiment

tested values of standard deviation are three: 20, 30 and 40 cents. Figure 3 shows clearly how this choice changes the shape of the templates.

Another factor that highly affects the performance is the distance metric used to match the pitch distribution with the templates. The distance metrics tested in the experiment are: City-Block (L1), Euclidean (L2), the inverse of the correlation and Canberra.

In order to easily reproduce the experiment, a Jupyter Notebook with the python source code is openly shared [3]. Furthermore, all the scores and the pitch distributions are downloadable in the same repository.

5. RESULTS

The overall results are provided in Table 2 and can be considered as a good starting point for this task with 11 classes. The best performance is obtained with the Euclidean (L2) distance measures and templates synthesized using standard deviation equal to 30 cents. In general, this distance metric results in higher performance than the others. Only with standard deviation 40 cents, Euclidean (L2) and City-Block (L1) distance have equal results. Considering the overall performance the standard deviation 20 and 30 cents, Euclidean (L2) distance results in higher performance for *nawba* recognition. However, the overall behavior of City-Block (L1) is similar to Euclidean (L2). Correlation leads to the lowest average performance, and seems not suitable for this kind of analysis. Performance observed for the Canberra distance is highly affected by the standard deviation with which the templates had been developed, leading to good results only when the standard deviation is set at 30 cents.

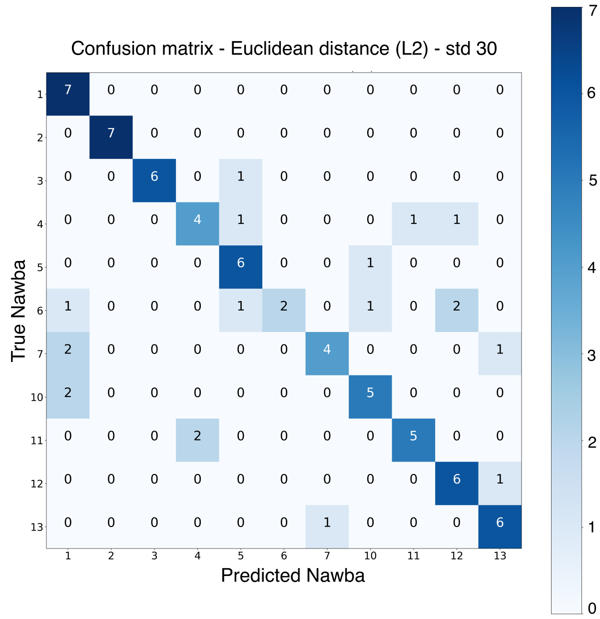

Figure 4. Overall confusion matrix for the experiments with Euclidean distance and 30 cents as standard deviation value.

Figure 4 shows the overall confusion matrix obtained by summing all the results of the seven folds for the best combination of standard deviation (30 cents) and distance metric (L2). Considering the overall result of 75% of recognized *nawabāt*, the distribution of the majority of the values in the diagonal was expected. In general, the values outside the diagonal seem not have a precise pattern. The worst results are obtained for *al-'uššāq* (with Dunya id 6), that is the only *nawba* with performance lower than 50%. Figure 5 shows an example of mismatch for a recording of this *nawba*.

All the results and the plots concerning the best experiment are available in the GitHub repository.

6. CONCLUSIONS

In this paper, we have presented for the first time in literature a system for automatic recognition of *nawba* for music recordings of the Moroccan tradition of Arab-Andalusian music. We followed a template matching strategy which

[3] https://github.com/MTG/andalusian-corpus-notebooks

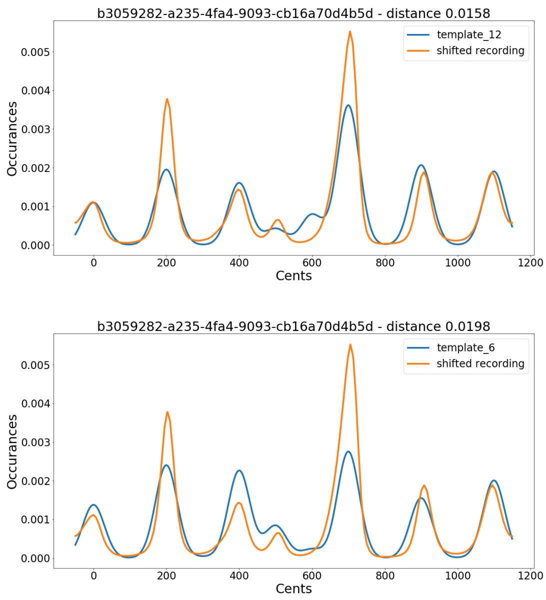

Figure 5. Example of incorrect recognition of the *nawba*. The track is recognized as belonging to *nawba* 12, but the correct *nawba* is 6.

has been previously used in mode recognition tasks for other music cultures successfully and we have reported its efficiency for our particular task. In our study, the templates have been created using a data-driven approach utilizing the score collection.

Being one of the first computational analysis on Arab-Andalusian music, this work targets promoting further studies on this music culture and for that aim shares all its data (including audio, metadata and scores) and code resources. As future works, we plan to consider automatic tonic frequency detection and *ṭāb'* recognition tasks.

Acknowledgments

The authors would like to acknowledge Mr. Amin Chaachoo, expert musicologist and experienced musician of this tradition, main curator of the Arab-Andalusian corpus and author of the music transcriptions, for his help with musicological explanations and transcription of Arab terms. This work was carried out with the support of the Musical Bridges project, funded by RecerCaixa.

7. REFERENCES

[1] M. del Amo, “Sistema de transliteración de estudios Árabes contemporáneos. Universidad de Granada,” *Miscelánea de Estudios Árabes y Hebraicos. Sección Árabe-Islam*, vol. 51, pp. 355–359, 2002.

[2] R. Brunelli, *Template matching techniques in computer vision: theory and practice*. John Wiley & Sons, 2009.

[3] J. T. Foote, “Content-based retrieval of music and audio,” in *Multimedia Storage and Archiving Systems II*, vol. 3229. International Society for Optics and Photonics, 1997, pp. 138–148.

[4] A. Uitdenbogerd and J. Zobel, “Melodic matching techniques for large music databases,” in *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*. ACM, 1999, pp. 57–66.

[5] C. L. Krumhansl and E. J. Kessler, “Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys.” *Psychological review*, vol. 89, no. 4, p. 334, 1982.

[6] K. Ng, R. Boyle, and D. Cooper, “Automatic detection of tonality using note distribution,” *Journal of New Music Research*, vol. 25, no. 4, pp. 369–381, 1996.

[7] D. Temperley and E. W. Marvin, “Pitch-class distribution and the identification of key,” *Music Perception: An Interdisciplinary Journal*, vol. 25, no. 3, pp. 193–212, 2008.

[8] P. Chordia and A. Rae, “Raag recognition using pitch-class and pitch-class dyad distributions.” in *ISMIR*. Citeseer, 2007, pp. 431–436.

[9] H. S. Arel, *Türk mûsıkîsi nazariyatı dersleri*. Kultur Bakanligi, 1991.

[10] B. Bozkurt, “An automatic pitch analysis method for Turkish maqam music,” *Journal of New Music Research*, vol. 37, no. 1, pp. 1–13, 2008.

[11] A. C. Gedik and B. Bozkurt, “Pitch-frequency histogram-based music information retrieval for Turkish music,” *Signal Processing*, vol. 90, no. 4, pp. 1049–1063, 2010.

[12] P. Chordia and S. Şentürk, “Joint recognition of raag and tonic in North Indian music,” *Computer Music Journal*, vol. 37, no. 3, pp. 82–98, 2013.

[13] P. Heydarian, “Automatic recognition of Persian musical modes in audio musical signals,” Ph.D. dissertation, London Metropolitan University, 2016.

[14] M. Sordo, A. Chaachoo, and X. Serra, “Creating corpora for computational research in Arab-Andalusian music,” in *Proceedings of the 1st International Workshop on Digital Libraries for Musicology*. ACM, 2014, pp. 1–3.

[15] X. Serra, “The computational study of a musical culture through its digital traces,” *Acta Musicologica*, vol. 89, no. 1, pp. 24–44, 2017.

[16] C. Poché, *La música arábigo-andaluza*. Móstoles: Akal, 2005.

[17] A. Chaachoo, *La música andalusí al-Ála*. Córdoba: Almuzara, 2011.

[18] ——, *La musique hispano-arabe, al-Ala*. Paris: L'Harmattan, 2016.

[19] H. S. Atlı, B. Uyar, S. Sentürk, B. Bozkurt, and X. Serra, "Audio feature extraction for exploring Turkish makam music," in *3rd International Conference on Audio Technologies for Music and Media, Ankara, Turkey*, vol. 12, no. 11, 2014, p. 2014.

[20] A. Porter, M. Sordo, and X. Serra, "Dunya: A system for browsing audio music collections exploiting cultural context," in *14th International Society for Music Information Retrieval Conference (ISMIR); 2013 Nov 4-8; Curitiba, Brazil: ISMIR; 2013. p. 101-6.* International Society for Music Information Retrieval (ISMIR), 2013.