2016-6

Detection of Melodic Patterns in Automatic Transcriptions of Flamenco Singing

Aggelos Pikrakis
University of Piraeus, pikrakis@unipi.gr

Nadine Kroher
University of Seville, nkroher@us.es

José-Miguel Díaz-Báñez
University of Seville, dbanez@us.es

DETECTION OF MELODIC PATTERNS IN AUTOMATIC TRANSCRIPTIONS OF FLAMENCO SINGING

Aggelos Pikrakis
University of Piraeus, Greece
pikrakis@unipi.gr

Nadine Kroher, José-Miguel Díaz-Báñez
University of Seville, Spain
nkroher@us.es, dbanez@us.es

ABSTRACT

The spontaneous expressive interpretation of melodic templates is a fundamental concept in flamenco music. Consequently, the automatic detection of such patterns in music collections sets the basis for a number of challenging analysis and retrieval tasks. We present a novel algorithm for the automatic detection of manually defined melodies within a corpus of automatic transcriptions of flamenco recordings. We evaluate the performance on the example of five characteristic patterns from the *fandango de Valverde* style and demonstrate that the algorithm is capable of retrieving ornamented instances of query patterns. Furthermore, we discuss limitations, possible extensions and applications of the proposed system.

1. INTRODUCTION

Flamenco is a rich music tradition from the southern Spanish province of Andalucıa. Having evolved from a singing tradition, the vocal melody remains the main musical element, accompanied by the guitar, rhythmical hand-clapping and dance. Gómez et al. (2016) mention, among others, the frequent appearance of glides and protamenti, sudden dynamic changes in volume and a small pitch range of less than an octave, as key characteristics of the flamenco singing voice. For a more detailed description of the genre, we refer to Gómez et al. (2014) and Gómez et al. (2016).

Flamenco singing is largely improvisational, in particular with respect to melody: during a performance, a melodic skeleton or a set of prototypical patterns are subject to spontaneous ornamentation and variation. Consequently, the automatic detection of modified instances of a given melodic sequence is a crucial step to a number of music information retrieval tasks. For example, most characteristic melodies are uniquely bound to a particular singing style. Consequently, detected melodic patterns provide important indications towards the style of an unknown recording. Furthermore, flamenco recordings often contain various songs and the location of pattern occurrences can assist the structural segmentation of a song. Moreover, the occurrence of common melodic patterns across tracks is crucial to characterising similarity among melodies which exhibit structural differences (Volk & van Kranenburg, 2012).

Given the absence of musical scores, related approaches in the context of flamenco (Pikrakis et al., 2012) but also in other non-Western oral music traditions (Gulati et al., 2014) have focused on the retrieval of melodic patterns from the fundamental frequency (f0) contour. The high degree of detail of this representation does not only increase computational complexity but is also prone to errors arising from micro-tonal ornamentations. In this study, we present a novel approach which operates on symbolic representations obtained from an automatic transcription system (Kroher & Gómez, 2016).

We provide a detailed technical description of the method in Section 2. The experimental setup is described in Section 3 and results are given in Section 4. We conclude the paper in Section 5.

2. METHODOLOGY

The core of our method is a modification of the well known Needleman-Wunsch (NW) algorithm (Needleman & Wunsch, 1970) from the area of bioinformatics. The NW algorithm was proposed as a global alignment method of molecular sequences. The term global alignment refers to the fact that when two sequences of discrete symbols are being matched, the objective is to align them from the beginning to the end, without omitting parts around the endpoints. During the alignment procedure, gaps are allowed to be formed. In the original NW formulation gaps are not penalized. Given two sequences of discrete symbols, the original NW algorithm can be formulated as a dynamic programming method that creates a dot matrix and finds the best path of dots on it, i.e., a path of dots (nodes) of increasing index that accumulates the largest score (number of dots). The dot matrix (also known as similarity grid) is formed by placing one pattern on the x-axis and the other one on the y-axis. An element of the dot grid is set equal to "1" if the symbols corresponding to its coordinates coincide.

The problem that we are dealing with in this paper cannot be treated as a global alignment task because our goal is to detect occurrences of a pattern in a significantly longer stream of notes. We are therefore proposing a modification of the NW algorithm, that preserves its fundamental characteristics and adds the capability to retrieve a ranked list of subsequences from an automatic transcription. Each retrieved result aligns, in some optimal sense, with the given prototype pattern. The novelty of our aprroach lies in the fact that it introduces a systematic way to: **(a)** extract iteratively occurrences of the reference pattern, ranked with respect to similarity score, **(b)** embed endpoint constraints in the NW method, **(c)** ensure invariance to key changes because the alignment takes place on the sequences of intervals derived from the pitch sequences that are being matched, and, **(d)** formulate transition costs between nodes of the

similarity grid as a function of intervalic differences. At a first stage, the proposed method operates on pitch sequences only, ignoring note durations. At a second stage, the results are refined by removing alignments that correspond to excessive local time-stretching. In the rest of this paper, we will use the abbreviation *mNW* for the proposed method.

In order to describe *mNW*, let $A = \{a_i; i = 1, 2, \ldots, I\}$ and $P = \{p_j; j = 1, 2, \ldots, J\}$ be the pitch sequences of the automatic transcription and the search pattern, respectively, where the a_i's and p_i's are pitch values in some symbolic (MIDI-like) format. We therefore ignore note durations at this stage. Sequence P is manually defined and reflects our musicological knowledge of the pattern to be detected. For example, pattern "A" of our experimental setup (Section 3) is represented by the following sequence of MIDI values:

$$\{64, 67, 65, 64, 67, 65, 65, 64, 62, 60, 58, 57\}$$

We now define that,

$$\delta_P(j_2, j_1) = p_{j_2} - p_{j_1}, j_2 > j_1,$$

is the music interval formed between the j_1-th and j_2-th note (pitch value) of the prototype pattern, which are not necessarily adjacent, and, similarly

$$\delta_A(i_2, i_1) = a_{i_2} - a_{i_1}, i_2 > i_1,$$

is the music interval formed between the i_1-th and i_2-th note (pitch value) of the automatically generated transcription. Therefore, the proposed *mNW* algorithm seeks a subsequence (chain) of a_i 's, of increasing index (not necessarily adjacent), such that the resulting sequence of intervals matches in some optimal scoring sense, a sequence of intervals formed by a subsequence of p_i 's of increasing index (also not necessarily adjacent).

To solve this problem from a dynamic programming perspective, A is placed on the vertical axis and P on the horizontal one, forming a scoring grid, S. Let

$$(i, j), i = 1, 2, \ldots, I, j = 1, 2, \ldots, J$$

be a node on this grid, which aligns the i-th note of A with the j-th note of P, and let $S(i, j)$ be the respective accumulated alignment score. The grid is initialized by setting the elements of the last row and column of the grid equal to zero, i.e., $S(I, j) = 0, j = 1, 2, \ldots, J$ and $S(i, J) = 0, i = 1, 2, \ldots, I$.

We then proceed row-wise, decreasing the row index and examining the nodes of each row at decreasing column index, which stands for a standard zig-zag scanning procedure. The accumulated score, $S(i, j)$, at node (i, j), where $i < I$ and $j < J$ is computed as follows:

$$h = \max\{S(i+1, k) + \gamma(\delta_A(i+1, i), \delta_P(k, j)); \\ k = j+1, \ldots, j+G_h\}, \quad (1)$$

$$v = \max\{S(m, j+1) + \gamma(\delta_A(m, i), \delta_P(j+1, j)); \\ m = i+1, \ldots, i+G_v\}, \quad (2)$$

$$S(i, j) = \max\{h, v\}, \quad (3)$$

where parameters G_h and G_v are positive integers that define the search radius for successors on the horizontal and vertical axis, respectively, and function $\gamma(.)$ is defined as:

$$\gamma(x, y) = \begin{cases} 1, if\ x = y, \\ -1, if\ \mid x - y \mid = 1, \\ -\infty, if\ \mid x - y \mid > 1, \end{cases}$$

The first two equations impose that the best successor of node (i, j) resides either on the next row (the $(i + 1)$-th row) or on the next column (the $(j + 1)$-th column). Parameters G_h and G_v control the horizontal and vertical gap length, respectively. In other words, they control how many pitch values can be skipped horizontally or vertically when searching for the best successor of the node. Function γ rewards equal intervals with a score equal to $+1$, penalizes with -1 any pair of intervals that differ by one semitone and forbids intervalic differences larger than a semitone to take place, hence the $-\infty$ penalty. After a node has been processed, the coordinates, (i_B, j_B), of its best successor, are stored in a separate matrix, Ψ, where $\Psi = \{\psi(i, j) = (i_B, j_B); i = 1, \ldots, I, j = 1, \ldots, J\}$.

After the whole grid has been scanned, the highest accumulated score on the first E_1 columns is selected and forward tracking on matrix Ψ reveals the best alignment path. However, this path will be rejected if it does not end in one of the last E_2 columns of the grid. Therefore, parameters E_1 and E_2 stand for the endpoint constraints of the alignment procedure, i.e., we permit that at most $E_1 - 1$ and $E_2 - 1$ notes are omitted from the left and right endpoints of the prototype pattern, respectively. If a path is rejected, we repeat from the second highest score until a valid path is detected or until all nodes of the first E_1 columns have been processed as candidate starting points of the best path. Obviously, if we want the algorithm to return two pattern occurrences, the procedure will be repeated until a second path is revealed, and, of course, this can be readily extended to address any number of desired occurrences.

Table 1 presents the best alignment result between pattern A of the experimental setup and a Valverde transcription. In this example, two notes are skipped from the automatically generated transcription (5th and 10th note from the first column) and this is shown with one inserted gap (symbol "-") per deleted note in the second column, in the respective rows. It is also worth observing that the matched subsequences are performed in different keys.

The example is further illustrated in Figure 1, where the dotted lines connect aligned notes and the two black notes are the ones that have been skipped on the automatic transcription sequence.

After the first processing stage has been completed, the obtained results are subsequently filtered at a second stage. More specifically, in order to restrain note duration variability, we compute the sequence of inter-onset differences of the notes of a formed path on both axes and discard any path for which at least two ratios of aligned inter-onset durations exceed a predefined stretching threshold (equal to 3 or $1/3$ in our study). This is equivalent to imposing, at a

Table 1: Best alignment result of pattern A against an automatically generated Valverde transcription: symbol "-" marks a skipped note (gap insertion).

transcription		query pattern (A)	
pitch	duration	pitch	duration
60	0.28	64	0.50
63	0.32	67	0.50
61	0.15	65	0.50
60	0.32	64	0.50
59	0.13	-	-
63	0.25	67	0.48
61	0.18	65	0.50
61	0.68	65	1.00
60	0.22	64	0.50
60	0.62	-	-
58	0.16	62	0.12
56	0.17	60	0.15
54	0.17	58	0.14
53	0.19	57	0.61

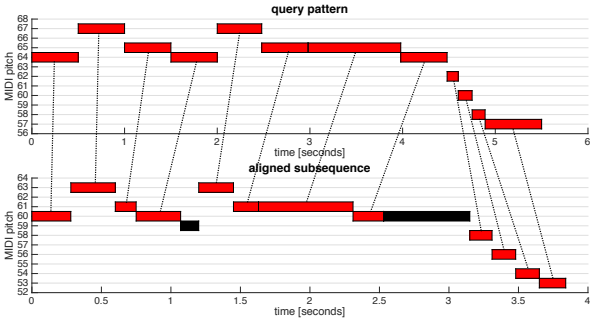

Figure 1: Illustration of the alignment shown in Table 1.

post-preocessing stage, a local time-warping threshold.

3. EXPERIMENTAL SETUP

We demonstrate the performance of the proposed algorithm in a query-by-example task. We aim at detecting occurrences of manually annotated MIDI sequences in a corpus of automatic transcriptions of polyphonic flamenco recordings. In this study, we focus on *fandangos de Valverde* (FV), a singing style belonging to the family of the *fandangos* (Kroher et al., 2016).

Like most *fandangos*, the *fandangos de Valverde* are bimodal in a structural sense (Fernández-Marín, 2011): solo guitar sections are set in *flamenco mode*, a scale with the diatonic structure of the Phrygian scale but with the dominant and sub-dominant located on the second and third scale degree, respectively (Figure 2). Singing voice sections are set in major mode and modulate only in the last verse back to *flamenco mode*.

Having evolved from Spanish folk tunes, songs belonging to this style are based on a particular melodic skeleton which, during interpretation, is subject to melodic and rhythmic modifications in terms of an expressive performance. The skeleton is composed of five distinct patterns (Figure 3) which occur in the form A-B-A'-C-A-D (where A' refers to a variant of A).

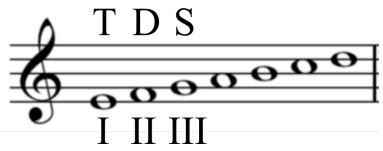

Figure 2: The flamenco mode: The tonic is located on the first, the dominant on the second and the sub-dominant on the third scale degree.

In this study, we use as query patterns manual transcriptions of the five phrases constituting the *fandango de Valverde* skeleton (Figure 3) and aim to retrieve their ornamented and modified occurrences in automatic transcriptions of real performances. To this end, we gathered a collection of 20 *fandangos de Valverde* taken from commercial recordings. The *cante100* dataset (Kroher et al., 2016) was added as noise to the collection: The contained 100 accompanied flamenco recordings cover a variety of singing styles and serve as a representative sample of flamenco music. None of the tracks contained in the *cante100* dataset belong to the *fandangos de Valverde* style. For each of the 120 tracks of the resulting collection we generated an automatic note-level transcription of the vocal melody using the algorithm described by Kroher & Gómez (2016).

The retrieved results are evaluated by means of the precision of the top 5 (P@5) and top 10 (P@10) ranking. A query result is considered relevant if its origin is a *fandangos de Valverde* recording and the detected melodic sequence corresponds to the query phrase.

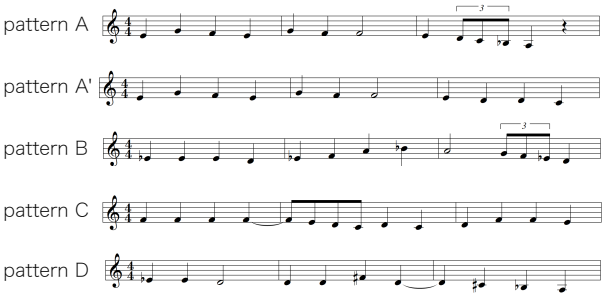

Figure 3: MIDI representations of the query patterns.

4. RESULTS

Table 2 gives the quantitative evaluation of all five query patterns and the top 5 results for pattern A are shown in Figure 4. It can be seen that the percentage of relevant melodic sequences in the top ranked results is significantly higher for patterns A, A' and B compared to patterns C and D. In particular, for patterns A' and B, all of the 5 highest ranked results are relevant with respect to the query, while for pattern D only one relevant result is retrieved.

A reasonable explanation for this behaviour is related to the amount of variation a pattern is subjected to during

performance: Pattern D, referred to as *caída* in flamenco terminology, constitutes the end phrase and, at the same time, the musical "highlight" of the interpretation. During this phrase, the melody modulates from major mode to flamenco mode and resolves in the Andalusian cadence. Consequently, singers tend to apply more expressive resources, which result in a higher performance variance. Within a lesser extent, the same applies to pattern C, where a high degree of ornamentation, in particular prolongation through a sequence of grace notes, tends to appear during the last two bars. Four examples of manual MIDI transcriptions of *caídas* are shown in Figure 5 in order to highlight observed performance variation, free of possible transcription errors. Furthermore, automatic transcriptions are particularly prone to errors in the end of the singing voice section, since the guitar accompaniment tends to significantly increase in volume. As a result, notes belonging to the singing voice melody might be missed and guitar notes might be transcribed instead.

Nevertheless, it can be seen from Figure 4 that the algorithm is capable of detecting ornamented and modified occurrences of a query pattern. It is also interesting to note that the obtained results contain a similar melodic sequence that was found in a recording of a different style (Figure 4 (b)), a *Bulería*. Despite this result being rated as not relevant in this task, it nevertheless demonstrates the potential of this tool for uncovering hidden structures and similarities in the context of large mining studies.

Table 2: P@5 and P@10 measures among queries.

query	P@5	P@10
A	80%	60%
A'	100%	70%
B	100%	70%
C	40%	40%
D	20%	10%

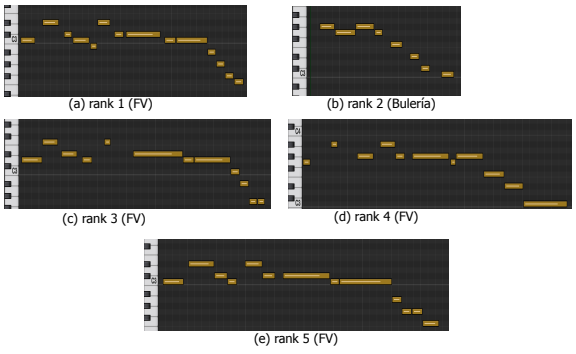

Figure 4: MIDI representations of the top 5 results for query pattern A.

5. CONCLUSIONS

We presented an algorithm for melodic pattern retrieval based on automatic transcriptions and demonstrated examples of the capabilities and limitations of the system. Future applications are expected to include the incorporation of the algorithm in a framework for unsupervised pattern detection, the retrieval of typical ornamentations from music recordings and the detection of short melodic guitar fragments (*falsetas*) in the melody of the singing voice.

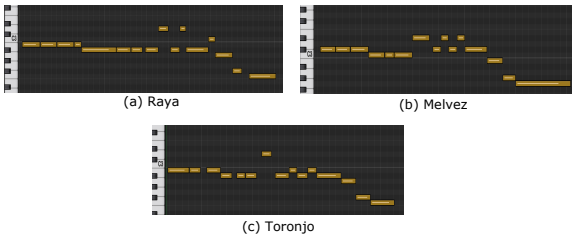

Figure 5: Manual transcriptions of pattern D for three singers: (a) A. Raya, (b) M. Vélez and (c) P. Toronjo.

6. REFERENCES

Fernández-Marín, L. (2011). La bimodalidad en las formas del fandango y en los cantes de levante: origen y evolución. *La Madrugá*, *5*(1), 37–53.

Gómez, F., Díaz-Báñez, J. M., Gómez, E., & Mora, J. (2014). Flamenco music and its computational study. In *BRIDGES: Mathematical Connections in Art, Music, and Science.*

Gómez, F., Mora, J., Gómez, E., & Díaz-Báñez, J. M. (2016). Melodic contour and mid-level global features applied to the analysis of flamenco cantes. *Journal of New Music Research*, *In Press*.

Gulati, S., Serrá, J., Ishwar, V., & Serra, X. (2014). Melodic pattern extraction in large audio collections of indian art music. In *International Conference on Signal Image Technology and Internet Based Systems - Multimedia Information Retrieval and Applications.*, (pp. 264–271).

Kroher, N., Díaz-Báñez, J. M., Mora, J., & Gómez, E. (2016). Corpus cofla: A research corpus for the computational study of flamenco music (in press). *ACM Journal on Computing and Cultural Heritage.*

Kroher, N. & Gómez, E. (2016). Automatic transcription of flamenco singing from polyphonic music recordings. *IEEE-Transactions on Audio, Speech and Language Processing*, *24*(5), 901–913.

Needleman, S. B. & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, *48*(3), 443–453.

Pikrakis, A., Gómez, F., Oramas, S., Díaz-Báñez, J. M., Mora, J., Escobar-Borrego, F., Gómez, E., & Salamon, J. (2012). Tracking melodic patterns in flamenco singing by analyzing polyphonic music recordings. In *13th International Society for Music Information Retrieval Conference (ISMIR).*

Volk, A. & van Kranenburg, P. (2012). Melodic similarity among folk songs: An annotation study on similarity-based categorization in music. *Musicae Scientiae*.