ADDING DESCRIPTORS TO MELODIES IMPROVES PATTERN MATCHING: A STUDY ON SLOVENIAN FOLK SONGS

Vanessa Nina Borsan **Mathieu Giraud**
Univ. Lille, CNRS, Centrale Lille
UMR 9189 CRIStAL, F-59000 Lille, France
{vanessa,mathieu}@algomus.fr

Richard Groult **Thierry Lecroq**
Univ Rouen Normandie, INSA Rouen Normandie,
Université Le Havre Normandie, Normandie Univ
LITIS UR 4108, F-76000 Rouen, France
{thierry.lecroq,richard.groult}@univ-rouen.fr

ABSTRACT

The objective of pattern-matching topics is to gain insights into repetitive patterns within or across various music genres and cultures. This approach aims to shed light on the recurring instances present in diverse musical traditions. The paper presents a study analyzing folk songs using symbolic music representation, including melodic sequences and musical information. By examining a corpus of 400 monophonic Slovenian tunes, we are releasing annotations of structure, contour, and implied harmony. We propose an efficient algorithm based on suffix arrays and bit-vectors to match both music content (melodic sequence) and context (descriptors). Our study reveals that certain descriptors, such as contour types and harmonic "stability" exhibit variations based on phrase position within a tune. Additionally, combining melody and descriptors in pattern-matching queries enhances precision for classification tasks. We emphasize the importance of the interplay between melodic sequences and music descriptors, highlighting that different pattern queries may have varying levels of detail requirements. As a result, our approach promotes flexibility in computational music analysis. Lastly, our objective is to foster the knowledge of Slovenian folk songs.

1. INTRODUCTION

Music pattern analysis in the field of Music Information Retrieval (MIR) is extensively studied. The challenges of this topic extend beyond algorithms, encompassing diverse music forms, representation (signal, symbolic, or textual), music content, and cultural metadata.

1.1 Content and Context in Ethnomusicology

Ethnomusicologists analyze recordings, live performances, and transcriptions (in various notations) to understand the composition of music. While transcriptions reveal the *what* of the music, cultural context is essential for comprehending the *how* and *why* behind these musical structures.

Initiated by Merriam [1], and many others [2–7], the music is to be observed *in* culture (in his later work, *as* culture), or as a multi-dimensional object, a direct consequence of the organization of social structures, and vice versa. Some studies [4, 6, 8–10] have primarily compared folk tunes based on their music content. In others, including the Slovenian Folk Song Collection [11], the categorization of the collection is organized according to the elements, such as lyrics and other textual content.

Considering *music material*, it can be explored as a general outline for music analysis [12], or through specific music descriptors, such as melodic contour (the melodic arch shape) [13]. Recent studies have expanded the use of descriptors to analyze folk songs, incorporating a broader range of attributes. De la Ossa [14] suggests basic music descriptors be included, such as scale types, range, several levels of rhythmic information, and so on. *Cantometrics*, introduced by Lomax [15, 16], proposed 37 descriptors, (almost) independent from usual Western music theory. His idea of representing datasets as a digital "Global Jukebox" was recently completed. [17–19]. Computational methodologies encourage us to process more data, including multiple layers of music *content*, and *context* as descriptors (music and/or metadata). Serra [20] exposed the presence of musical entities (performer, music, instrument, etc.) that "are linked by various types of relationships," which contribute to the understanding of music as a whole. Conklin and Neubarth also stressed the importance of non-musical information, such as region and genre [21], extended Densmore's and observed (super)area and (super)type information [22]. These non-musical phenomenons, although limited, were always correlated with different types of music content (rhythm, melody, pattern, and antipattern [23]) of folk songs. Although focusing on musical content, especially on cadences, van Kranenburg also considers lyric, perceptual information, as well as other information [24–26].

1.2 Pattern Inference and Matching

Music pattern searching or matching is most commonly approached from a music analysis perspective [12, 27–30], by addressing music structures, such as melody, harmony, and rhythm. Previous contributions focused on a single or a couple of features or the computational representations and matching of multi-feature music patterns. The Mongeau-Sankoff algorithm [31] simultaneously explores multidimensional music features, as it defines the distance between any two melodies depending on the pitch, tonal contour and rhythmic structure. The pattern similarity is ranked by the number of transformations, including consolidation and fragmentation.

Other dynamic programming methods for melodic sequence alignment were proposed [6, 32, 33], as well as other methods on the general melodic or pitch-related queries [33–36]. Some research added the rhythmic [37], or, especially with multipart music, the harmonic information [38]. In another solution, Marsden adapts the *hierarchical or tree structures* for representing and comparing melodies [39]. Lartillot, conversely, matched melodic sequence (or motives) by using *heterogeneous patterns* [30], whose occurrences can be located through multiple parametric dimensions – including contextual ones, such as implied harmony.

1.3 Motivation and Contents

These studies indicate that a melodic pattern is not isolated from other musical elements, such as a phrase, rhythmic or harmonic structure, ornamentation, and so on. While most distinguish between *music material* and *cultural metadata*, we instead split the first into *melodic sequence* and *descriptors* (see Table 1). Our objective is that *melodic phrase should never be detached from its context.* Hence, we focus on segmented melodic phrases that never lose their identifier (connecting them to all supplementary (meta)data), nor their position within a tune (first, middle, last). This enables the *tune description by phrase position, contours, labels, and rough harmonic tendencies*, and to easily access and apply any combination of other (meta)data information to the pattern-matching process.

In Section 2 we introduce the corpus as well as our annotation methodology on metadata and descriptors, including structure, contour, and implied harmony. Section 3 introduces a pattern-matching algorithm that utilizes suffix arrays (for all melodies) and bit arrays (a selection of descriptors) to return matched results based on melodic *content* and controlled descriptor context. Sections 4 and 5 discuss the implementation and the results of examples of combined melody/descriptor queries, and Section 6 provides concluding remarks and addresses open perspectives.

2. THE ANNOTATED CORPUS OF SLOVENIAN FOLK SONGS

2.1 The Corpus

We are expanding the digital world of folk songs ([17, 40–42] and others) with a limited selection of tunes from the largest collection of Slovenian folk songs – SLP, *Slovenske ljudske pesmi*), which consists of 5 critically edited physical books, issued between 1970 and 2007 [11, 43–46]. Tunes belong to *narrative song* genre (Figure 1) or hybrids between narrative and lyric genres (resemble narrative form, but much shorter). These are divided into types (by lyric resemblance) and their variants.

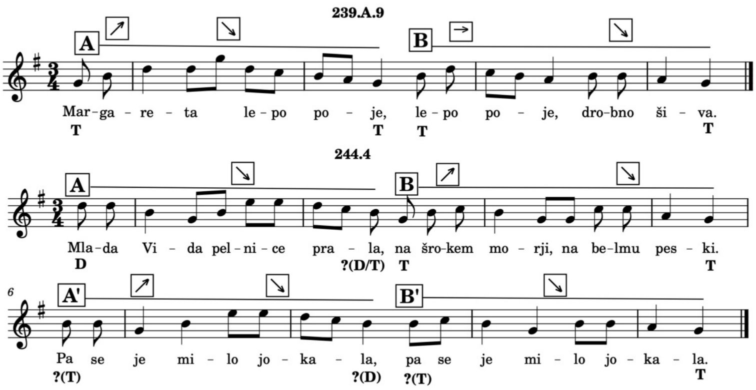
The fifth and last book [11] was edited and issued by the Slovenian Ethnomusicological Institute (Research Centre of the Slovenian Academy of Sciences and Arts, consecutively mentioned as GNI), which later digitized (OCR-ed pdf, musicXML, sib) by Matija Marolt and Matevž Pesek (FRI, University of Ljubljana). Most tunes are transcriptions of field recordings collected members of the institute, and external colleagues, such as Franc Kramar (1890–1959), Josip Dravec (1912–1996), and Stanko Vraz (1810–1851). Most tunes were collected and/or transcribed between the 1950s and 1970s. The earliest two (notated) transcriptions date back to 1819 and 1839, while the most recent transcription was completed in 2001. Together, there are 650 tune variants of 54 types and belong to family fates and conflicts topic, which represents about 34% of all Slovenian ballads [11]. Some tunes have one or two, while the most popular have 100+ variants (*Infanticide Bride*, *A Widower at His Wife's Grave*, *Step-mother and an Orphan*, and *Convicted Infanticide*). which mostly belong to Styria (166), followed by Upper Carniola (133) and Lower Carniola (99) regions.

Out of those, 418 are transcribed as monophonic, 218 [1] as homophonic, and 8 as mixed. About 70% were performed by a solo female singer. Instrumentally-accompanied examples are very rare. All tunes were transposed to G (major/minor) by the editors. Recent annotations include special notation symbols indicating deviations, such as slight disparity in pitch (higher/lower), duration (shorter/longer), and more [11].

2.2 Descriptors and Corpus Annotation

Out of the 418 monophonic tunes, 18 were excluded due to incomplete score information or incompatible encoding. Our final corpus contains 400 monophonic tune variants with detailed manual annotations, made by the first author and reviewed by all contributors. An average tune has about 9 bars and 30 notes. *Phrase boundaries* were annotated by curating the output of a simple heuristic relying

[1] Some melodies may have been harmonized by the annotator, making it unclear which singing line represents the original tune. This is common in transcriptions by Franc Kramar (1890-1959). The opposite can be true for monophonic transcriptions.



ID	POS	LBL	CT	H_S	H_E
239.A.9.1	F	A	$\nearrow\searrow$	T	T
239.A.9.2	L	B	$\rightarrow\searrow$	T	T

ID	POS	LBL	CT	H_S	H_E
244.4.1	F	A	$\searrow$	D	?(D/T)
244.4.2	M	B	$\nearrow\searrow$	T	T
244.4.3	M	A'	$\nearrow\searrow$	?(T)	?(D)
244.4.4	L	B'	$\searrow$	?(T)	T

Figure 1. Tunes from the SLP corpus with structural, contour, and implied harmony annotations. (Top) *The Death of the Bride Before Her Wedding*, 9th variant of tune type 239(A), transcribed by GNI in 1960. The first phrase (F), labeled A has a convex ($\nearrow\searrow$) contour, whereas the second phrase has a horizontal-descending ($\rightarrow\searrow$) contour as the first pitch of B phrase is about at the average compared to consecutive pitches of the phrase. Starting (upbeat) and ending implied harmonies (H_S, H_E) can be clearly labeled as a tonic, even though the strong beat on the first full measure is a D pitch. (Bottom) *The Widower at His Wife's Grave*, 44th variant of tune type 252, transcribed by Franc Kramar in 1913. Phrases A and B are approximately repeated as phrases A' and B', with changes in melody, contours, and harmonic functions. Harmonically, the first phrase starts on a clear dominant but is somewhat ambivalent at the end. The second phrase is the most stable on a tonic. The rest is slightly more ambivalent between the two degrees again, while, at the very end, the tune concludes on a tonic. (Right) A few descriptors that are associated with these tunes (see Table 1).

on pauses and punctuation marks, yielding 1502 phrases (median of 4 per tune, min. 2, max. 8).

Musical descriptors, listed in Table 1, describe either the full tune or phrases. The descriptors can carry both, *non-musical* and *musical* information.

Tune metadata [4] relies on transcribers' information, and the format aligns with the original sources, with some conversions made for data analysis convenience [11]. Tune phrases were annotated with *descriptors* across the following categories:

- *Phrase position.* This central annotation category establishes the relationship between other descriptors. Each phrase is annotated with a sequence number and position, such as first, middle, or last.
- *Structure.* Each phrase is assigned a label that describes the repetition of its melodic material within the verse. The first label is always A, followed by A, A', A+ (similar to B), A(X) (refrain-like A), or B. The same alphabetical progression is applied to subsequent phrases. The tunes have an average of 2.82 different labels (1 to 6 with symbols or 4 with letters only) and infrequent repetitions (Table 2). Each label appears only 1.34 times on average per tune.
- *Implied Harmony.* Using Western harmony to describe folk songs may be biased and controversial, but it is likely that Slovenian folk tunes and their transcribers have been exposed to the Western music system to some extent [47]. Approximate functions of tonic (T) or dominant (D) were annotated for about 60% of phrase beginnings and 50% of endings, with ambiguous cases marked as "?". To evaluate the validity of individual annotations, the inclusion of scale information (the count of distinct pitch classes) is provided.
- *Contour.* Diverging from only comparing the unreliable note-to-note melodic representation of oral music tradition, we use Huron's 9 types of melodic arches [13] (the most frequent being the convex contour $\nearrow\searrow$), where the starting and ending MIDI pitch value is compared against an average value of all intermediate MIDI pitch values.

[4] Metadata was collected for all 650 songs, including polyphonic compositions.

3. MATCHING MELODY WITH DESCRIPTORS

3.1 Pitch and Descriptor Representation

Each tune is subdivided into *phrases* as *pitch sequences* with *descriptors*, which are considered as a set of n phrases $P = \{p_1, p_2, ..., p_n\}$, and m phrase descriptors $\Delta^1, \ldots, \Delta^m$. Each descriptor Δ^t has a finite set of values $V(\Delta^t)$. A phrase p_i is associated with a *descriptor sequence* $d_i = (d_i^1, d_i^2, \ldots, d_i^m)$, where each d_i^t is in $V(\Delta^t)$.

For example, the following options:

$$\begin{cases} \{\Delta^1, \ldots, \Delta^5\} = \{\text{POS}, \text{LBL}, \text{CT}, \text{H}_S, \text{H}_E\} \\ V(\text{POS}) = \{\text{F}, \text{L}, \text{M}\} \quad V(\text{H}_E) = \{\text{T}, \text{D}, ?\} \\ V(\text{LBL}) = \{\text{A}, \text{B}\} \qquad V(\text{H}_E) = \{\text{T}, \text{D}, ?\} \\ V(\text{CT}) = \{\nearrow\searrow, \rightarrow\searrow, \nearrow, \ldots\} \end{cases}$$

can describe phrase 239.9.A.1 (Figure 1) as:

$$p_{239.9.A.1} = gbddgdcbag \quad d_{239.9.A.1} = (\text{F}, \text{A}, \nearrow\searrow, \text{T}, \text{T}).$$

		Non-musical Metadata
tune ID	$\triangleright$	Type, variant, other
type title	$\triangleright$	Title or label
region	$\triangleright, \star$	Region
annotator	$\triangleright$	Name of the initial collector/transcriber
year	$\triangleright, \star$	Year [2] of initial annotation.
singer	$\triangleright, \star$	Singer sex and ensemble size
lyric	$\triangleright, \star$	First line [3], first verse, structure
		Musical Descriptors
POS	$\star$	Phrase position (**F**irst, **M**iddle, **L**ast)
NUM	$\star$	Phrase number (1, 2, 3, 4, ...)
LBL	$\star$	Phrase label (A, B, C, ...)
SYM	$\star$	Phrase symbol (A**'**, A**"**, A **+**, ...)
CT_SPEC	$\diamond$	Huron's contours (↗, ↘, ↗↘, ↘↗, →, ↗→, ↘→, →↗, →↘)
H_S	$\star$	Starting harmony (T, D, ?, ...)
H_E	$\star$	Ending harmony (T, D, ?, ...)
TS	$\diamond$	Time signature (simple duple/triple ...)
SCALE	$\diamond$	Scale (8, 7, 6, ...)
		Musical Content
MEL		Melodic sequence (example: *gbddg*)

Table 1. Metadata, musical descriptors, and musical content. Manual annotations were done for phrase boundaries and descriptors marked with $\star$, while computed descriptors are marked with $\diamond$. Descriptors marked with $\triangleright$ were collected by the initial transcribers and/or the GNI. Other descriptors, such as general contour, tone set, and leading tone, are also present in the dataset but not discussed here.

3.2 Melody and Descriptor Pattern Matching

The goal of *melody-and-descriptor matching* is to find all phrases (associated with their descriptors) matching in both the given *pitch pattern* and selected variation of *descriptor pattern*.

A *pitch pattern* pp is also a sequences of pitches. It matches a phrase p_i when pp is a factor of p_i, matching note-to-note. This definition currently permits no kind of deviation. For example, $pp = dcb$ matches $p_{239.9.A.1} = gbddg\underline{dcb}ag$.

A *descriptor pattern* is $dp = (dp^1, \ldots, dp^m) \in (V(\Delta^1) \cup \{\star\} \times \cdots \times V(\Delta^m) \cup \{\star\})$, where $\star$ is a "don't-care" symbol. It determines which descriptors are to be checked, and matches a descriptor sequence $d_i = (d_i^1, \ldots, d_i^m)$ if, for every $t = 1 \ldots m$, either $dp^t = \star$ or $dp^t = d_i^t$. For instance, the descriptor pattern $dp = (\star, \mathrm{A}, ↗↘, \star, \star)$ checks if the phrase label matches A, and contour matches ↗↘, but ignores the phrase position and harmonic functions. Thus dp matches $d_{239.9.A.1} = (\mathrm{F}, \mathrm{A}, ↗↘, \mathrm{T}, \mathrm{T})$ but does not match $(\mathrm{F}, \mathrm{B}, ↗↘, \mathrm{T}, \mathrm{D})$.

NP	instances		
2	20% (78)	AB (60)	AA (18)
3	6% (23)	ABC (11)	ABB (6)
4	**65% (261)**	**ABCD (112)**	ABAB (45)
5	2% (8)	–	–
6	6% (24)	ABCDCD (7)	AABABA (5)
8	1% (6)	ABCBCBCB (2)	–

Table 2. Out of 400 tunes, sorted according to the number of phrases (NP), the most common structure "ABCD" is present in 28% of all tunes. Label variants are ignored (A' is considered as A). There are no tunes with 7 or more than 8 phrases. Unique structures (–) are not reported.

3.3 Algorithm

For this matching problem, we first retrieve phrases and positions from a suffix array, in linear time, then filter these matches with bit-wise operators.

Pitch sequence matching with suffix array. Pitch sequences of P are concatenated to one sequence, separated by a symbol, such as $S_P = p_1\$p_2\$\ldots p_n\$$. An index data structure such as a compressed suffix array is computed and stored to retrieve all occurrences of a pitch sequence. When a query is matched at position k in S_P, the corresponding phrase p_i and its position in p_i is retrieved using a (pre-computed) bit-vector $\overline{S_p}$, and functions $rank_1(x, k)$ and $select_1(x, k)$ (respect. the number of occurrences of $\mathtt{1}$ in the prefix of length k of a bit-vector x, and the index of the k-th $\mathtt{1}$ in x). The bit-vector $\overline{S_p} = b_1 \ldots b_{|S_p|}$ is defined as $b_i = \mathtt{1}$ if $S_p = \$$, otherwise $b_i = \mathtt{0}$. Hence, the query occurs in phrase p_i at position j (within p_i) with $i = rank_1(\overline{S_p}, k)$, and $j = k - select_1(\overline{S_p}, i)$. Retrieving the list of phrases and positions of a query, pitch sequence q of length m is done in time $O(m + occ)$, where occ is the number of occurrences of q in S_p provided that $rank$ and $select$ operations are performed in constant time.

Descriptor pattern matching with bitwise operators. Each descriptor Δ^t can be represented by b^t bits, with $b^t = \lceil \log_2 |V(\Delta^t)| \rceil$, and each value $v \in V(\Delta^t)$ is associated to a bit-vector $\overline{v}$. Each descriptor sequence $d_i = (d_i^1, d_i^2, \ldots, d_i^m)$ is then stored as a bit-vector $\overline{d_i} = \overline{d_i^1} \ldots \overline{d_i^m}$. A descriptor pattern $dp = (dp^1, dp^2, \ldots, dp^m)$ is associated to two bit-vector masks $\mu(dp) = \mu^1 \ldots \mu^m$ and $\pi(dp) = \pi^1 \ldots \pi^m$, where

$$\begin{cases} \mu^t = \pi^t = \mathtt{0} \ldots \mathtt{0} \ (b^t \text{ bits}) & \text{if } dp^t = \star \\ \mu^t = \mathtt{1} \ldots \mathtt{1} \text{ and } \pi^t = \overline{dp^t} & \text{otherwise.} \end{cases}$$

Then, a descriptor pattern dp matches one descriptor d if and only if $(\overline{d} \ \mathsf{xor}\ \pi(dp))$ and $\mu(dp) = 0$. For example, if $dp = (\star, \mathrm{A}, ↗↘, \star, \star)$, then

$$\begin{aligned} \overline{d_{239.9.1.A}} &= \mathtt{00 \cdot 0 \cdot 0010 \cdot 01 \cdot 01} \\ \mu(dp) &= \mathtt{00 \cdot 1 \cdot 1111 \cdot 00 \cdot 00} \\ \pi(dp) &= \mathtt{00 \cdot 0 \cdot 0010 \cdot 00 \cdot 00} \\ \hline \overline{d_{239.9.1.A}} \ \mathsf{xor}\ \pi(dp) &= \mathtt{00 \cdot 0 \cdot 0000 \cdot 01 \cdot 01} \\ \overline{d_{239.9.1.A}} \ \mathsf{xor}\ \pi(dp) \ \mathsf{and}\ \mu(dp) &= \mathtt{00 \cdot 0 \cdot 0000 \cdot 00 \cdot 00} \end{aligned}$$

Checking whether a descriptor dp matches a descriptor d is done in $O(1)$ time, provided that the bit-vectors fit in one machine word.

	$\nearrow\searrow$	$\searrow$	$\nearrow$	$\searrow\nearrow$	$\nearrow\rightarrow$	$\searrow\rightarrow$
First	24%	16%	**30%**	15%	6%	1.2%
Middle	**36%**	21%	15%	8%	10%	1.3%
Last	**44%**	32%	2%	5%	4%	11%
Total	**35%**	23%	16%	9%	7%	1.5%

Table 3. We show the most frequent Huron's contour types according to phrase position. Regarding general contour (data not shown), first phrases are mostly ascending (53%), while middle and last phrases are predominantly descending (63% and 89% respectively).

		T	D	$?_{\mathrm{T}}$	$?_{\mathrm{D}}$	?
First	H_S	25%	**54%**	9%	<1%	12%
	H_E	22%	27%	15%	7%	**29%**
Middle	H_S	16%	**36%**	14%	6%	28%
	H_E	21%	16%	18%	6%	**40%**
End	H_S	19%	32%	10%	5%	**34%**
	H_E	**60%**	<1%	20%	None	19%
Total	H_S	19%	**40%**	11%	5%	25%
	H_E	**32%**	15%	18%	5%	32%

Table 4. Starting (H_S) and ending (H_E) harmonic functions in relation to phrase positions demonstrate a consistent pattern. Phrases typically initiate on a dominant (D) and conclude on a tonic (T). However, there is ambiguity with the functions ?, $?_{\mathrm{T}}$, and $?_{\mathrm{D}}$, as they can be interpreted as either T or D, which arises from the influence of previous and following pitch values or bars, making the exact annotation spot unclear.

4. IMPLEMENTATION AND AVAILIBILITY

The descriptors and the algorithm were implemented in Python using `music21` [48] for music data manipulation, `bitarray` library for descriptors matching and a C++ library `sdsl-lite` [49] (used in Python with `pysdsl` library) for melodic matching. From the latter, we used `rank` and `select` methods, and `BitVector` and `SuffixArrayBitcompressed` classes. On a standard laptop from 2022, building suffix arrays and bit vectors on 1502 phrases takes less than 0.5 s. A single suffix array takes 49.4 KB, and bit vectors 1.7 KB. The longest melody/descriptor queries take about 100 ms. The annotations and the code are available on a git repository through open licences (Open Database License, Database Contents License, GPLv3+) at `algomus.fr/data` and `algomus.fr/code`. We are collaborating with partners in Slovenia to prepare the release of 400 melodies and additional ethnomusicological research.

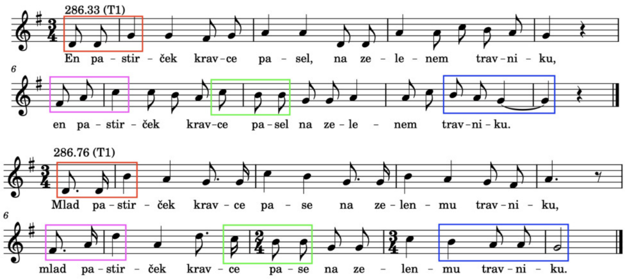


Figure 2. Two variants (out of 34) of the subtype 286.T1 with similar melodies with short melodic patterns in coloured squares.

5. RESULTS AND DISCUSSION

Almost all tunes (84%) in our collection revolve around a minimum of six distinct pitch classes, indicating the likely utilization of Western major/minor scales and modes. Approximately 67% of the tunes fall within the range of a major sixth (M6) to an octave (P8). Additional statistics in the annotated corpus examine phrase positions (Section 5.1). Specifically, the analysis focuses on the prevalent tune subtype 286.T1, annotated for melodic similarity. Findings from combined melody and descriptor queries are presented in Section 5.2.

5.1 Descriptors and Phrase Positions

The 400 tunes consist of 1502 phrases, split into first (400), middle (702), and last (400) positions. The labels, contours, and implied harmonies are strongly influenced by the phrase positions, as evident from Tables 2, 3, and 4. In general, phrases are convex or descending (see Table 3), while the *first phrases* are mostly ascending or convex, and less harmonically stable at their ends than beginnings. Conversely, the *last phrase* is almost never ascending, and more harmonically stable at its ends than beginnings. The *middle phrase* group is more divided and is more unstable.

We assume that the contrasting beginnings and endings of each verse offer pitch orientation for the singers, given the repetitive structure of narrative songs. The contour relationship between the first, middle, and last phrases supports the notion that "what goes up is likely to come down," as proposed by Huron [13].

5.2 Case Study: Subtype 286.T1, *Infanticide Bride*

Our dataset includes 103 monophonic variants of the widely known "Infanticide Bride" theme in European folk song tradition. Subtype 286.T1 consists of 34 tunes selected for their melodic similarity (Figure 2), which often exhibit similar patterns, such as the *fad* as a start middle phrase pattern or the *bag* as a last phrase ending pattern.

We have developed combined melody/descriptor queries to represent certain phrases of subtype 286.T1. These queries are evaluated as a binary classification problem: Can we accurately identify the 34 initial phrases of 286.T1 and distinguish them exclusively from others?

Pattern design and matching. Table 5 demonstrates that simple melody queries with 1 to 3 notes achieve reasonable recall rates (50%-80%) but limited precision. Refining the queries with descriptors improves precision and relevance, leading to enhanced F_1 measures. The ddb melody query alone produces 93 matches, but 75 of them are "false positives" unrelated to the first phrase of 286.T1 tunes. Incorporating a phrase position descriptor (F, first) improves the query, while adding relevant contour ($\nearrow$) and starting harmonic information further enhances specificity. This comprehensive query results in only 2 false positives, achieving a precision of up to 0.88, with minimal sensitivity loss. The ag pattern in the last phrase, characterized by a convex contour and a harmonic ending, is a noteworthy example. Given the enhanced harmonic stability typically found in verse endings, the inclusion of the $\mathrm{H}_S\mathrm{T}$ as a stable harmony descriptor proves to be effective in this context. Including too many or irrelevant descriptors leads to poor results. For instance, the cbb pattern is primarily found at the end of the middle phrase. However, requiring a stable harmonic framework ($\mathrm{H}_E\mathrm{T}$) for middle endings reduces precision, as it is less common in those positions (Table 4). Another interesting instance is the fad, occurrences of which are almost evenly split into two contours. If we matched (fad, M, $\nearrow\searrow$ or $\nearrow\rightarrow$), we would get 23 true positives and a precision of 0.82 with a recall of also 0.82. The algorithm should be extended to accommodate the matching of a subset of multiple descriptors within the same category, rather than solely relying on one descriptor.

Patterns as building blocks. Melody/descriptor patterns have versatile applications beyond classification. In our case, the most effective queries incorporate position descriptors, indicating that we should view phrase *building blocks* as patterns. It is noteworthy, that studying the "false positives" (matches outside of 286.T1) is expected to yield intriguing results, shedding light on the transmission of music material among tunes and vice versa. For instance, the ag pattern in the last phrase, exhibiting a $\nearrow\searrow$ contour and ending with $\mathrm{H}_E\mathrm{T}$, is not only specific to 286.T1 but also appears in 14 tunes of type 252 (*A Widower at His Wife's Grave*). The shared section of the melodic line in the two tunes has identical descriptors, although its positions may vary (scores not shown). Comparing outcomes across multiple corpora would provide insight into the unique musical characteristics of (Slovenian) folk songs.

5.3 Discussion

Our study explored tune structures and melodic patterns, finding that combining melodic content with descriptors provides valuable insights into the characteristics of Slovenian folk songs. However, not all descriptors fit universally to describe all content, and vice versa. In contrast to the usual transferability of folk song melodies, our case study indicates that the melody of 286.T1 was not easily transferable, possibly due to its popularity in multiple regions. Further inter- and intra-corpus research is needed to investigate this distinctive characteristic. We also show, that individual melodic extracts alone lack the specificity required for a comprehensive description of a full phrase or tune. Strong correlations were found for descriptors like contour. Expanding the dataset is needed to comprehensively explore the relationship between lyrics and melodies, including the observed descending shape in the last phrase, potentially reflecting or corresponding with speech characteristics [50, 51].

Query (melody + descriptors)			**TP**	**FP**	**FN**	**Prec.**	**Rec.**	**$\mathbf{F}_1$**
d	None	1 (34)	28	1191	6	0.02	**0.82**	0.04
d	F		28	327	6	0.08	0.82	0.14
d	F, $\mathrm{H}_S\mathrm{D}$		27	189	7	0.12	0.79	0.22
d	F, $\nearrow$, $\mathrm{H}_S\mathrm{D}$		21	48	13	0.30	0.62	**0.41**
ddb	None	1 (34)	18	**75**	16	0.19	**0.53**	0.28
ddb	F		18	32	16	0.36	0.53	0.43
ddb	F, $\mathrm{H}_S\mathrm{D}$		17	21	17	0.45	0.50	0.47
ddb	F, $\nearrow$, $\mathrm{H}_S\mathrm{D}$		14	**2**	20	**0.88**	0.41	**0.56**
fad	None	3 (33)	24	24	9	0.50	**0.73**	0.59
fad	M		24	14	9	0.63	0.73	**0.68**
fad	M, $\nearrow\searrow$		**11**	2	22	**0.85**	0.33	0.48
fad	M, $\nearrow\rightarrow$		**12**	3	21	**0.80**	0.36	0.50
cbb	None	3 (33)	25	71	8	0.26	**0.76**	0.39
cbb	M		25	39	8	0.39	0.76	**0.52**
cbb	M, $\nearrow\rightarrow$		11	3	22	0.79	0.33	0.47
cbb	M, $\mathrm{H}_E\mathrm{T}$		**1**	3	32	0.25	**0.03**	0.05
ag	None	4 (34)	27	481	7	0.05	**0.79**	0.10
ag	L		27	165	7	0.14	0.79	0.24
ag	L, $\nearrow\searrow$		23	54	11	0.30	0.68	0.41
ag	L, $\nearrow\searrow$, $\mathrm{H}_E\mathrm{T}$		23	51	11	0.31	0.68	**0.43**

Table 5. Evaluation of melody/descriptor queries seen as classification queries intended to match phrases 1, 3, and 4 of the melodic tune subtype 286.1 (34 first and last phrases, 33 third phrases) against all 1502 phrases of the dataset. We computed True Positives (TP), False Positives (FP), False Negatives (FN), and from those, precision, recall, and F_1-score. Bold values are discussed in the text.

6. CONCLUSION AND PERSPECTIVES

By integrating descriptor information into melodies, we gain a deeper understanding of the observed music. Our findings indicate a strong dependency of many descriptors on phrase positions, and that combining melody and descriptors enhances precision compared to using melody alone. Our algorithm efficiently matches melodies and descriptors, which can be extended beyond our proposed selection. Lastly, we released annotations of Slovenian folk songs, a yet underrepresented corpus in the MIR community.

Our current plans primarily involve releasing the corpus of these tunes, accompanied by comprehensive ethnomusicological commentary. In addition, future work should prioritize improving the algorithm's usability for non-computational users, expanding the existing annotations of descriptors, and implementing the capability to perform combined query searches with approximate matching for melodies and descriptors. Our study (and corpus) may be used as supporting data for new algorithms of phrase segmentation, tune structure analysis, and harmony tasks including semi-automatic annotation.

Acknowledgements. We extend our gratitude to Matija Marolt, Matevž Pesek, and current as well as past members of the Institute of Ethnomusicology ZRC SAZU for their pivotal role in digitizing and curating the corpus from field recordings, notations and notes. We also appreciate the valuable input and comments provided by Dinh-Viet-Toan Le and the Algomus team, Patrick E. Savage and the members of Comp Music Lab, and the anonymous reviewers, which greatly contributed to the development of this paper.

7. REFERENCES

[1] A. P. Merriam, *The Anthropology of Music*. Northwestern University Press, 1964.

[2] K. Blaukopf, *Musical Life in a Changing Society: Aspects of Music Sociology*. Hal Leonard Corporation, 1992.

[3] I. Mills, "The heart of the folk song," *Canadian Folk Music Journal*, vol. 2, pp. 29–34, 1974.

[4] S. P. Bayard, "Prolegomena to a study of the principal melodic families of British-American folk song," *The Journal of American Folklore*, vol. 63, no. 247, pp. 1–44, 1950.

[5] Z. Kumer, *Vloga, zgradba, slog slovenske ljudske pesmi*. Založba ZRC, 1996.

[6] P. van Kranenburg, A. Volk, and F. Wiering, "On operationalizing the musicological concept of tune family for computational modeling," *Proceedings of Supporting Digital Humanities: Answering the unaskable*, 2011.

[7] T. Rice, *Modeling Ethnomusicology*. Oxford University Press, 2017.

[8] J. R. Cowdery, "A fresh look at the concept of tune family," *Ethnomusicology*, vol. 28, no. 3, pp. 495–504, 1984.

[9] C. Pendlebury, "Tune families and tune histories: Melodic resemblances in British and Irish folk tunes," *Folk Music Journal*, vol. 11, no. 5, pp. 67–158, 2020.

[10] A. Volk and P. van Kranenburg, "Melodic similarity among folk songs: An annotation study on similarity-based categorization in music," *Musicae Scientiae*, vol. 16, no. 3, pp. 317–339, 2012.

[11] M. G. Kaučič, M. Klobčar, Z. Kumer, U. Šivic, and M. Terseglav, *Slovenske ljudske pesmi V.: Pripovedni pesmi*. Založba ZRC, 2007, vol. 5.

[12] S. Ahlbäck, "Melody beyond notes: A study of melody cognition," Ph.D. dissertation, Göteborgs Universitet, 2004.

[13] D. Huron *et al.*, "The melodic arch in Western folksongs," *Computing in Musicology*, vol. 10, pp. 3–23, 1996.

[14] S. d. l. Ossa, *A Basic Guide to Folksong Analysis*. Budapest: Liszt Academy of Music, 2019.

[15] A. Lomax, *Folk Song Style and Culture*. Routledge, 2017.

[16] A. Lomax and N. Berkowitz, "The evolutionary taxonomy of culture: A few behavioral factors account for the regional variation and evolutionary development of culture," *Science*, vol. 177, no. 4045, pp. 228–239, 1972.

[17] A. Wood, K. R. Kirby, C. Ember, S. Silbert, S. Passmore, H. Daikoku, J. Mcbride, F. Paulay, M. Flory, J. Szinger *et al.*, "The Global Jukebox: A public database of performing arts and culture," *PLOS One*, vol. 17, no. 11, 2021. [Online]. Available: https://doi.org/10.1371/journal.pone.0275469

[18] A. L. Wood, *Songs of Earth: Aesthetic and Social Codes in Music*. University Press of Mississippi, 2021.

[19] P. E. Savage, "Alan Lomax's cantometrics project: A comprehensive review," *Music & Science*, vol. 1, 2018.

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

[21] K. Neubarth, I. Goienetxea, C. G. Johnson, and D. Conklin, "Association mining of folk music genres and toponyms," *International Society for Music Information Retrieval Conference (ISMIR 2012)*, pp. 7–12, 2012.

[22] K. Neubarth and D. Conklin, "Contrast pattern mining in folk music analysis," in *Computational Music Analysis*. Springer, 2016, pp. 393–424.

[23] D. Conklin, "Antipattern discovery in folk tunes," *Journal of New Music Research*, vol. 42, no. 2, pp. 161–169, 2013.

[24] P. van Kranenburg, "Computational approach to content-based retrieval of folk song melodies," Ph.D. dissertation, Utrecht University, 2010.

[25] P. van Kranenburg and F. Karsdorp, "Cadence detection in western traditional stanzaic songs using melodic and textual features," in *International Society for Music Information Retrieval Conference (ISMIR 2014)*, 2014, pp. 391–396.

[26] C. McKay, J. Cumming, and I. Fujinaga, "jSymbolic 2.2: Extracting features from symbolic music for use in musicological and MIR research," in *International Society for Music Information Retrieval Conference (ISMIR 2018)*, 2018, pp. 348–354.

[27] D. Conklin and M. Bergeron, "Discovery of contrapuntal patterns," in *International Society for Music Information Retrieval Conference (ISMIR 2010)*, vol. 2010, 2010, pp. 201–206.

[28] I. Y. Ren, H. V. Koops, A. Volk, and W. Swierstra, "In search of the consensus among musical pattern discovery algorithms," in *International Society for Music Information Retrieval Conference (ISMIR 2017)*, 2017, pp. 671–678.

[29] E. Cambouropoulos, "Musical rhythm: A formal model for determining local boundaries, accents and metre in a melodic surface," in *Joint International Conference on Cognitive and Systematic Musicology*, 1996, pp. 277–293.

[30] O. Lartillot, "Automated motivic analysis: An exhaustive approach based on closed and cyclic pattern mining in multidimensional parametric spaces," in *Computational Music Analysis*, 2016, pp. 273–302.

[31] M. Mongeau and D. Sankoff, "Comparison of musical sequences," *Computers and the Humanities*, vol. 24, no. 3, pp. 161–175, 1990.

[32] P. E. Savage, "Measuring the cultural evolution of music: Cross-cultural and cross-genre case studies," in *PsyArXiv*, 2020 (preprint).

[33] P. E. Savage, S. Passmore, G. Chiba, T. E. Currie, H. Suzuki, and Q. D. Atkinson, "Sequence alignment of folk song melodies reveals cross-cultural regularities of musical evolution," *Current Biology*, vol. 32, no. 6, pp. 1395–1402, 2022.

[34] T. Eerola and M. Bregman, "Melodic and contextual similarity of folk song phrases," *Musicae Scientiae*, vol. 11, no. 1, pp. 211–233, 2007.

[35] C. Anagnostopoulou, M. Giraud, and N. Poulakis, "Melodic contour representations in the analysis of children's songs," in *International Workshop on Folk Music Analysis (FMA 2013)*, 2013, pp. 40–43.

[36] J. McBride, A. T. Tierney, P. Pfordresher, J. Six, S. Fuji, and P. E. Savage, "Are pitches discrete? an information-theoretic framework and a corpus study," in *International Conference on Music Perception and Cognition and Triennial Conference of the European Society for the Cognitive Sciences of Music (ICMPC-ESCOM 2021)*, 2021.

[37] C. Finkensiep, K. Déguernel, M. Neuwirth, and M. Rohrmeier, "Voice-leading schema recognition using rhythm and pitch features," in *International Society for Music Information Retrieval Conference (ISMIR 2020)*, 2020, pp. 520–526.

[38] E. Cambouropoulos, "The harmonic musical surface and two novel chord representation schemes," in *Computational Music Analysis*. Springer, 2016, pp. 31–56.

[39] A. Marsden, "Representing melodic patterns as networks of elaborations," *Computers and the Humanities*, vol. 35, no. 1, pp. 37–54, 2001.

[40] H. Schaffrath, "The Essen Associative Code: A code for folksong analysis," in *Beyond MIDI: The handbook of musical codes*. MIT Press, 1997, p. 343–361.

[41] T. Eerola and P. Toiviainen, "Suomen kansan esävelmät – finnish folk song database," 1999. [Online]. Available: http://esavelmat.jyu.fi/

[42] P. van Kranenburg, M. De Bruin, and A. Volk, "Documenting a song culture: the Dutch song database as a resource for musicological research," *International Journal on Digital Libraries*, vol. 20, no. 1, pp. 13–23, 2019.

[43] Z. Kumer, M. Matičetov, B. Merhar, and V. Vodušek, *Slovenske ljudske pesmi I: Pripovedni pesmi*. Ljubljana: CGP Delo, 1970, vol. 1.

[44] Z. Kumer, M. Matičetov, and V. Vodušek, *Slovenske ljudske pesmi II: Pripovedni pesmi*. Ljubljana: Slovenska matica, 1981, vol. 2.

[45] M. Terseglav, I. Cvetko, M. Golež, and J. Strajnar, *Slovenske ljudske pesmi III: Pripovedni pesmi*. Ljubljana: Slovenska matica, 1992, vol. 3.

[46] ——, *Slovenske ljudske pesmi IV: Pripovedni pesmi*. Ljubljana: Slovenska matica, 1992, vol. 4.

[47] V. Vodušek, *Etnomuzikološki članki in razprave*. Založba ZRC, 2003.

[48] M. S. Cuthbert and C. Ariza, "music21: A toolkit for computer-aided musicology and symbolic music data," *International Society for Music Information Retrieval Conference (ISMIR 2010)*, pp. 637–642, 2010.

[49] S. Gog, T. Beller, A. Moffat, and M. Petri, "From theory to practice: Plug and play with succinct data structures," in *International Symposium of Experimental Algorithms (SEA 2014)*, 2014, pp. 326–337.

[50] Y. Ozaki, A. Tierney, P. Pfordresher, J. Mcbride, E. Benetos, P. Proutskova, G. Chiba, F. Liu, N. Jacoby, S. Purdy *et al.*, "Similarities and differences in a global sample of song and speech recordings [stage 2 registered report]," *PsyArXiv*, 2022 (preprint).

[51] P. Albouy, S. A. Mehr, R. S. Hoyer, J. Ginzburg, and R. J. Zatorre, "Spectro-temporal acoustical markers differentiate speech from song across cultures," *bioRxiv*, 2023 (preprint).