UNIVERSITÉ DE LILLE
ÉCOLE DOCTORALE MADIS-631

DOCTORAL THESIS

Melodic and Descriptor Patterns: A Computational Approach for Digitisation, Annotation, and Analysis of Slovenian Folk Song Ballads

Author:
Vanessa Nina BORSAN

Thesis to be defended in December before the following examination committee:

Kevin PAGE	Senior Researcher, University of Oxford	*reviewer (rapporteur)*
Emilios CAMBOUROPOULOS	Professor, Aristotle University of Thessaloniki	*reviewer (rapporteur)*
Marc CHEMILLIER	Directeur d'études, EHESS	*examiner (examinateur)*
Urša ŠIVIC	Research Associate, Ethnomusicology Institute ZRC SAZU	*examiner (examinatrice)*
Maude PUPIN	Professor Université de Lille	*examiner (examinatrice)*
Mathieu GIRAUD	Directeur de recherche CNRS, Université de Lille	*co-supervisor (co-directeur)*
Thierry LECROQ	Professor, Université de Rouen Normandie	*co-supervisor (co-directeur)*
Richard GROULT	Maître de conférences, Université de Rouen Normandie	*co-supervisor (co-encadrant)*

UNIVERSITÉ DE LILLE

Abstract

École doctorale MADIS-631

PhD in Computer Science (Informatique et applications)

Melodic and Descriptor Patterns: A Computational Approach for Digitisation, Annotation, and Analysis of Slovenian Folk Song Ballads

by Vanessa Nina BORSAN

This thesis focuses on combining digital archiving and music analysis derived from ethnomusicological practices and methods that are rooted in computational sciences, especially those found in the field of Music Information Retrieval (MIR). The analysis and other tasks focus on the collection of Slovenian folk song ballads, which trigger a number of questions in terms of methodology and data interpretation. In the broadest sense, the thesis provides three contributions.

First, it digitises and publishes a dataset of 402 monophonic transcriptions of Slovenian folk song ballads, which had previously been accessible only in physical archives and editions. These transcriptions, enriched with metadata and annotations, are made available through the open source platform Dezrann. The dataset, along with its accompanying materials, includes detailed information on the history of the corpus and its melodic transcriptions, as well as a variety of music annotations.

Second, it introduces pattern matching algorithms specifically designed for this folk song dataset. These algorithms, implemented using Python libraries, address four distinct tasks: melodic sequence matching, two descriptor set-based matching tasks, and a combination of both. By using bitwise operators and compressed suffix arrays, the method proves as a time- and space-efficient solution for pattern discovery and classification. The flexibility of these methods encourages their applicability beyond introduced corpus to other types of (music) collections.

Finally, it critically engages with the disciplinary intersections of MIR and ethnomusicology, examining the challenges and opportunities for collaboration between these fields. While MIR provides powerful tools for large-scale music analysis, this research highlights the difficulties in adapting computational techniques to the nuanced, context-rich materials often encountered in ethnomusicology. Through case studies and evaluations, the thesis demonstrates the potential for integrating algorithmic methods with traditional musicological approaches, while also addressing the institutional and disciplinary barriers that have hindered such cross-disciplinary work.

Contents

Chapter 1

Introduction

Contemporary computational (ethno)music(ology) research has introduced novel ways of exploring musical materials, from comparative and big corpora studies to encouraging new queries and research questions that were unimaginable in the pre-computational era.

This thesis bridges ethnomusicology, computer science, and text algorithmics in an effort to establish an efficient method for music pattern matching and provide a dataset that can accommodate such tasks.

Recognising the scarcity of well-documented, freely accessible datasets, the thesis first took one step back to digitise, model, and publish materials that, until now, existed only in printed and physical form within archives. The thesis provides a dataset of 402 monophonic transcriptions of Slovenian folk song ballads with metadata and annotations.

In developing a methodology for organising, annotating, and analysing these materials, as well as designing time- and space-efficient algorithms for handling specific melodic patterns and matching the query descriptor set, several challenges emerged. These challenges expanded the scope of the thesis, encompassing fields such as digital archiving, music libraries, and even media studies. The latter served as a foundation for exploring how such computational contributions benefit both Music Information Retrieval (MIR) and the ethnomusicological community.

The topics addressed within the scope of this interdisciplinary challenge are organised around five overarching objectives, which intend to:

- Objective 1: Digitise, organise, curate, and thoroughly document the materials of selected Slovenian folk song ballads.
- Objective 2: Annotate the materials and create a publicly accessible digital dataset on the Dezrann platform, completed with these annotations.
- Objective 3: Develop algorithms for melodic and descriptor pattern matching tailored to these materials, while remaining flexible enough to accommodate other research queries.
- Objective 4: Provide a multi-step evaluation of these contributions and conduct two case studies to test the algorithms' performance as well as (contextual) flexibility.

- Objective 5: Examine the role of this contribution, and similar computational methods, within the frameworks of ethnomusicology and related fields, while critically assessing the pitfalls, benefits, miscommunications, and impact. Additionally, to evaluate potential next steps for such methodologies.

Although each objective warrants individual attention, it is important to note that they are closely interrelated, with most objectives overlapping throughout each chapter of the thesis. To offer the reader an overview of the thesis structure, I provide a brief summary that follows the chapters in a somewhat linear fashion.

1.1 Thesis Chapter Organisation

1.1.1 Related work

At the outset, in Chapter 2, I introduce the relevant work that supports the majority of the stated objectives. First, I focus on ethnomusicological approaches to collecting, organising, and structuring datasets for analysis, highlighting different principles, from adapting materials to fit somewhat Western classical systems, to modifying these systems or offering alternative solutions to accommodate various types of music, such as *cantometrics*.

The discussion then shifts from broader analytical models to more specific ideas on *how certain musical elements should be symbolically represented*, highlighting music notation (it being the centre of this thesis), such as David Huron's concept of viewing melodic lines as contours or arches. Both sections conclude by exploring the computational application of these principles in contemporary research, addressing various hierarchical levels of music representation—from interrelated elements within a score to larger systems that connect scores and corpora based on shared features.

Speaking of computational methods, the next step was to introduce the existing *methods for digitising and analysing* these materials. This part includes a summary of available datasets, noting the various approaches to addressing these challenges. For instance, while some collections focus on gathering diverse materials in one place, others adapt *digital archiving* techniques to suit specific musical traditions. I also address the varying usability of such datasets, depending on the quality, quantity, accessibility, and format of the digital materials.

The final section of this introductory chapter focuses on a specific type of digital music analysis of such materials: *pattern matching* using principles from text algorithmics. It begins by providing an overview of key contributions in the field, covering both online and offline, as well as exact and approximate, pattern matching methods. These approaches, along with their adaptations for music tasks, are then introduced through examples of single and multi-parameter music pattern matching, highlighting the most significant contributions from MIR.

1.1.2 Slovenian Folk Song Ballads

The thesis then addresses the digitised dataset of 402 Slovenian folk song ballads, including explaining the terminology behind the dataset's name (Chapter 3). Since the songs were collected, transcribed, and curated over different historical periods, from the early 1900s to the late 1990s, I also examined the changes and advancements in practice, tools, and methodology, and how these developments may have influenced the materials.

Before diving into pattern matching analyses on these songs, I introduce the structure, annotation system, and general statistics, along with initial explanations, to provide a clearer understanding of the potential results and how they should be interpreted. Lastly, for easier orientation, I also explain how all of these materials are organised on the digital platform Dezrann, as well as provide notes on the availability and licenses.

1.1.3 Pattern Matching Tasks

Following the related work and dataset introduction, I elaborate on the key pattern matching problems and outline methodologies for the proposed algorithms, categorised as *melodic sequence*, *descriptor*, and *mixed* pattern matching, split into 4 different problems (Chapter 4). *Problem 1* involves identifying occurrences of a short melodic sequence in the dataset using a compressed suffix array and bitarray. *Problem 2* focuses on single descriptor set matching using bit vectors and bitwise operators. *Problem 3* allows multiple descriptor elements per query, using inverted dictionaries to retrieve corresponding IDs. Finally, *Problem 4* combines melodic sequences and descriptors to identify instances where both match the query. The chapter provides further problem definitions for each of these, as well as ensures the data encoding and pattern matching task explanations with examples.

1.1.4 Evaluation

Next, I provide a brief overview of what should be evaluated when considering computational tasks on music, highlighting key points from various MIR discussions (Chapter 5). The evaluation is then divided into three parts. First, performance evaluation assesses time and memory usage for data conversion and the four pattern matching algorithms, alongside a brief commentary on their implementation in Python. Second, a case study on Slovenian folk song ballads involves a classification task using single descriptor set (Problem 2) and melodic sequence (Problem 1) pattern matching. Third, another case study introduces a different pre-existing corpus, focusing on a new classification task with multiple descriptor set matching (Problem 3), meaning to allow multiple descriptor elements per descriptor to be considered in a query, and adapted melodic sequence matching. Finally, the chapter concludes with evaluation on the broader use of these methodologies in real-life music research, addressing both advantages and limitations.

1.1.5 MIR and (Ethno)musicology

Before drawing conclusions, the thesis expands on the points raised in the evaluation, focusing on the final objective: examining the role of this contribution, and similar computational methods, within the frameworks of ethnomusicology and related fields (Chapter 6). It critically assesses the pitfalls, benefits, miscommunications, and impact of these approaches. The discussion seeks to identify the underlying causes of limitations in computational music research, while also considering broader challenges in knowledge transfer between various actors and subdisciplines, even those not directly connected, yet very relevant, to the pattern matching methodology. To address these issues, the discussion focuses on three main areas: (1) the relationship to new media and technologies (in research), (2) the foundations of specific disciplinary theoretical concerns, and (3) the structure of disciplines and supporting institutions.

1.1.6 Conclusion

The thesis concludes with a brief summary of the key topics covered in the preceding chapters, along with an assessment of each of the five objectives. It finishes by presenting potential future directions that yet need to be or should be explored in similar lines of research.

Chapter 2

Related Work

Statement: This chapter includes an extended version of literature review from my published paper at ISMIR 2023 [32] and a submitted paper to Ethnomusicology Forum [33] (in review).

Rooted in the confluence of ethnomusicology, digital humanities, and computational analysis, this thesis delves into the intersections of digital archiving, the innovative use of text algorithmics, particularly in music pattern matching tasks, and Slovenian folk song. Beginning with an overview of ethnomusicological thought to contextualise the materials and methodology of the thesis, the section outlines the cultural nuances shaping musical traditions, segueing into the structures and organisations of digital archives where preservation meets computational processing and a variety of material formats. The thesis provides a detailed review of existing projects with commentary on their contents, usability and accessibility. Laying the groundwork for the representation and annotation of digitised musical data, the review of related work also focuses on the way data is represented in such datasets or archives, along with their pros and cons. Lastly, by first acknowledging the contributions of pattern matching in text algorithmics, the chapter concludes with an exploration of music pattern discovery and similar tasks, and their potential as a computational methodology to elucidate musical structures.

2.1 Ethnomusicological Framework

In the introductory segment of the state-of-the-art review, I will explore pivotal insights from ethnomusicology and related fields, acknowledging their essential influence on the development of computational systems in music research. Due to the extensive contributions from these disciplines, my examination will focus specifically on the elements that are most pertinent to the thesis topic, its resources, and methodological approaches.

In the last few decades, a large attention has been paid to identify the effective computational research methods for non-Western art music, many of which trace back to the pre-computational era. For example, the foundational works of Samuel Bayard [14] and James R. Cowdery [59], Kurt Blaukopf [25, 324], Isabelle Mills [206], and also the contributions of Alan Lomax and colleagues [181, 260, 322], Alan P. Merriam [205], Mantle Hood [114], and Timothy Rice [249], have laid the groundwork for contemporary ethnomusicological studies. These pioneering efforts continue to inspire modern computational ethnomusicology, particularly in tune matching, pattern matching tasks, and the development of music ontology, taxonomy, and topology.

To provide the needed context for the key tasks of this thesis, it is important to trace the historical development of ethnomusicological methodologies. Understanding their evolution sheds light on how these early contributions shaped current state of research and influenced the integration of computational techniques in music research. This historical perspective highlights the progression from traditional ethnomusicological approaches to the sophisticated computational methods employed today.

2.1.1 On Music Data Representation

Starting with Bayard, his theoretical framework addressed the nuances in tune development, by explaining the relationship between recording and origin time, and the selection of descriptors for understanding music patterns, emphasising the subjectivity of these observations [14]. He discussed tonal range, rhythm, melodic progression, the sequence of stressed notes, and notably introduced the concept of *tune family*, a notion closely aligned with computational music research. The concept *tune family* [14], was used to categorise groups of tunes that share a common ancestor in oral transmission. More precisely, Bayard defined a *tune family* as "a group of melodies showing basic interrelation by means of constant melodic correspondence, and presumably owing their mutual likeness to descent from a single air that has assumed multiple forms through processes of variation, imitation, and assimilation"[14].

"Although Bayard laid much important groundwork, he did not provide sufficient models for these potential applications [59]". Instead of combining descriptors for the tune as a whole, Cowdery suggested three principles for working with folk song material - the outlining, conjoining, and recombining principle. The first one "allows us to compare wholes to wholes, and the second provides for comparing sections to sections [59]". The third one combines the first and second by taking the melodic combinations as a system of potentialities instead of a fixed chain of events, showing that "motives can recombine in various ways, expanding or contracting, to make new melodies which still conform to the traditional sound." Both contributions were acknowledged in computation, particularly in folk music pattern matching and sequence alignment [313, 310, 309, 318, 262, 28, 34, 244, 233, 264, 32].

Computational work with data also demands a comprehensive systematic breakdown of what and how we annotate and analyse, such as methodological handbook of defining basic music parameters for folk song analysis by Sergio de la Ossa [223]. He discussed the analysis of song, melodic, rhythmic and lyric structure, scale types, ranges, and other music descriptor categories. Revised concept of Alan Lomax, which was recently enhanced by Anna Lomax Wood and others, proposed an alternative method for encoding musical things [181, 322] - *cantometrics*, a numerical encoding system that can identify patterns in singing styles and musical structures across cultures, aiming to understand cultural and its social aspects. It correlates music and cultural factors, offering insights into human behavior and societal relationships. Similar principles were close to many ethnomusicologists of the time, for example, Valens Vodušek [308].

In contrast to the all-encompassing systems, David Huron, among others, focused on a single descriptor, such as melodic contour [120]. He emphasised the perceptual relevancy and importance of melody direction over absolute pitch values, and identified nine common melodic archetypes for melodic phrases. These were based on the relationships between initial and final pitch values and a median of middle pitch values, which in turn were supposedly a more perceptually relevant description of melodic character than absolute note values (Figure 3.6).

In contemporary computational ethnomusicology, with accumulated data over the decades and digitisation improvements, researchers are able to manage datasets considerably larger than those available to earlier studies, which has heightened interest in quantitative analyses. Nonetheless, there remains a strong emphasis on qualitative aspects, incorporating both content (referring to any number of musical descriptors) and context (encompassing discursive, socio-political, cultural, performance, and other practices related to the music). Many computational studies understand that music entities responsible for the observed materials "are linked by various types of relationships," which contribute to the understanding of music as a whole [271], or incorporate the information on "genre" and regions [219], or include multiple layers of music *content*, and *context* as descriptors (music and/or metadata) by observing a number of phenomenons, for example, including (super)area and (super)type information [218] and other "non-musical" traits correlated with different types of music content, such as rhythm, melody, patterns, and anti-patterns [48]. Most studies, however, remain within these realms, or extend their non-musical features to at most lyric and/or perceptual ideas, as done in [143, 147, 201]. An extensive overview of various manual and computational approaches, including several of the studies mentioned, was presented by Maria Panteli [226, 225]. Panteli distinguishes between manual and computational methods and examines both audio recordings and music notation for each approach. Her overview provides details on the size and brief description of each corpus, the number and type of descriptors used, and a summary of the main findings from each study, along with the primary references.

The results, annotations, and data representations in computational ethnomusicology are often presented as is, thus as the ground truth. However, this becomes problematic, particularly with verbally transmitted information. The limited symbolic representations and artificial classifications rarely capture the actual concept of those materials. At best, they are biased textual summaries—if one considers Western music notation and annotations as quasi-text. This can lead to significant distortions in understanding the complexities of the studied music matters (Section 2.3).

In Cowdery's words, "[a] traditional musician will not evaluate a new tune or version by comparing it to some faceless archetype [...]. To understand this process, we must look for the overlapping and flexible ways in which musicians work with their materials rather than looking for categories to impose from outside" [59]. And, while observing from the distance without imposing a certain type of a non-compatible classification could be the key contribution of computation in the long term (instead of considering only the takes of the "natives" or originate from our own biased position), this is a task yet to be mastered in computational music analyses. What is already possible, however, is to include some contextual meaning. Pendlebury stressed, that "[t]he examination of tunes in the contexts of their source documents elucidates the cultural factors that influenced their reuse over history, such as the development of print technology, military campaigns, trends in commercial theatre, and the mass production and use [233]".

The concerns raised in the presented studies, as well as in those that follow, do not have straightforward solutions. It is important to understand that the critique should not be framed as a conflict between ethnomusicology and newer computational approaches, as both fields have their own shortcomings and face their own sets of criticisms [225]. Many of the recognised challenges remain, particularly those related to restricted access to data, unbalanced collections, ambiguity in defining the ground truth, incorrect or misinterpreted metadata, and the methodologies of corpus creation, among others [225]. To address some of these issues, I will begin by examining collection practices, their influence on (meta)data representation and curation, and the challenges of their digitisation and accessibility.

2.1.2 The Fieldwork: On Collectors and Annotators

Behind the representation and analysis of the materials, there is the indispensable role of transcribers and collectors. First, they influence the materials by using the symbols (music notes and text) to transcribe a tradition that is typically transmitted orally.

Second, as Jack Goody pointed out, the transcriber becomes the "new audience," effectively distancing the informant from their usual performance context. This creates a layer of interpretation and potential distortion between the original expression and its recorded, staged, or transcribed form. As a result, the informant may be inclined to choose a performance or version that is most likely to entertain the audience, easiest for them to understand, or simplest for the informant to recall in the new context [97]".

Third, there is the influence of socio-political considerations (what should or should not be collected, and the methods used) as well as the impact of technological advancements, such as recording devices, which continually alter the way material is transmitted from its original oral form to newer media (written, transcribed, recorded, photographed, and so on). Dictation, as opposed to recorded transcription, posed even greater challenges for capturing music and dance compared to spoken text or lyrics, especially before the advent of tape recorders and cameras. This is because the notation used for documenting these standardised communicative acts was non-linguistic, less widely known, and therefore more challenging to use [97]. Although recording devices mitigated some (but certainly not all) obstacles, such as inaccuracies in repeated songs and time constraints for transcription, the symbolic notational system remained in use, along with many of its inherent problems (some of which have already been discussed).

Only by comprehending the assembled materials and the specific ideas of collection processes, along with their organisation and structuring, one can begin to explain and synthesise. Until then, these encodings and the algorithms that compare the encoded materials are, akin to Floridi's concept of genes, "a type of predictive and procedural information [...] [which as] dynamic procedural structures [...] together with other essential environmental factors, contribute to controlling and guiding the development of organisms. [...] [They represent] information for something, not [yet] about something [85]."

2.1.3 Final Remarks: On Research Framework

The processes of fieldwork, transcription, and data conservation in different periods consider different practices. Hence, when examining collected materials, the more information we gather, the closer we come to Luciano Floridi's knowledge [85]. While some points may seem distant from music research, his basic structure aids our understanding of why stopping at a primary level is insufficient. He stresses, that "[t]he idea is that information can be quantified in terms of a decrease in data deficit. [...] These are raw data, not yet semantic information. [...] When they become meaningful, they constitute semantic content [...]. When semantic content is also true, it qualifies as semantic information." The truth remains a subject for a broader discussion; nevertheless, the greater the diversity of information, the more intriguing the inquiry becomes.

The choice of collected information is not only practical but also a theoretical concern. Within the realms of ethnomusicological pursuits, as Rice points out, various truths can emerge "from different social and historical positions, interpretations of meaning, plumbing reflexively the depths of individual experience [...]." This integral and indispensable "paradigm shift," if overlooked, can become a source of confusion in ethnomusicological theory [249], as several collections are "a product of its time" [93], meaning that they are collected, transcribed, and analysed within various historical settings.

Similar ideas influenced Rice's concept of ethnomusicological theorising, which consists of 1 plus 4 steps [249]. The first, data collection, is occasionally considered the preface, a precondition that must be fulfilled before research. However, by inevitably attributing meaning to data during collection, he understood it as the initial stage. Rice defined the next four stages, *organising*, *structuring*, *explaining*, and *synthesising*, by further dividing them into 15 separate activities [249]. Among all, he was particularly critical of the common omission of the final step, synthesis, which has become even more challenging to achieve with the increasingly popular big data analyses.

2.2 Digital Archives and Datasets

The organisation, explainability, and accessibility of digital archives, along with the format, quality, and quantity of their contents, are pivotal for the computational exploration and synthesis of music materials. Although my focus is on ethnomusicological content, the following principles are applicable to other types of materials as well. By Rice, the process of organising involves four key steps: "(1) analysing and describing the data, (2) classifying and categorising the data, (3) labelling categories and classes, and (4) listing categories in some orders' [249]. In the fields of comparative and computational ethnomusicology, however, the approaches to organising music are often inconsistent [238], which can affect the quality and utility of research outcomes.

Some researchers apply Western music organising and analysis theories as their framework [295, 53, 7, 66, 1, 223], while others tailor their analysis approaches to the unique characteristics of their research material [44, 37, 222, 47]. Then there are those, who strive for a consensus between different materials, striving towards multicultural comparisons [179, 261, 133, 322, 224].

The methodologies used to describe music, such as classification, categorisation, and labelling, as well as the hierarchical ordering of these categories and classes, depend heavily on the organisation of the observed material (as shown in the referenced studies). To establish a digitised music dataset, it is thus sensible to first develop a comprehensive framework that addresses both the collection of new materials and the structuring mechanisms.

Excluding the digital archives that only consist of lists of indexed works or scans of physical material, there are currently more than 15 digitised folk song datasets, some of which are publicly available. Then there are those who discuss the state of folk song materials in digital archives [176, 292, 298, 238], or review existing datasets [190]. While it is not feasible to detail every project, I will briefly introduce the various methods used for collecting and organising practices of digital folksong datasets.

Datasets can be distinguished by content's diversity, quantity and quality of (meta)data, the motivation for digitisation, content type (notations, audio, lyrics), and varied online availability. The content can be separated into multiple tradition datasets [258, 22, 236, 322], or a selection of culturally resembling content, such as Greek [190, 228], Indian [282, 279], Dutch [146], Latin [276, 256], Basque [49], and Georgian [253, 268], purposefully focusing on ethnomusicological research [291, 237, 146, 322], pattern matching tasks [53, 149, 48, 218, 243, 52, 264, 32], and others, mood or emotion recognition tasks [190, 234, 103], and similar.

Not all data is necessarily of the same type. They can consist of any combination of text, scores, audio, lyrics (if applicable), even images or video content, and metadata, and are released under various accessibility and reproducibility licenses. Thus, after taking these factors into account and assessing the context and reliability of the data, it becomes more feasible to form a clearer understanding of the future use of digitised materials, which play a significant role in scientific discovery and practice. For instance, a dataset with downloadable scores but minimal metadata is valuable for tasks focusing on music structure like melodic pattern matching and machine learning for similarity recognition. However, without adequate metadata and supporting research, this material may not be suitable for ethnomusicological studies beyond the specific collection project. Conversely, materials with abundant metadata, images, and other computationally unreadable or non-downloadable formats are limited in the scope of computer processing. The data reluctant to the open-source principles is commonly considered to be an inadequate contribution to the scientific community [320, 185, 316]. Below, I provide a summary of available datasets that go beyond publicly available scans, with information on content, accessibility and other commentary.

2.2.1 An Overview of Digital Archives and Datasets

The Appendix A lists 46 distinct collections, ranging from broad compilations like The Essen Folksong Collection, Dunya, and RISM to those focused on specific topics such as Jingju, Dutch folk songs, and Carnatic music. While most datasets are publicly accessible online, many come with usage restrictions. Of these collections, the majority is licensed under some form of Creative Commons Attribution, while some provide no copyright or licensing information, and the remaining are partially open, typically available for research and educational purposes, but with stricter limitations on commercial usage (Figure 2.2). The materials include a combination of notated scores, audio, lyrics (where applicable), metadata, and annotations (Figure 2.1a), and they vary in the level of research conducted to further elucidate or promote each collection.

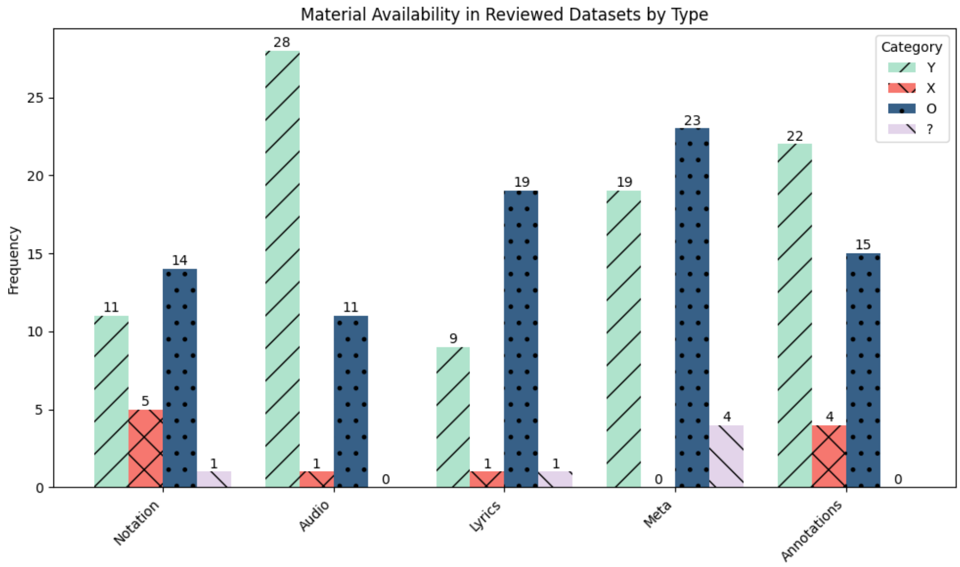
The Appendix A provides an overview of selected digital datasets and archives, together with a primary reference, URL, as detailed information on licensing as possible, a short description of the content, the specification of content type (notated scores, audio, lyrics, metadata, annotations), as well as a set of informal annotations on usage categories (for additional information on music corpora and relevant studies, refer to [225, 226]).

The latter assume 4 different user groups based on the following criteria: *Enthomusicologically appropriate* materials are materials that are well-documented and accompanied by sufficient metadata. The interface and sources should be accessible without requiring computational skills and should be readable in standard programs such as Excel, PDF viewers, and similar tools. Availability to download the data is not necessary. *Education* materials must comply with the similar standards that apply to ethnomusicological research. An interactive platform and/or well-explained materials are valued, and the availability of annotations on the platform is considered an added benefit. *MIR* research requires at least some data, such as music scores, recordings, or metadata, to be easily downloaded without special effort, and this data should not be copyrighted, at least for research purposes. The dataset should contain more than just a few examples, and annotations of technical descriptors of those sources are an added advantage. For *general public*, the interface or webpage must be user-friendly, and data descriptions should be clear and easy to understand for everybody. Video and audio recordings are highly beneficial. In certain cases, digital archives can serve as "reference" tools for searching, such as index databases.

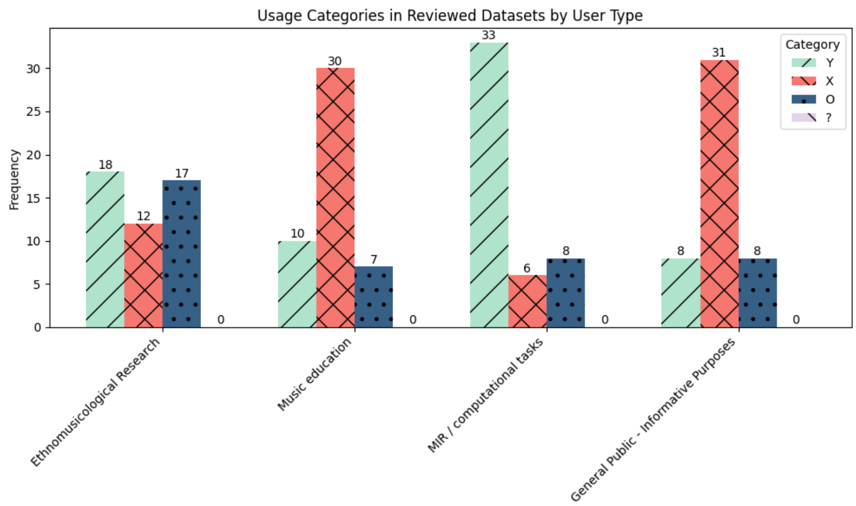
To name a few examples (see also statistics on these in Figure 2.1b), the CompMusic Dataset of Indian Music Tonic, which features annotated tonic pitches of the lead artist, is primarily suited for MIR research, while the Dutch Song Database, with its comprehensive descriptions (but somewhat inaccessible scores and annotations), better suits ethnomusicological and related studies. Only a few of these collections, such as Etnofon or the Digital Collections of the Library of Congress, are usable for the general public. The content types and these are sorted into 3 general categories (Y=available/appropriate; X=unavailable/not appropriate, O=partially available/appropriate).

In addition, I provide a detailed description of the key dataset for this thesis–Slovenian folk song ballads–explaining the collection practices that influenced the materials (Section 3.2.1), digitisation process and its representation on digital platform Dezrann with data availability and licensing details (Section 3.4), as well as introduce the annotation system and statistic on these annotations (and other (meta)data) (Section 3.2.2 and 3.3).

Before moving to the said corpus, I will highlight related work that had an immense impact on structuring such datasets, including the discussions on (digital) music representation and annotations, as well as introduce grounds for exploring and analysing these materials, focusing on pattern matching methodology.



(A) Most common types (Y=yes and O=limited) of materials in the reviewed datasets are audio recordings, while the least present are the scores or music notation. Some type of metadata is available in all but four examples, where this was not possible to determine (?=undetermined).



(B) The reviewed datasets and archives most commonly tailored to the needs of MIR and other computational tasks, while they are not suited for general public or music education purposes.

FIGURE 2.1: Statistics on datasets' and digital archives' material availability and potential usage by categories. See Section 2.2.1 for further explanation on the categories and labels.

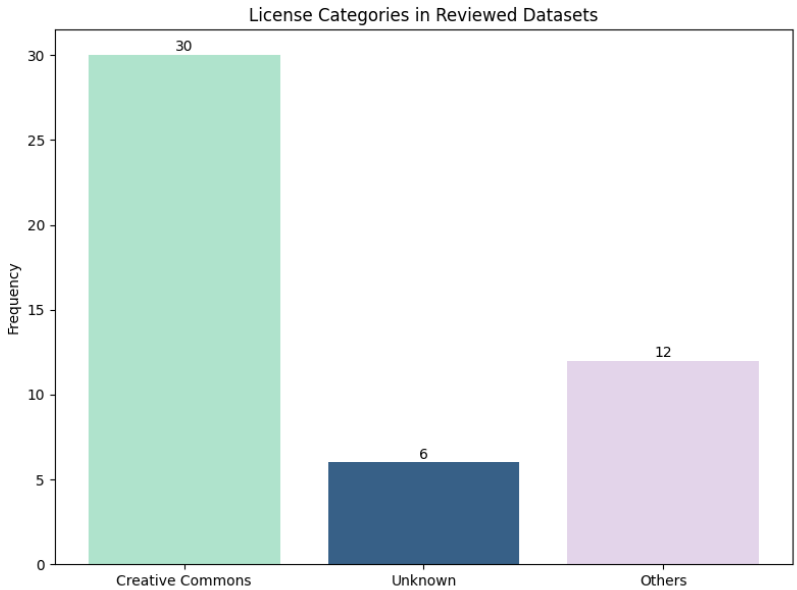


FIGURE 2.2: The types of licensing or copyright protecting the materials in the reviewed datasets in Appendix A. The most common form is some type of Creative Commons license (leftmost, light green) for at least part of the materials, followed by other arrangements such as specific usage descriptions, alternative licenses, and similar agreements (rightmost, light purple) . In six cases, the copyright information is not provided (center, dark blue).

2.3 Music Representation and Annotation

The diversity of music information has given rise to numerous debates regarding its representation at all levels, ranging from music notation and audio signals to broader, non-music-specific structures, such as the development of typologies and taxonomies.

Starting with the larger, typologies are conceptual frameworks primarily employed for systematic comparisons. They are constructed categories that define objects in opposition to something else, such as in distinguishing contours [2, 120, 7, 58], cadences versus non-cadences [147], and patterns versus anti-patterns [48].

Taxonomies, on the other hand, are broader, quantitative classifications. They sort individual objects into large categories based on measurable characteristics and define their hierarchical or relational positions within these categories, forming homogeneous groups. Examples in music include the Hornbostel-Sachs musical instrument classification [116, 169], and classifications of folk music genres [243, 219, 261].

Additionally, it is worth mentioning music ontology, which comprises structured frameworks for organising and categorising musical knowledge, defining relationships between various musical concepts such as genres, instruments, and performances. These frameworks facilitate tasks in Music Information Retrieval (MIR), music archiving, and analysis, thereby significantly advancing digital (ethno)music -ology and enhancing music recommendation systems, as elaborated in [242, 241, 250, 238], among others.

In this section, I will primarily address three categories–typologies, quasi-taxo -no-mies, and ontologies–in relation to music annotations. None of these represent fixed or stable structures; therefore, to fully understand them within the "digital" realm, it is essential to begin by exploring various types of music data encodings and data representations.

2.3.1 Music Representation

In the context of annotating and analysing musical content, a significant challenge emerges concerning the representation of data. This issue is exacerbated by the inherent ambiguity of the concept of music itself, which eludes any uniform description. Images and texts are represented through better-defined means—pixels and alphanumeric symbols, respectively. In contrast, the domain of music lacks a universally accepted "best" representation, and thus faces many challenges [65, 113].

Music can also manifest in the forms of various notational systems, text, speech, images, video, or audio, each necessitating distinct representation strategies and further fragmenting the unity of music representation. According to [306], the representation levels of music matters can be roughly split into four categories, which are either *physical*[1], *signal*[2], *symbolic*[3], or *knowledge*[4]. Their system omits the direct incorporation of image or video.

In the scope of this thesis—which primarily focuses on music notation and metadata—I find the use of their *symbolic* category to be the most crucial for further consideration.

The choice of music representation, much like the datasets that emerge from them, is influenced by the context of its application—whether for archival purposes, digital production, performance settings, or other uses. Some of the most common representation systems in computational contexts include MusicXML (which is also a predominant input format for this thesis), .sib, Finale, MIDI [254, 76, 117], Base-40 [110], ABC, LilyPond [220], GUIDO music notation [115, 246], and MEI [250].

Technological advancements, artificial intelligence and machine learning are continually driving the evolution of music representation, enabling a more nuanced analytical capabilities and making any standardisation in the scope of data representation challenging. Furthermore, there are cultural, socio-economic, and ethical considerations, particularly in how music from diverse traditions is represented and preserved. Last but not least, music representation has been greatly impacted by the

[1] These are partial characteristics of sound objects and scenes in terms of geometrical descriptions and include other physical properties such as mass, elasticity, and viscosity.

[2] Signal representations treat music as a continuous flow of information both in time and amplitude. This category typically includes audio signals that can be digital or analog and are essential for applications that require the raw audio data without any abstraction or symbolic interpretation.

[3] These are discrete musical events such as notes, chords, and rhythms. Symbolic representations are content-aware and describe events in relation to formalised music theory concepts.

[4] They are generally textual structured formalizations of knowledge about musical objects, often utilised in digital music libraries and MIR. As they consist of both music representation formats AND music annotations. The more complex the structures, the further they move away from topology and turn to taxonomy. These are further treated in the Section 2.3.2.

interdisciplinary connections between music and fields such as mathematics, psychology, and linguistics, each of which once again re-defined the way music should be represented and used (see different dataset usages in aforementioned [226]).

All of this combined presents challenges for data annotation, as the content of the representation dictates how annotations are formulated and what they encompass. Moreover, these are fundamental requirements for computational analysis, where diverse representations directly affect the complexity of data modeling and analysis, thereby influencing the outcomes of research endeavors.

2.3.2 Music Annotations

Music annotations involve adding remarks, symbols, or comments to music representation (score, audio representation, image, text, and other). They serve as a guide for music interpretation, performance, as well as are found useful for most algorithmic processes. They specify dynamics, articulation, technical instructions, and emotional or stylistic ideas, encompassing both musical and non-musical elements. Often referred to as descriptors or features, they detail various aspects and provide critical insights for executing and understanding the score (and the music matters behind it).

They serve as fundamental ground for most computational music research [45, 299], and are, within the realms of MIR, commonly categorised into two main types. The first type includes *foundational* components of music such as melody, harmony, rhythm, and texture, which outline the formal structure and other musical elements. Along with others, these aid in constructing higher-level representations, often formulated as *multi-dimensional or multi-layered systems or protocols*. Such frameworks behave similarly to a taxonomy or rather an ontology, as they not only represent individual elements but also delineate and interlink relationships within individual compositions or across multiple corpora. They enhance the efficacy of computational analysis and technically (although not always efficiently) facilitate the exchange of data among corpora and researchers.

Recent scholarly efforts broadened the scope of semantic annotations in music research, extending beyond traditional links between metadata and musical components to include annotations that capture mood and emotion. This expansion has enriched the field of music matters, as indicated by studies such as those detailed in [103].

Foundational Music Annotations

This category pertains to annotations which, within the domains of MIR, computational music analysis, and digital libraries for musicology, are regarded as fundamental for the exploration of various musical patterns. These annotations delineate and identify structural elements such as motifs, phrases, and sections, and they also label a diverse array of musical events.

In certain instances, the annotation process[5] is entirely automated [232, 104, 38, 193, 214]. Conversely, there are several cases, where the annotation is (partially) manually curated [265, 312, 263, 310, 229, 51, 18].

The structure of music can be expressed as in describing isolated events, individual layers, or, on the other hand, summaries or groupings of music events, for example by using Schenkerian analysis [128, 194, 90, 15] or the Generative Theory of Tonal Music [172, 86, 108, 90, 111]. The latter relies on both, music theory and a cognitive, perceptual understanding of music structures.

It is not uncommon to see Western classical music theory applied to define these annotations. However, the ethnomusicological domain, as well as more recent contributions in MIR [98], indicate that there may be discrepancies. When considering the perception of music across different languages and cultures, verbally-transmitted music, or music outside the more "established" music theories, there are issues related to the naming and the purpose of naming various aspects of the musical phenomenon in such applications and generalisations.

Phrases. It is commonly understood that a phrase represents a segment of a larger music piece. This segmentation is delineated through a range of musical descriptors such as melodic, rhythmic, harmonic, and lyrical elements (if present), along with perceptual characteristics and other musical factors. Hence, although being understood as a segment of something, its definition can vary greatly among corpora, thus there is no single universal conception of this descriptor.

To annotate phrase segments in a music piece, the first appearance of a certain sequence is usually considered as a reference for all subsequent ones. The borders of phrases in music that is close to Slovenian folk songs can be determined by either rhythmic or lyric breaks, cadence detection [147], and similar, or a combination of multiple descriptors.

In monophonic music, phrases are primarily compared by their melodic material. Although the rules for annotating variations are not entirely standardised, the relationships between phrases within a single music piece are usually marked using (capital) letters of the Latin alphabet in alphabetical order, starting with A. The next contrasting part is labeled as B, and so forth [223, 81, 32]. To describe minor variations, extensions or further subdivisions, notations such as A', A+, lowercase letters (a, b, ...), or similar, are used [223].

The same approach can be utilised for annotations of phrases based on rhythmic properties and verse/rhyme structure [93]. In cases where multiple patterns, such as melody and lyrics are observed separately within the same piece, the annotation system can be adapted to differentiate between them. One approach is to use capital letters for one pattern and lowercase letters for the other (if not already applied to another annotation type). Alternatively, letters from the beginning of the alphabet

[5] An extensive study on computational methods for music structure analysis and segmentation was conducted by [257, 182].

can be used for one pattern type, while letters from the middle of the alphabet for the other, enabling a clear distinction between the two types of patterns [156, 93].

Melodic Contours. The annotations of phrases can be labelled according to the note-to-note or interval-to-interval melodic sequences; however, they can also be annotated by examining the broader melodic structures, such as melodic contours. In the case of the Slovenian folk song ballads corpus, the transcriptions of many songs were unreliable (Section 3.2.1). Even those that closely resembled the field recordings only stand in place of approximate transcriptions of sung pitches and merely refer to the approximate structure of all verses, rarely annotating the minor differences among individual verses[6].

The contour annotations, providing an overview of the pitch movement of individual melodic phrases, are thus much more suitable for analysing verbally-trans-mitt-ed music and recordings, where determining absolute pitch values is challenging or even redundant (Section 2.1). Although this can become more difficult if one is to consider non-monophonic examples, as well as music with rather extensive melodic lines that may not be easily split into smaller segments.

Annotations of contours can be symbolically represented as sequences of the characters 0 (unchanged), + (ascending), and - (descending), with each symbol indicating the pitch relationship to the subsequent tone [7]. Alternatively, annotations can utilise cosines [58], curves [37], or predefined types of arches [120]. This thesis adopts the latter, Huron's nine types of contours, these being ascending, descending, horizontal-ascending, horizontal-descending, descending-horizontal, ascending-horizontal, convex, concave, and horizontal. They are calculated by analysing a completed sequence (in our case, a phrase), by taking the first and the last value (in our case, MIDI value) of the pitch sequence and comparing it to the average of all the intermediate pitches (see Section 3.2.2).

Rhythm, Beat, Meter. Similarly to melodies, rhythmic patterns can be annotated based on their structure, length, and combinations.

The simplest or highest differentiation among rhythmic structures is to divide and annotate them into binary or ternary groups based on the meter. These may pertain to an entire tune or a specific phrase or measure, which is common in tunes with multiple meters.

Rhythmic values can be annotated as individual temporal events (especially in terms of audio analysis), or with values relative to the meter or a fundamental unit (typically, a quarter note equals 1.0, an eighth note equals 0.5, and so on). They can also be annotated as approximations, where subsequent rhythmic events are described in relation to the preceding one—whether they are longer, shorter, or of the same duration. Moreover, specific systems tailored to individual corpora's nuances

[6] Meaning, if verses 3 and 10 were sung differently, but the rest followed a similar structure, the latter would be favoured and transcribed.

can facilitate this process. For instance, [38] undertook a study annotating various types of Turkish usuls based on their rhythmic combinations. On the other hand, [7] investigated the nuances of upbeats and downbeats at different levels. Furthermore, [26] explored additional models for computational analysis of rhythmic patterns.

If the music is not taught within strict theoretical frameworks, such as the aforementioned Turkish classical music, Western classical, jazz, or popular music, it can be challenging to annotate exact rhythmic patterns. This difficulty is already present in some forms of jazz and related music, let alone entirely verbally-transmitted folk songs. In line with what is described in Section 2.3.2, rhythmic transcriptions of field recordings and live performances (of folk songs) are rarely a reflection of a complete performance, but rather a sketch or a summary. In our case, this is particularly evident in the rhythmic aspect, as many subjects did not maintain a strict rhythmic structure while singing. They would often pause briefly to breathe, provide spoken explanations, return to the beginning of a verse, or recall the subsequent lines and verses, and then accelerate to avoid forgetting, and so forth.

Harmony and Scale. In MIR, as well as in the practices of music theorists and musicologists, various types of harmony annotations exist. These include chord and key recognition labels, as well as scale analysis annotations, employed to extract and annotate harmonic information from the provided music data (see examples of annotating scalar and harmonic degrees in Figures 2.3, 2.4, and 2.5). While these annotations are predominantly conducted on polyphonic music (for example, [235]), the formal functions can also be analysed in monophonic tunes, based on the prevalent scale or mode.

In the case of annotating Slovenian folk songs, although some harmonic labels were initially incorporated and analysed [32], comparisons among different understandings of harmonic frameworks revealed substantial differences. Firstly, these differences in various styles of annotations were too significant to provide a basis for deriving "averages". Secondly, these songs were not conceived within the frameworks of Western classical music forms at their time/origin. Consequently, harmonic labelling was omitted, raising questions about the purpose and potential outcomes of such an endeavour.

The same issue occurred with scales. Following [223, 179] and similar studies, the provided annotations include only the number of different tones or pitch classes, rather than names of the scales. While theoretically possible, naming the scales would offer little insight and could even be misleading in relation to the actual object of observation, as well as would unnecessarily impose labels foreign to that particular music practice.

Multidimensional Annotation Models and Protocols

Regardless of the content type, be it music collections, literary works, or any other form of data, the principles of annotation have developed to ensure the accessibility,

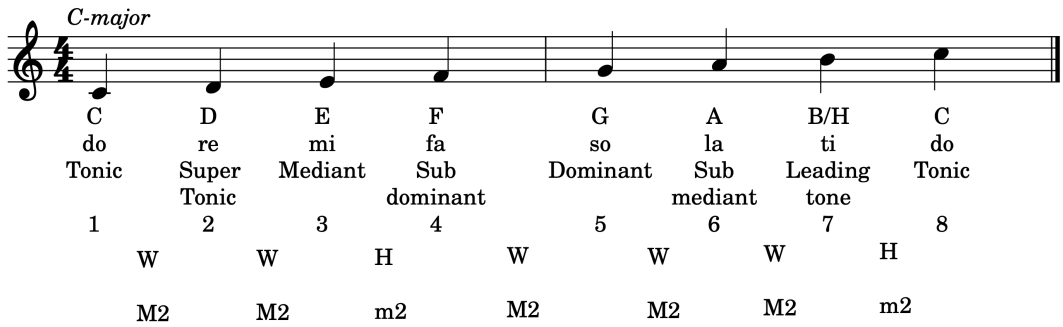


FIGURE 2.3: This figure illustrates various methods for annotating scale degrees, presented from top to bottom. These include representation as C-major as a single descriptor, or representing each scalar degree alphabetically; using a solfege-based system (with "do" indicating either the first degree of any scale or fixed to a specific pitch, such as C, regardless of its position in a scale); through scale degree notation; enumerated based on position within the scale (using Arabic numerals or alternatively, Latin letters); depicted through chromatic relationships (denoted as W for whole step, H for half step); and via intervallic relationships, specifically showing only major seconds (M2) and minor seconds (m2), among others.

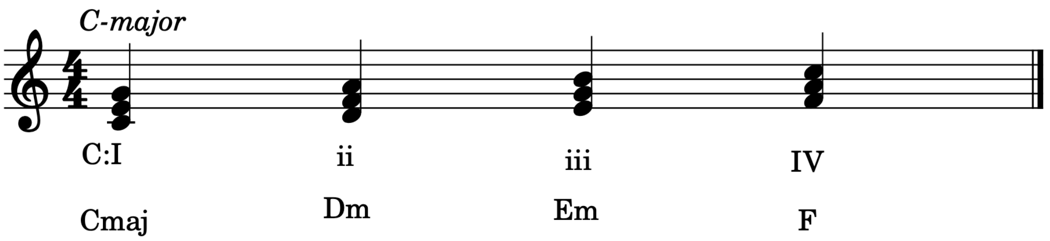


FIGURE 2.4: This is a brief example of annotating chords with either scale degrees in a certain key or by using the chord names. An extended version of harmonic annotations is seen in Figure 2.5

preservation, and usability of music data. Unlike the annotations discussed so far, *multidimensional* annotations encompass and connect multiple hierarchically varied musical events, condensing them into a single annotation (per score) or a bundle of annotations (per corpus).

Most digitised material becomes particularly valuable when it extends beyond the function of mere "storage" to become a foundation for computational processing and other research endeavors, such as searching, analysis, and generation. To facilitate this, especially with diverse data or corpora, an additional set of protocols is required. An example of this is the OWL (Ontology Web Language) [200] and its implementation of the CIDOC Conceptual Reference Model. "The OWL facilitates encoding and reasoning over a genre ontology, while the CIDOC enables a representation of complex spatial containment and proximity relations among geographic regions," navigating through different types of cultural heritage information or, for instance, study different toponyms [92], and systematically arrange our content based on its "complexity".

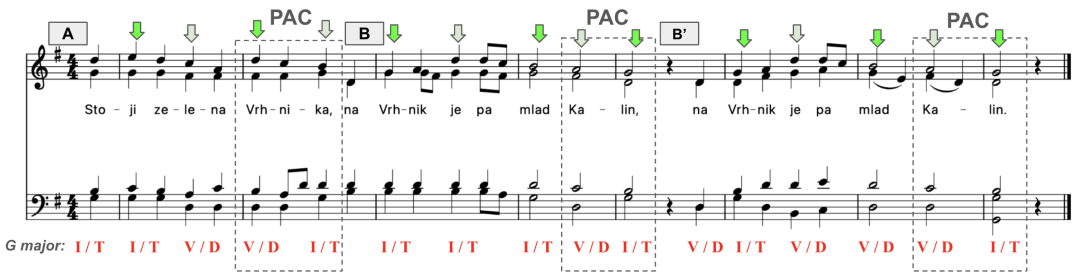


FIGURE 2.5: An example of a polyphonic Slovenian folk song variant titled *Prisilno daleč omožena / Forcibly Wed from Afar*, sourced from the archives of the Ethnomusicology Institute ZRC SAZU, with harmonisation by Franc Kramar. The figure illustrates various methods of annotating musical elements, including strong beats (indicated by green arrows), types of cadences (PAC=perfect authentic cadence), and two methods of annotating harmonic degrees (Figure 2.4). Additionally, phrase labels (ABB') are shown. While a polyphonic song provides a solid basis for annotating harmony, this approach may not always be feasible or appropriate.

The music ontology of the semantic web has been addressed by several authors, such as [241, 238], of whom some pointed out that ontological and taxonomical categories might suffice for a number of music data, however a more detailed labelling to specific cultures and use cases is required. This inevitably results in (partial) data incompatibility across studies, despite using the same reference models (CIDOC, FRBR, etc.).

Although this thesis does not focus on constructing multilayered annotation systems, it acknowledges the importance of several common systems in the field. These systems are particularly relevant to the organisation of materials on platforms such as Dezrann (Section 3.4), where our data is stored and made available for interaction.

One of the longest-standing structures in music is MEI (Music Encoding Initiative) [250], with its more recent extended version, MELD [315, 317] (Figure 2.6). Both originate from the TEI (Text Encoding Initiative) and construct a web of events in an XML-like format, which are interrelated. These systems support textual, music, audio, video, and other multimedia description entries, all connected through interrelations. While these do not suffice for all types of research, such annotations enable a high level of transparency in corpus description and allow for a precise definition of the relationships between a musical piece and all accompanying data (metadata, album cover, video, lyrics, collection, and so on).

Lastly, it is worth mentioning Measure Maps [99], a versatile and streamlined format developed for representing bar-level information in digital musical scores. This is one of the approaches, which enhances the accuracy and computational intelligibility of comparisons and analyses across various encodings. As an interoperable standard, Measure Maps significantly bolster the functionality of detailed music encoding systems such as MEI and others.

All these ideas relate to what Oliver Brown and others, particularly in cognitive

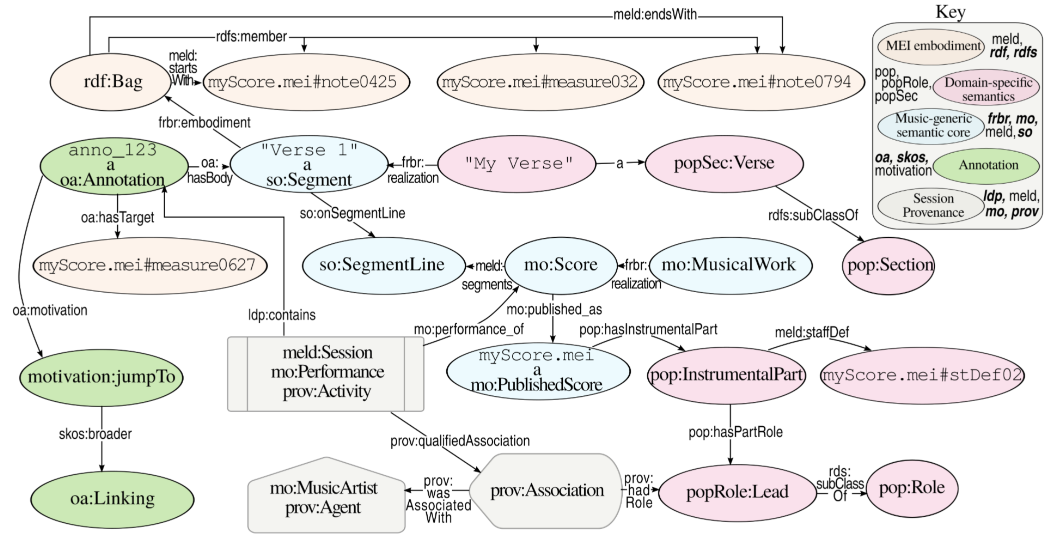


FIGURE 2.6: The MELD semantic framework [315]. Annotations address music segments embodied in a published score (MEI) resource. "A music-generic core is linked to but separable from domain-specific entities instantiating concrete music sections. **(Top right)** Key: External ontologies in bold italics. *frbr:* Functional Requirements for Bibliographic Records Core; *ldp:* Linked Data Platform Vocabulary; *meld:* MELD Vocabulary; *mo:* Music Ontology; *motivation:* MELD-specific oa motivations; *oa:* Web Annotation Ontology; *pop, popRole, popSec:* Pop music domain-specific semantics; *prov:* PROV Namespace; *rdf:* RDF Concepts; *rdfs:* RDF Schema; *skos:* Simple Knowledge Organization System; *so:* Segment Ontology."

science and AI, discuss within the framework of *conceptual spaces*. In these multidimensional spaces, broadly speaking, we perceive some objects as being closer to each other than others. Brown illustrates this with the example of pitch and harmony in music, where certain notes (such as octaves and fifths) are perceived as harmonically closer than others (like semitones), despite their physical proximity on instruments like the piano or guitar. The situation becomes vastly more complicated when comparing different kinds of music, forms, and cultures. Therefore, it is clear that these spaces, however broad, do not encompass every situation but rather provide an understanding of limited material within specific frames of music reasoning [36].

2.4 Pattern Matching in Text Algorithmics

We are constantly seeking patterns to organise the data we perceive daily, whether it pertains to ordinary tasks or complex scientific inquiries. One field that has significantly advanced in pattern discovery is bioinformatics, which merges disciplines such as biology, chemistry, physics, computer science, software engineering, mathematics, and statistics to address the complexities inherent in vast biological datasets. In bioinformatics, a wide array of computational, statistical, and programming methods are applied to model biological phenomena and tackle essential questions. Notably, standardised analytical frameworks, particularly in genomics, are utilised for tasks like gene identification and related analyses, such as elucidating the genetic underpinnings of diseases, distinctive adaptations, and desirable traits. These methods also aid in examining variations across populations and organising protein sequences.

Bioinformatics employs various modelling principles, including text mining, image processing, and signal processing, to approach diverse data formats. These principles have extended beyond biology and genetics to fields like clinical medicine, agriculture, environmental science, pharmaceutics, finance, and digital humanities. In digital humanities, they aid in tasks such as text analysis, historical research, linguistics, speech recognition, audio compression, and music analysis. The latter, which forms the central focus of this thesis, frequently employs string matching or similar alignment approaches to identify patterns in melodic and rhythmic sequences.

This thesis will focus on two main types of pattern matching: *exact* and *approximate* string matching (for extensive details on these, see [102, 60]). Initially, it will provide a brief overview of selected pattern matching tasks in bioinformatics. Subsequently, it will explore how these methods can be adapted for music pattern matching tasks in MIR. The selection between exact and approximate string matching depends on the complexity of the task, as well as the type and context of the data used. Each category encompasses several methods that differ in performance, capability, and efficiency, achieving various levels of accuracy.

In the following subsections, I will briefly describe each of the two categories, accompanied by some examples. Initially, I will explore these categories within the realms of text algorithmics, and subsequently, within the realms of MIR research.

The general task we address here is that, given:

- A pattern p of length m starting at position 1,
- A text t of length n starting at position 1,

We try to find all occurrences of the pattern p within the text t. This involves determining every position i in t such that the substring of t starting at i and of length m is equal (or similar, see Section 2.4.2) to p.

2.4.1 Exact String Matching

There can be two further subdivisions of exact string matching[7] - *online* and *offline*.

First, *online* exact string matching methods process the input text incrementally, thereby providing immediate result without necessitating or storing and sorting the entire text in advance. Some of the algorithms for this task are Aho-Corasick, Boyer-Moore, Rabin-Karp and Knuth-Morris-Pratt (KMP).

Conversely, the *offline* exact string matching algorithms, which typically preprocess (and store) the text, utilising data or indexing structures such as suffix tries,

[7] The categorisation presented here is not the sole method for distinguishing between various exact string matching algorithms. An alternative categorisation, accompanied by explanations of several algorithms discussed in this chapter, can be found in [107].

suffix trees, and suffix arrays. These structures allow for a more time efficient repeated searches within the same string structures. Suffix trees, for instance, facilitate quick searches for substrings, while the Burrows-Wheeler transform (BWT) with FM-index is especially useful for rapidly matching patterns in large, compressed text databases.

Online String Matching

Knuth-Morris-Pratt (KMP). One of the most known algorithms for this type of pattern discovery was proposed by Donald Knuth, James H. Morris and Vaughan Pratt [137, 61]. The idea behind their algorithm is that upon encountering a mismatch after a series of matches, there is already knowledge of certain characters in the text for the subsequent window[8]. This allows the algorithm to bypass rematching the characters that are guaranteed to match, thereby enhancing the search process.

First, KMP pre-processes the pattern p to create a partial match table, which indicates the longest proper prefix[9] of the substring $p[1...m]$ which is also a suffix[10] of this substring. Then, in contrast to naive approaches[11], it uses the partial match table to skip unnecessary comparisons in the string t (Figure 2.7).

It operates with a worst-case time complexity of $O(n+m)$, where n is the length of the text and m is the length of the pattern. The space complexity for storing the partial match table is $O(m)$.

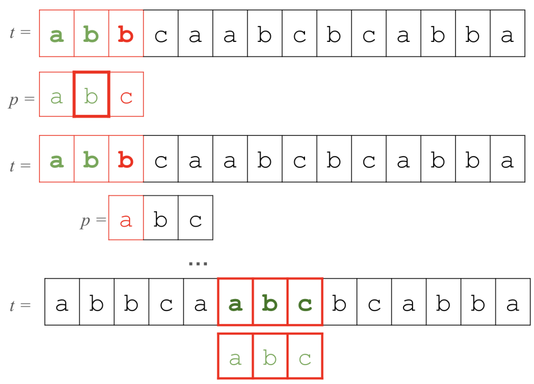


FIGURE 2.7: An example of the KMP algorithm. A pattern p = abc is searched in text t = abbcaabcbcabba. The algorithm initially checks from position 1 of the text t, and upon encountering a mismatch at position 3, it shifts to the next possible position, allowing it to resume matching from the third position onward. Consequently, the algorithm successfully identifies p at position 6 in t.

[8] In this context, the "window" refers to the current segment of the text that is being compared to the pattern as the algorithm slides the pattern across the text.

[9] A "prefix" here is any substring that starts from the beginning of the pattern and extends to any intermediate point.

[10] A "suffix" is any substring that ends at the end of the pattern and begins from any intermediate point.

[11] The naive approach to pattern matching involves straightforward, brute-force methods that compare each element of the pattern to each segment of the text or sequence, without any preprocessing or optimisation.

Rabin-Karp. It is another online exact pattern matching algorithm proposal by Richard M. Karp and Michael O. Rabin, which employs hashing to search for one or multiple patterns within a text [240]. It does so by calculating the hash of the pattern and compares it against the hash values of substrings in the text of equal length to the pattern. If a hash match is detected, due to collision handling, it performs a verification of the match through a detailed comparison to finally confirm the exact match. Unlike naive methods and similar to KMP (with its partial match table), it allows the algorithm to filter out non-matching substrings quickly, reducing the number of direct character comparisons (Figure 2.8).

This algorithm has an average and best-case time complexity of $O(m+n)$, where m is the pattern length and n is the text length, and its worst-case complexity rises to $O(mn)$.

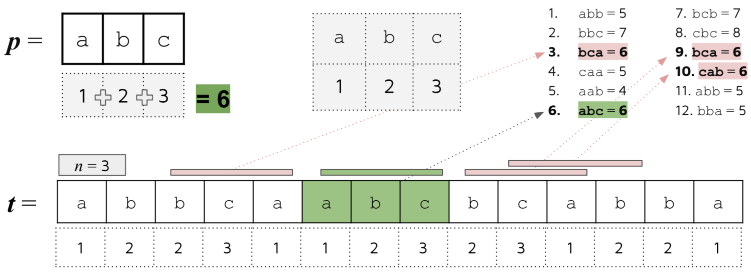


FIGURE 2.8: An application of the Rabin-Karp algorithm, searching for the pattern p = `abc` in t = `abbcaabcbcabba`. In both cases, the algorithm first calculates the hash values for the query pattern p and the subsequences using predetermined hash values for each character: `a` = 1, `b` = 2, and `c` = 3. For text t, the hash is computed for every 3-gram, meaning a substring of length 3. For example, given a string of "`abcdef`", the 3-grams are "`abc`", "`bcd`", "`cde`", and "`def`". Within all hash values of trigrams of t, it detects four instances, where the hash values match that of the query p. The algorithm then verifies these matches by comparing each candidate subsequence (meaning each trigram, a portion of t) directly with the query pattern p. Only those matches that also match the sequence of characters in trigrams are returned as valid matches. Thus, in this case, the sequence "`abc`" is confirmed wherever it appears exactly as such in the text - at position 6 of t.

Offline String Matching

Suffix Trie, Suffix Tree, and Suffix Array. A *suffix trie* is constructed as an uncompressed structure where each node represents a single character, with every suffix represented as a path from the root. The time complexity for building a suffix trie is $O(n^2)$, and the space complexity is also $O(n^2)$, due to the need to store each character of every suffix separately.

On the other hand, a *suffix tree* (Figure 2.9, also [197]) is formed as a compressed structure with branches for each suffix. The time complexity for building a suffix tree is $O(n)$, and the space complexity is also $O(n)$, where n is the length of the string.

Third, unlike the suffix trie and tree, which are hierarchical data structures with nodes and edges, a *suffix array* is a linear array of integers representing the starting positions of the suffixes in lexicographical order (Table 2.1). The time complexity for

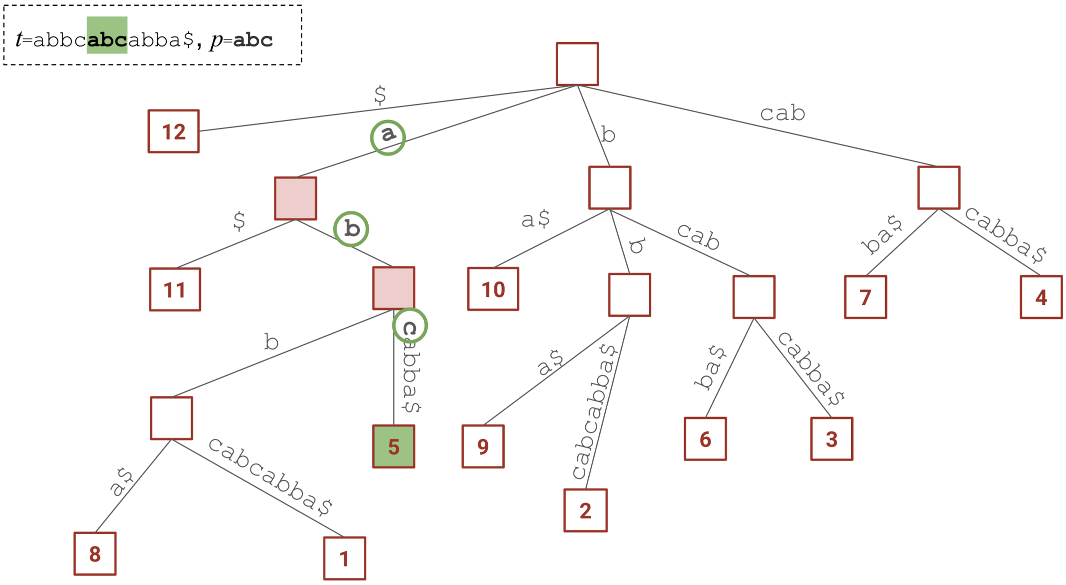


FIGURE 2.9: An offline pattern matching algorithm searches for the pattern p = abc in a suffix tree of t = abbcabcabba$. Appending "$" (or any other "special" symbol) to the string ensures proper lexicographic ordering of suffixes, makes all suffixes unique, ensures correct ranking, and clearly defines the end of the string for pattern matching. From this string, we first build a suffix tree (as opposed to a suffix trie, where each character in the suffix is a node, and each suffix is a path from the root), where, to achieve linear space complexity, labels of branches are pairs indicating the starting position and length of the substring they represent, rather than storing each substring explicitly. This way, the space used is proportional to the length of the string, not the number of suffixes. The algorithm starts at the root and progresses to p_1="a". It follows the nodes of t down the branch by progressing through letters in p, until matching all letters of p with (a part of) t. Due to its efficiency, the algorithm only needs to iterate through 2 nodes until finding the exact starting position (5) of p in t.

constructing a suffix array is $O(n \log n)$ or $O(n)$ for constant size alphabet, using advanced algorithms, while the space complexity is $O(n)$. A more detailed description of variations of such arrays was provided by [61].

Burrows-Wheeler Transform and FM-index. Burrows-Wheeler Transform (BWT) algorithm [40], also known as block-sorting compression [192], is used to compress the input data. It re-organises the characters of a text in a reversible fashion. Unlike some of the other sorts, it has the ability for the original text (or string) to be reconstructed in reverse by storing only the data in the last column (observe the last column of the sorted rotations in Table 2.2). The transformed text, denoted $BWT(t)$, is composed by concatenating the final characters from each lexicographically sorted cyclic rotation of t (Table 2.2). This transformed text is not only easier to compress but can also be efficiently stored using $nH_k(t) + o(n)$ bits, where H_k is the k-th order empirical entropy of t [251]. Its convenience and potential applications as well as optimisation were reviewed, explored, and extended by [192, 3, 255, 251], among others.

Building upon the BWT, the FM-index (Full-text index in Minute space) by [79,

Index	**Start Position in** t	**Sorted Suffix**
1	15	`$`
2	14	`a$`
3	5	`aabcbcabba$`
4	11	`abba$`
5	1	`abbcaabcbcabba$`
6	6	`abcbcabba$`
7	13	`ba$`
8	12	`bba$`
9	2	`bbcaabcbcabba$`
10	3	`bcaabcbcabba$`
11	9	`bcabba$`
12	7	`bcbcabba$`
13	4	`caabcbcabba$`
14	10	`cabba$`
15	8	`cbcabba$`

TABLE 2.1: Suffix array for the string t = `abbcaabcbcabba$`. Searching the pattern p = `abc` by using the suffix array, starts at the first position (position 6 in t) at **index 6**, and in other positions at **index 3**, **index 5**, **index 9**, **index 10**, and **index 13**. To find the **exact** occurrences of p, we start the binary search for the substring `abc` within the suffixes listed, and then locate the suffixes that start with `abc` by checking the sorted suffixes. For example, we begin by comparing `abc` with the middle suffix in the array. If the suffix matches, note the position. If the suffix is lexicographically smaller, move to the upper half of the array, or else, if the suffix is larger, move to the lower half of the array.

78, 278], integrates a count table C and a function $Occ(p, i)$. The table $C[p]$ counts lexicographically smaller letters of each character, and $Occ(p, i)$ tracks the occurrences of p up to position i in $BWT(t)$ (see Tables 2.3 and 2.4). This architecture enables efficient pattern searching through fast backward searches and substring queries without fully decompressing the string. The FM-index optimises both time and space, performing searches in $O(m)$ time, where m is the length of the query p and n is the length of the text t, while requiring less memory compared to earlier methods.

The FM-index normally consists of BWT string, first column (F), or rather that (F) represented by the count array (C) (Table 2.3), and/or the occurrence table (Occ) (Table 2.4). The latter logs the frequency of each character in the BWT up to each index i, which is used for dynamically determining character positions during backward searches. The count array, if used, can tabulate cumulative counts of characters in the BWT that are lexicographically smaller than each character, and facilitates swift range calculations for backward search operations. However, given the potential for high space consumption, these are frequently replaced by bit arrays as a less space-consuming alternative to achieve similar functionality.

Thus, bit arrays (Table 2.5) can, especially when integrated with $rank\_c(w, i)$[12], which calculates the number of occurrences of character c up to a specific position i in

[12] w= string, i = position

the string w, replace the aforementioned components (an example of that is exemplified by the functions available in the SDSL library [91]). Rank operations determine the number of occurrences of a character up to a specified position, such as using $rank_b(\text{``}a\mathbf{b}aca\mathbf{b}caaca\text{''}, 10) = 2$ to count how many times $b$ appears up to the 10th position in a selected string, while select operation finds the position of a specific occurrence, for example, $select_a(\text{``}abac\mathbf{a}bcaaca\text{''}, 3) = 5$ identifies the position of the third occurrence of a.

While both approaches offer distinct benefits, bit arrays equipped with `rank` and `select` structures may sometimes prove more space-efficient than an FM-index built out of count, particularly when handling texts in which certain characters occur infrequently or when the alphabet size remains relatively modest—such as in text sequences encoding up to 12-tone melodies.

Cyclic Rotations	**Sorted Rotations**
abbcaabcbcabba$	$abbcaabcbcabb**a**
bbcaabcbcabba$a	a$abbcaabcbcab**b**
bcaabcbcabba$ab	aabcbcabba$abb**c**
caabcbcabba$abb	abba$abbcaabcb**c**
aabcbcabba$abbc	abbcaabcbcabba**$**
abcbcabba$abbca	abcbcabba$abbc**a**
bcbcabba$abbcaa	ba$abbcaabcbca**b**
cbcabba$abbcaab	bba$abbcaabcbc**a**
bcabba$abbcaabc	bbcaabcbcabba$**a**
cabba$abbcaabcb	bcaabcbcabba$a**b**
abba$abbcaabcbc	bcabba$abbcaab**c**
bba$abbcaabcbca	bcbcabba$abbca**a**
ba$abbcaabcbcab	caabcbcabba$ab**b**
a$abbcaabcbcabb	cabba$abbcaabc**b**
$abbcaabcbcabba	cbcabba$abbcaa**b**

TABLE 2.2: Cyclic Rotations and Sorted Rotations of `abbcaabcbcabba$` with BWT (`abcc$abaabcabbb`) highlighted in bold.

Component	**Value**
First Column (F)	`$aaaaabbbbbbccc`
BWT	`abcc$abaabcabbb`
Count Array (C)	$C[a] = 1$, $C[b] = 6$, $C[c] = 12$, $C[\$] = 0$

TABLE 2.3: FM-index Structure for `abbcaabcbcabba$`. It stores BWT as well as count array (C), $C[p]$ counts lexicographically smaller letters than each character in BWT up to any given position (retrieved from the first column (F), see rotations in Table 2.2). Meaning, $C[\$] = 0$, because there is no character smaller than \$, $C[a] = 1$, because there is only 1 occurrence of 1 character smaller than a, and so on. FM-index also stores the occurrence array (occ) as displayed in Table 2.4.

Index	a	b	c
1	1	0	0
2	1	1	0
3	1	1	1
4	1	1	2
5	1	1	2
6	2	1	2
7	3	1	2
8	4	1	2
9	5	1	2
10	5	2	2
11	5	3	2
12	5	4	2
13	5	5	2
14	5	6	2
15	5	6	3

TABLE 2.4: The cumulative frequency of each character a up to position i in the BWT string T of initial string t, where T = abcc$abaabcabbb and t = abbcaabcbcabba$. Each entry $occ(c, i)$ in the table for a character c at position i is calculated as: $occ(c, i) = occ(c, i\text{-}1) + 1$, if the character at position i is c, otherwise $occ(c, i\text{-}1)$.

Index	BWT	BA (''a'')	BA (''b'')	BA (''c'')
1	`a`	1	0	0
2	`b`	0	1	0
3	`c`	0	0	1
4	`c`	0	0	1
5	`$`	0	0	0
6	`a`	1	0	0
7	`b`	0	1	0
8	`a`	1	0	0
9	`a`	1	0	0
10	`b`	0	1	0
11	`c`	0	0	1
12	`a`	1	0	0
13	`b`	0	1	0
14	`b`	0	1	0
15	`b`	0	1	0

TABLE 2.5: This table shows an example of bitarrays (BA) for characters `a`, `b`, and `c` in `abcc$abaabcabbb`. The method is an alternative approach to the occurrence array (*Occ*) (Table 2.4). It can be used to efficiently mark specific positions of characters and enable `rank/select` operations within the BWT.

2.4.2 Approximate String Matching

When discussing approximate string matching, which implies that our pattern query p may only partially match the string t due to variations in length, additional or fewer characters, among other factors, three primary considerations emerge. First, there are the *metrics*, which broadly quantify the differences between the two strings, describing how close or distant the compared strings are.

Following these, two broad types of algorithms for the approximate matching problem were gradually developed. First, the *exact solutions*, which provide a precise answer by computing the edit distance or the exact number of operations required to transform the pattern into a substring of the text (or vice versa). Conversely, *approximations* estimate the string rather than determining the exact measure. The latter, unlike exact solutions, are much less computationally demanding, as they can significantly reduce computational requirements, especially when dealing with extensive datasets. However, they compromise a degree of precision, rendering them potentially unsuitable in certain contexts.

Metrics

Richard Hamming introduced the so called *Hamming distance*, a "geometrical model" for the 2^n points in $\{0,1\}^n$ space. It is a metric by setting $D(p,t)$, both of the same length, to equal the number of coordinates, where $p_i \neq t_i$, effectively counting the substitutions required to change p into t.

As summarised by [21], Vladimir Levenshtein [174] expanded the concept to incorporate insertions and deletions in addition to substitutions. This was further developed to include transpositions or swapping, as introduced in [64], and is now known as the *Damerau-Levenshtein distance*. The *Levenshtein distance* (Figure 2.10) counts the minimum number of these operations needed to transform p into t, and in the Damreau-Levenshtein version, also considers counting reversals, which involve interchanging the order of two adjacent symbols.

The Levenshtein distance $\text{lev}_{a,b}(i,j)$ is computed as:

$$\text{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} \text{lev}_{a,b}(i-1,j) + 1, \\ \text{lev}_{a,b}(i,j-1) + 1, \\ \text{lev}_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

where:

- a: The first input string.
- b: The second input string.
- i: The current index in the first string a.
- j: The current index in the second string b.

Apart from Hamming and Levenshtein distance, there are other types of edit distance approaches, such as the *longest common subsequence (LCS)*, which also computes insertion and deletion, but does not include the substitution, the *Jaro distance*, which only considers transposition, and others.

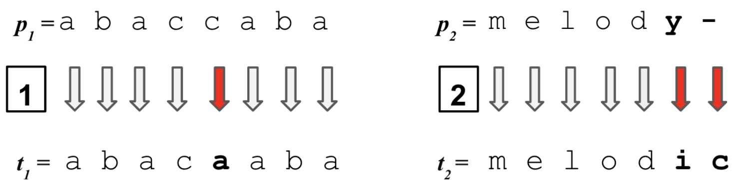


FIGURE 2.10: Two examples of the Levenshtein distance (or edit distance). Using the described method, we can compute the minimum number of deletions, insertions, or substitutions required to align p with t or to transform p into t. In the first example, only one substitution is required, resulting in a distance of 1. In the second example, one substitution and one insertion are necessary, yielding a distance of 2.

Exact Solutions

Expanding upon simple metrics of computing distances between two strings, other algorithms compute the cost or score differences between two text sequences and also consider their context[13]. This allows the scoring system to become more complex and adaptable to different scenarios, meaning that the cost of a substitution could be adjusted to be more "costly" than an insertion, if that is sensible for the specific case. These algorithms are typically implemented using dynamic programming and can accommodate both global and local variations or alignments.

The first of the two approaches is best represented through the *Needleman-Wunsch algorithm* [215]. It was designed for the alignment of the entire text from start to end. Similarly to distance metrics, it initialises the first row and column of a matrix based on gap penalties, and then fills in the rest of the matrix based on scores calculated from matches, mismatches, and gaps. The goal is to find the highest possible score, which represents the best way to align the entire text sequence (Figure 2.11).

Resembling Needleman-Wunsch, the *Smith-Waterman algorithm* [280], developed 11 years later, supports both global and local alignments. Specifically tailored for local alignment, it identifies regions of similarity within longer text, making it particularly effective for aligning the pattern or text wih multiple subsequences within the initial text. The matrix used is quite similar to Needleman-Wunsch, but, unlike its predecessor, it allows for scores to be reset or for negative scores to be zeroed. This feature enables alignments to start and end anywhere within the text, ultimately pinpointing the best local alignment (Figure 2.12).

Both algorithms are highly accurate for text sequence alignment tasks; however, due to their matrix-based approach, where the computation of each cell alone is computed in constant time, the overall time complexity of these algorithms is typically quadratic. Consequently, they are computationally expensive and impractical for large-scale applications unless modified. Alternative approaches are often used in such cases, but these typically involve other compromises, such as reduced "accuracy".

[13] In text sequence comparison, the cost/score quantifies differences based on operations like insertion, deletion, and substitution. Advanced algorithms may also consider the context of these differences and adjust the scoring system, accordingly.

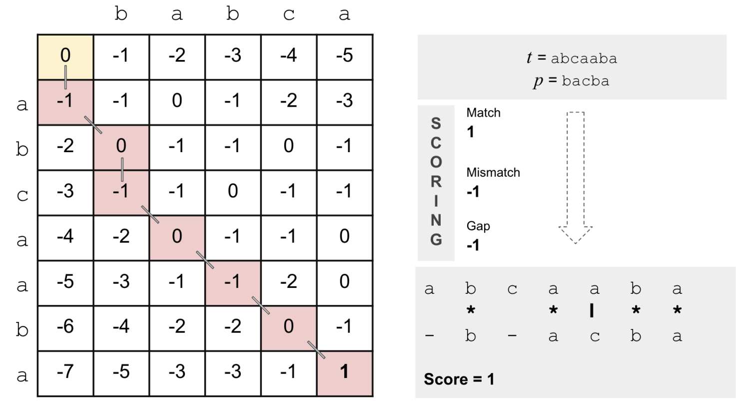


FIGURE 2.11: An example of the Needleman-Wunsch algorithm for global sequence alignment, using sequences `abcaaba` and `babca`. This method uses a scoring system instead of distance metrics (like in Levenshtein distance) to determine alignment. The matrix calculates alignment scores based on a scheme of +1 for matches and -1 for mismatches or gaps, initialized with sequential gap penalties. Scores are derived from the best values of neighboring cells (left, top, or diagonal top-left), with red diagonal lines indicating the optimal alignment path from the matrix's start to the bottom-right cell, where the total score (1) represents the best alignment.

Approximations

As previously noted, algorithms that provide exact solutions are often computationally intensive. Therefore, approximations are commonly employed in pattern matching tasks. Most widely used algorithms for this approach simplify the decision-making process by employing heuristics. Alternatively, some algorithms may begin with a potential solution and be iteratively refined. The particular methodology was not the primary focus of the pattern matching task in this thesis; thus, the description of these algorithms will be relatively brief.

One of the possible executions for such pattern matching is *BLAST (Basic Local Alignment Search Tool)* [6]. This tool performs text sequence alignment queries against both public and personal databases through web-based interface, application, or by using the BLAST code, of which many low-level functions can perform independently from the provided interface [187]. As reviewed in the cited handbook, BLAST uses heuristic algorithm to identify the local alignments. It details on each pairwise sequence alignment starting with the sequence identifier and a full definition line followed by the length of the sequence, and provides both the E-value[14], and bit score[15] with additional information on the number of identities, substitutions, and gaps. The alignment is visually laid out with the query sequence on top, the text sequence below, and markers such as dashes for gaps and pluses

[14] *E-Value:* The e-value quantifies the number of expected hits by chance when searching a database, serving as a measure of the probability that a particular sequence alignment is significant. Lower e-values indicate more significant alignments [187].

[15] *Bit Score:* The bit score measures the statistical significance of a sequence alignment, with higher scores indicating greater similarity between sequences. Scores below 50 are generally considered unreliable [187].

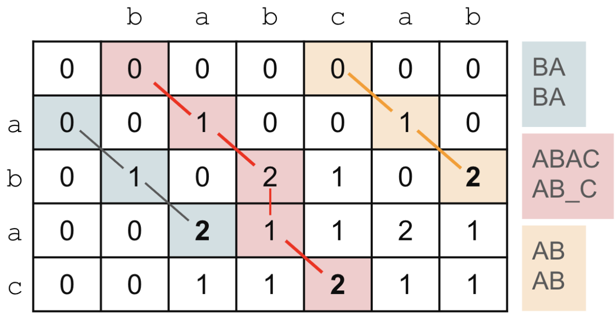


FIGURE 2.12: An example of Smith-Waterman algorithm for local sequence alignment, using sequences `babcab` and `abac` aligned vertically and horizontally, respectively. The matrix initialises with zeros and computes scores based on a pre-set scoring scheme of +1 for matches and -1 for mismatches or gaps. Each cell's score is derived from the maximum of three possible predecessors (left, top, or top-left diagonal), adjusted by the match or mismatch penalty. Among others, we display 3 possible alignments, all scoring in 2. The first two (blue and orange) display alignment of a short substring (`ba` or `ab`) without indels or mismatches. The third, red diagonal lines indicate the optimal path of the two substrings `abac` and `abc`, where one insertion or deletion, thus one indel, is required for alignment.

for substitutions, designed to facilitate human readability and quick analysis. As observed by [199], however, BLAST is no exception when it comes to trading off between speed and sensitivity.

A faster and more sensitive search tool for sequence alignment *PatternHunter* proposed by [186] uses a seed-based heuristics approach, where the seeds (or small instead of larger search strings) are optimally spaced between one another. Unlike BLAST, which typically uses contiguous (successive) k-mers (substrings of length k) as seeds, PatternHunter uses alternating k-mers. For example, it selects every 11th letter as part of a seed pattern. This alternative seeding strategy helps in skipping over certain positions, potentially offering a different sensitivity and specificity profile in detecting sequence similarities. It follows a two-step analysis process as well as has the ability to use multiple seeds simultaneously, both of which increase the algorithm's sensitivity. Similar solutions were proposed by *FASTA* [177], *BLAT* [131], and similar.

2.5 Pattern Discovery and Matching in MIR

Music pattern analysis in the field of MIR is extensively studied. The challenges of this topic extend beyond algorithms, encompassing diverse music forms, representations (signal, symbolic, or textual), extended music content, and cultural metadata. In music, as well as in other cases, a pattern usually describes an event—be it a melodic or rhythmic sequence, repeated words or phrases in lyrics, metadata, or other elements—that is repeated more than once. There is no single definition of a music pattern, as this heavily relies on what we understand as music, what we consider as events in that music, and what we recognise as repetition (see [52] for an extended overview of this matter).

When considering symbolic music representation, such as music notation, authors addressing the pattern matching problem generally rely on existing or adapted principles in music theory and music analysis [50, 5, 245, 42, 164] by focusing on music elements such as structure (phrases, parts, etc.), melody, harmony, and rhythm. It can consider them individually, in pairs or as multi-dimensional objects.

When working with these music materials, MIR is covering a large palette of tasks, which either relate to pattern discovery/inference, pattern matching or comparing sequences. All of these depend on how the music elements are represented. Thus, these processes begin by encoding the music data representations into *single-dimension strings of symbols*, such as *basic symbolic strings* or slightly more complex *indexed structures*, such as n-grams, suffix tries and trees, as well as suffix arrays and compressed structures, such as BWT with FM-index. On the other hand, there are *multi-dimensional object* that cover anything from tuples (pitch, time) and few-parameter schemes to larger models mentioned in (Section 2.3.1). These can be structured as a set, but also as geometry-based schemes.

2.5.1 Single Dimension Pattern Discovery

This category focuses on single-dimension pattern discovery and related tasks. It is challenging to clearly distinguish between different representations and approaches, as many studies, while concentrating on a specific task, often engage in supplementary research activities such as classification, encoding, annotation, supplementary analyses, and similar. Additionally, discussing pattern discovery or matching in isolation is not practical, as the objectives in MIR, such as archiving, comparing datasets, identifying patterns within a single dataset, or optimising algorithm efficiency, can vary significantly from one study to another. Nonetheless, based on the discussions in Sections 2.3 and 2.4 on music data representation and general pattern-related tasks, this section will be divided into three subcategories with intention to summarise the topic as systematically as possible.

First, *dynamic programming* for single-dimension string matching breaks the problem into simpler sub-problems, solving each once and storing the results. This method is effective for sequence alignment, using matrices to efficiently compute optimal alignments. Notable examples include the Needleman-Wunsch and Smith-Waterman algorithms, previously acknowledged in Section 2.4.2. As already mentioned in that section, while these methods prove useful for pattern matching tasks, they usually run in quadratic time, meaning they are not especially time-efficient.

In contrast, *indexing structures* like suffix trees, suffix arrays, and inverted indexes pre-process the text to create a data structure that allows for fast pattern searching and retrieval. These structures enable quick lookups by efficiently organising the data, making them ideal for large databases and repeated queries. These can be as time-efficient as linear time.

Additionally, *n-grams* and k-mers are often used in conjunction with indexing structures to break sequences into smaller parts, and potentially, if used in such (or similar) manner, further enhancing the speed and accuracy of pattern matching.

Dynamic Programming

Dynamic programming techniques, such as the Longest Common Subsequence (LCS) and edit distance algorithms, are often employed to efficiently compare and align these sequences, accommodating variations and transpositions. By leveraging these methods, researchers can uncover recurring motifs, thematic material, and structural similarities within musical compositions.

One of the most well-known and at the same time one of the early computational approach for sequence alignment of monophonic melodies in MIR was done by [207]. They used and extended the Levenshtein distance to accommodate music-related scenarios (Figure 2.13).

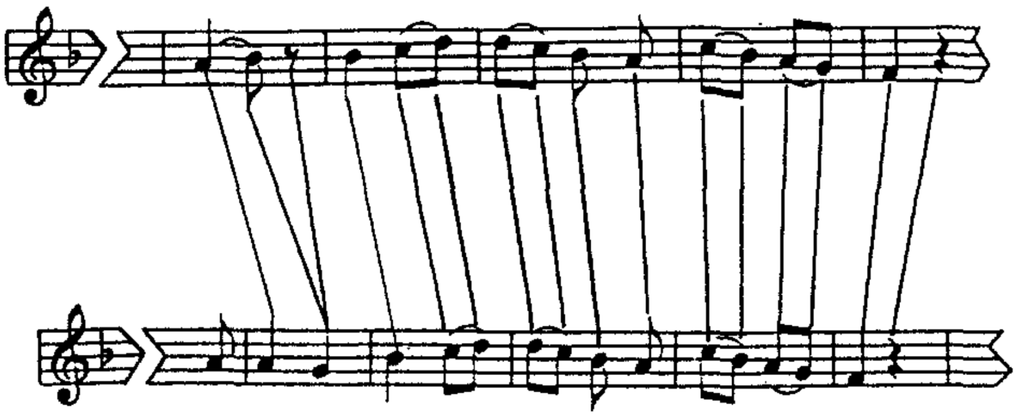


FIGURE 2.13: The Mongeau-Sankoff method [207] extends the concept of approximate pattern matching (Section 2.4.2) to melodic sequences by considering both, pitch and duration. Unlike, for example, the Levenshtein distance, it also includes operations such as consolidation (replacing several elements with a single one; for instance, the second and third unit in upper example to a single note unit in bottom example) and fragmentation (replacing a single element with multiple), in addition to the standard operations of deletion, insertion, and replacement.

Other examples are based on exact solutions for local and global alignment, such as Smith-Waterman and Needleman-Wunsch. Although originally utilised in text algorithmics for sequence alignment, they offer good robust methods for pattern matching tasks in MIR. I will highlight two examples. that utilised the Needleman-Wunsch algorithm. First, [261, 264], detailed in Figure 2.14, compared Japanese and English folk song melodies, while the second [305, 149, 304, 28, 123, 303], partially detailed in Figure 2.15, have conducted pattern marching on several studies on Dutch folk song corpus, focusing on the phenomenon of tune families. Authors of the latter explored the distance measure of melodic similarity through different variants of local and global alignments.

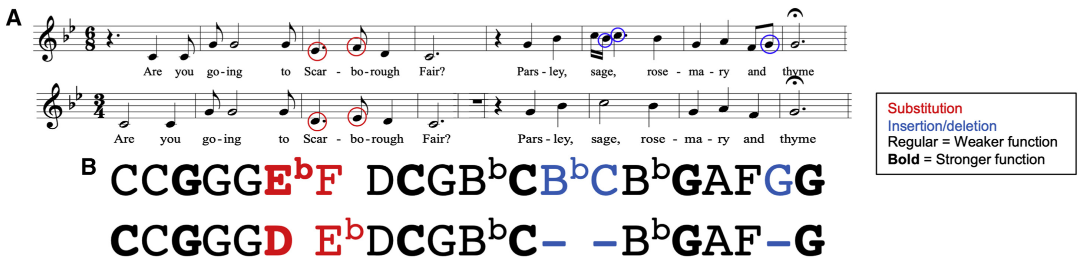


FIGURE 2.14: This paper [264] explored the similarities between Japanese and English folk songs. (A) Two english folk songs as scores. (B) The same two songs represented as a sequence of alphabetic characters with marked differences (red for substitution, e.g., a note in one example has to be substituted to match the first example, blue for insertion or deletion, and bold/regular, to indicate whether the note function in each song is stroger or weaker). Among other methods, they used Needleman-Wunsch algorithm, not only to pursue displayed score alignment, but to confirm that evolution of songs behaves very similarly across both folk song types. By observing the dynamics of insertions, deletions and substitutions, they indicated that the first tend to be more common than the latter in both cultures. The substitutions are, on the other hand, more likely to occur between neighboring notes.

Pattern Indexing

Indexing approaches are a relevant continuation of the methods described above, as they are able to involve more information, as well as allow for a more varied approach and faster searching.

Suffix Tries, Trees & Arrays. These structures exhibit improved time efficiency compared to non-indexed string configurations. By storing varying amounts of data, they eliminate the need for redundant information storage with each iteration. The sequential musical data is thus represented in a tree-like or comparable indexed structure, ensuring optimal data management and retrieval.

A very early example in this direction is [124], who executed approximate sequence-based pattern matching by employing suffix arrays to identify all transpositions, inversions, retrogrades[16], and retrograde inversions[17] of melodic segments across Palestrina's masses, allowing for the detection of varied forms of core melodies.

Burrows-Wheeler Transform (BWT) in Music Information Retrieval (MIR). As previously discussed in Section 2.4.1, the BWT [40, 192, 3, 255] paired with the FM index [278] enhances text compressibility and search efficiency for string-based pattern matching. As a methodological extension of the previous two approaches, this method is advantageous for rapid retrieval, particularly in the context of music streaming services and digital libraries. However, it has seen very limited application in music research.

[16] Jeppesen defines a retrograde as a musical sequence that is played in reverse order. This means taking a melodic line and playing the notes from the end back to the beginning, thus creating a mirror image of the original sequence in time.

[17] Jeppesen's retrograde inversions involve reversing the order of a musical sequence while also inverting the intervals. This means that not only are the notes played backward, but the direction of each interval (up or down) is also reversed. For example, if the original sequence goes up a major third, the retrograde inversion would go down a major third at the corresponding point in the sequence.

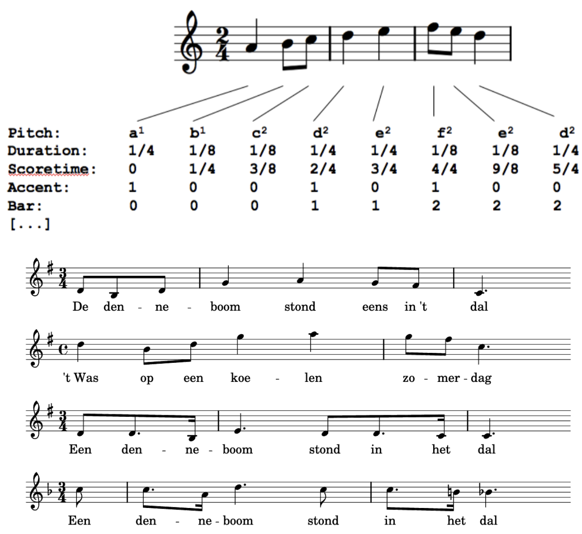


FIGURE 2.15: The proposal by [149] treats musical scores as sequences of multi-feature representations. In addition to pitch and duration, it considers features such as score time, accent, and bar. Their pattern discovery algorithm, based on an extended version of the Needleman-Wunsch algorithm (Section 2.4.2), applies single substitution scoring functions (focused on individual melodic features), combination functions (that can align multiple features simultaneously), and gap penalty functions. The pattern matching aims to identify tune families in Dutch folk songs, using previously manually identified "average" representations of each expected family. (Right) Image indicates the matches to the uppermost score to the 3 incipits below, in order of resemblance as sorted by the algorithm.

One music-centred study [46] utilised this for the hierarchical sorting of music by genres, composers, and similar categories. Further extending its application to music pattern analyses, [26] demonstrated how BWT and its inverse (iBWT) are useful in both analysing and generating rhythmic variations, thus proving invaluable for systematic organisation and creative composition in music. Additionally, the potential of the BWT for music analysis is recognised in [323]; however, that particular study does not directly apply BWT to music.

*N***-grams.** Another approach is to consider the *n*-grams, a rather common technique in computational musicology [208] that derives from the field of computational linguistics (similar to *k*-mers in bioinformatics and related fields). It has been widely used in statistical natural language processing, as well as in music research. In the latter, it usually proceeds to divide musical sequences into overlapping events, facilitating both exact and approximate string matching. For exact matching, these substrings enable segment-by-segment comparison, streamlining the search for identical sequences. In approximate matching, *n*-grams quantify similarity by counting the overlap of these substrings between sequences, allowing for a defined tolerance of variation. This method is used to handle specific variations in musical data, such as slight deviations in melody or rhythm [231, 327, 152, 326].

To detail one example, [300] uses the *n*-gram (or note) indexing and hashing by first splitting numerous music sequences into smaller *n*-note segments. This method allows for approximate matching, accommodating variations in musical sequences, which is essential for robust music retrieval in large databases. The use of hashing and indexing enables searches to be conducted in sub-linear time, making it particularly effective for searching through extensive music collections.

Another instance is the Melodic Signature Index (MSI) from the NEUMA project [56]. Their system supports both exact and approximate string-based searches using an algebraic signature-based index for efficient pattern matching across large music databases. It uniquely handles transposition and rhythm adjustments in exact searches, as well as employs a similarity function for approximate queries (see Figure 2.16).

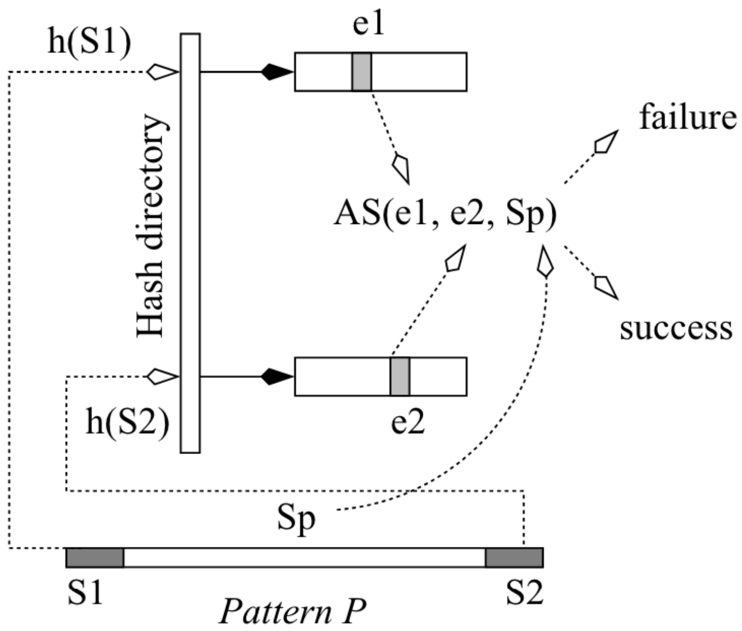
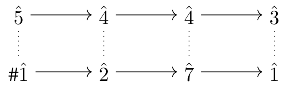
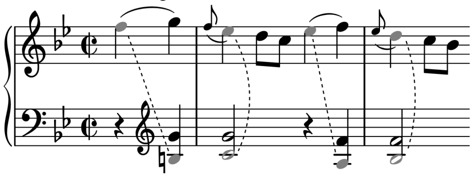


FIGURE 2.16: Figure 3 in [56] demonstrates an exact search method for matching sequences of pitch intervals akin to a given pattern p, whilst disregarding rhythm. The pattern p is preprocessed to generate three signatures: the initial n-gram s_1, the final n-gram s_2, and the intermediate section s_p that follows s_1. By hashing s_1 and s_2, one identifies records matching these signatures, and only considers pairs of records in the same voice and at the appropriate distance based on offsets. It should be noted that although this thesis typically uses p and t for pattern and text comparison, this figure retains the original annotations for easier reference to the original paper, thus substituting t with s.

Adapted to variation and changes in music, skipgrams allow for the inclusion of non-adjacent elements in sequences, capturing relationships between items that are not directly contiguous. This capability is particularly valuable in music analysis, where it enables the identification of patterns and structures in polyphonic music that involve non-simultaneous events. This extended version of n-grams, displayed and described in Figure 2.17, was adapted by [80]. Skipgram architecture was also proposed by [109]. The latter employed skipgram architecture of Word2vec to model polyphonic music by treating segments of Beethoven's piano sonatas as analogous to words in text.



(a) The "Fonte" schema as characterized by a typical outer-voice movement in scale degrees



(b) A realization of the Fonte in a piece

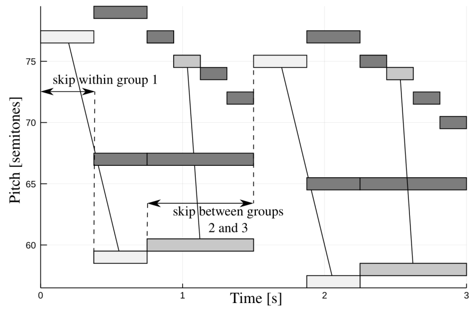


FIGURE 2.17: In [80], authors proposed a pattern discovery algorithm, extending the *n*-gram method by employing the so-called skipgrams. These constitute a method for identifying patterns within sequential data. The fundamental concept involves searching for recurring sequences of items, wherein the items within a sequence need not be directly adjacent. The skipgram algorithm operates by iterating through the input data one item at a time, maintaining a record of "prefixes," which are partial sequences that possess the potential to be extended into complete skipgrams. For each item encountered, the algorithm assesses whether it can be appended to any of the existing prefixes without exceeding a predetermined sequence length. If feasible, a new, elongated prefix is generated. Upon a prefix attaining the desired length, it is classified as a skipgram and incorporated into the output. (Left) A "voice-leading schema" (Figure 1 [80]) shown in two ways: (a) as a simplified representation using scale degrees, and (b) as it might appear in an actual piece of music. The notes in the schema do not all have to happen at the same time. (Right) Figure 2 [80] shows an example of how skipgrams can be applied to polyphonic music. The highlighted notes are a part of a skipgram, and the lines between them show the different "stages" of the pattern. The distance between notes is measured by the time between their onsets. This skipgram represents the same pattern as the schema in Figure 1.

2.5.2 Multi-dimensional Tasks

Multi-descriptor

A multi-parametric or multi-descriptor pattern matching employs a larger context of a certain melodic event by including more descriptors, such as contour, rhythmic values, onsets, cadences, as well as a number of metadata information. As it considers multiple different representations of music, the methods of encoding and pattern matching may be varied and/or combined (including the aforementioned string-based pattern matching, *n*-grams, and similar).

Some examples in music research are [167, 166, 165, 164] (Figure 2.18), all of which explored pattern discovery task through multi-parametric closed pattern and cyclic sequence mining with a one-pass approach to efficiently identify complex, repeating patterns, including heterogeneous patterns, in large datasets.

In a different series of studies, Darrel Conklin et al. explore the possibilities of describing music events by pitch, rhythm, and contour, as well as adding a number of metadata classifications. Based on his earlier work [54], each sequence in the corpus is stored to and searched within a suffix tree, they continue their research on a corpus of Cretan folk songs [53], adding information on the so-called (super)type and (super)area (for example, if *syria* is the area on a Cretan island, then *east* or *west*

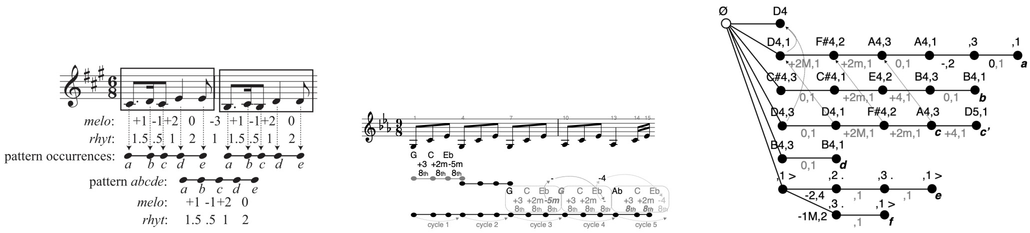


FIGURE 2.18: Cyclic pattern matching. The two figures illustrate three examples of cyclic representation and mining models by Olivier Lartillot, which consider repeating patterns as cycles. (**Left**) The Figure 1 from [167] (a similar example is displayed in Figure 4 of [166]) sets the initial pattern *ABCDE* as the basis for the cycles. As the music progresses, the pattern is subjected to variations and embellishments, yet, due to its multi-dimensional description, the underlying cyclic structure is able to persist. (**Center**) The Figure 11.9 from [164], the three-note pattern *GCEb* repeats twice, establishing a cycle. The algorithm recognises this and represents it as a single cyclic construct, simplifying the representation and avoiding the proliferation of redundant patterns. It also adapts to variation, as in cycle 4 the interval (-5m) is replaced by a more general descending contour (-), while in cycle 5, the algorithm adapts the model to include the new regularity of the pitch Ab and the descending perfect fifth (-4). (**Right**) A Figure 11.3 from [164] shows a trie-like pattern (Section 2.4.1) representation, where the white dot is the root, while the other nodes are shown in black dots. Each note description is beside its node, except interval descriptions, which are shown along the edges between nodes. Arrows point from more specific to less specific patterns.

part of the island would be super-areas), to further differentiate between rhythmic-melodic patterns (Figures 1 and 2 in [53]).

Another example is the utilisation of the algorithm MGDP (Maximally General Distinctive Pattern), which originated from Conklin's earlier study [50]. This method identifies musical parameters that are strongly associated with a particular genre, composer, region, or period, but are infrequent in other contexts.

While these examples were generally monophonic, Conklisn also pursued a multi-descriptor pattern matching on polyphonic music, especially counterpoint, by using the vertical viewpoint technique on a pre-defined feature set (Figure 2.19).

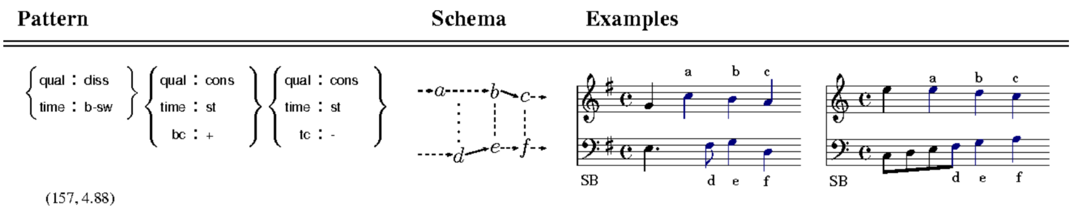


FIGURE 2.19: This paper [55] applies a vertical viewpoint (VVP) technique, which uses a feature set (left) to describe the bits of scores by using *qual*, (harmonic interval: `cons` (P1, m3, M3, P5, m6, M6) and `diss` (other intervals)), *time* (temporal relationship between two voices: `st` (starting together), `b-sw` (bottom voice starts), `t-sw` (top voice starts)), and `bc`, `tc` to describe melodic contour of either top or bottom voice: ++ (leap up) , - - (leap down), + (step up), - (step down), = (unison). The extracted patterns are then scored using an odds ratio that compares their probability in the corpus with their probability in a background model, to ensure the inclusion of distinctive rather than merely frequent patterns. For the pattern discovery task, this paper applied 2 methodologies, *antipattern* method (they created anticorpus out of contrasting pieces), and *null model* method, which estimated the probability of patterns based on its constituent features in the corpus. Their technique was proved to be rather efficient for exploring polyphonic music.

He also extended his approach to analyse Basque folk songs, discovering patterns within a specific “genre” by identifying anti-patterns (sequences less frequently present in that sub-corpus) [48, 218]. This approach derived from [11] and has been variably used to define patterns versus anti-patterns, or corpora versus anti-corpora, across different datasets. Examples include the aforementioned Basque folk songs, as well as Dutch folk songs studied by another research group [145].

Geometry-based pattern matching

A widely used approach for encoding data in multi-parametric pattern matching within music is through geometry-based representation. This technique represents musical features as geometric entities and utilizes distance metrics like Euclidean distance or cosine similarity. By organising musical data into geometric spaces, these methods facilitate tasks such as music similarity detection and content-based music retrieval, allowing for efficient comparison and retrieval. Furthermore, advanced techniques like nearest neighbor search and graph-based methods are particularly effective in identifying similar patterns within polyphonic music collections.

For example, [203, 202, 204] proposed a method wherein musical scores were transformed into multidimensional vector-based representations, treating each as points in Euclidean space. They employed SIA(TEC), COSIATEC (Compression-Oriented Suffix Array Interval Technique with Equivalence Class Transformation), and SIAMESE (Suffix Array Interval Matching with Equivalence class Extraction and Suffix Extensions) algorithms for approximate geometric pattern matching. These techniques facilitate the detection of both complete and partial matches within the dataset, with COSIATEC offering a more compact representation of significant patterns.

In another example, [183] used a geometric representation of onset and pitch of melodic segments (as MIDI). They pursued approximate pattern matching by weight matrix W based on the score and the given weight function, which measures deviations of the translated pattern from the score.

Next, [294] applied a stream segregation algorithm to a Johann Sebastian Bach corpus to identify streams in music compositions represented as MIDI files. These files are converted into musical graphs where vertices represent music notes, and edges indicate melodic and harmonic relationships between the notes.

Then, [77] worked with a subset of scores from the Essen folk melody collection [265], represented in multiple dimensions, including pitch, interval, content, and contour. They performed (predominantly geometric) approximate pattern matching using MDS (multidimensional scaling), which compared these multiple dimensions, including pre-processed normalized melodic phrases and extracted contour vectors (using equidistant sampling points).

Last but not least, [171] analysed polyphonic classical music, represented as a set of points in an Euclidean plane (onset, pitch). They used exact and approximate pattern matching of those points by geometric hashing and vector-based indexing.

2.6 Conclusion

This chapter aimed to summarise the majority of studies related to the topic of this thesis. It began with an overview of ethnomusicological methods, leading to more specific issues related to digital archives that support the thesis's data collection, the digitisation of materials along with their annotations, and the construction of the analytical model.

The second part focused on supporting the computational methodology, highlighting general contributions in the field of text algorithmics as a basis for the chosen methodology on music pattern discovery and related tasks.

The issues discussed in the first part will be further developed in the next Chapter 3, where I will introduce the dataset of Slovenian folk song ballads, providing primary explained analyses results and statistical information. These topics will also be partially addressed in the last Chapter 6, offering a critical discussion on the integration of ethnomusicological studies within MIR, and vice versa.

The pattern-matching contributions, on the other hand, will primarily support Chapters 4 and 5, where the main focus will be on introducing and evaluating the developed algorithmic methodology for pattern matching tasks. This methodology combines melodic sequences with descriptor queries, aiming to integrate knowledge on the organisation of materials and, more importantly, to combine single-parameter string algorithms with multiparametric pattern matching.

Chapter 3

Dataset: Introducing Slovenian Folk Song Ballads

Statement: This chapter includes an expanded contribution from my published paper at ISMIR 2023 [32] and a paper submitted to Ethnomusicology Forum [33].

This chapter provides a gradual exploration of the Slovenian Folk Song Ballads dataset. It begins with a definition of key terminologies, including *folksong*, *ballad*, and *Slovenian*, to establish a foundational understanding essential for the usage and analysis of the data.

The text proceeds with a comprehensive introduction to the corpus. I first collect the palette of thoughts that have been made on these songs to explain their historical context. Then, I detail the latest contribution we [33] made to the current, now digitised dataset's organisation along with a number of new annotations. This is followed by statistical analysis of the dataset, offering insights into various patterns and trends present in the established digital collection.

Additionally, a subsequent part is devoted to a commentary on music sequence analysis, addressing potential misinterpretations of the corpus with supporting socio-historical context on how these were transcribed and curated.

Finally, the chapter delves into a thorough overview of the dataset's integration into the Dezrann platform, outlining its organisational structure and accessibility for further scholarly research and analysis. It also describes the scripts used to integrate the corpus and its annotations to the platform.

3.1 Terminology

3.1.1 What is a (Folk) Song?

There are many different perspectives on what constitutes a folk song. Generally, it is interpreted as an orally-transmitted song, composed of a melodic line and vocals with lyrics, whose primary function revolves around human events. The definition of a folk song has undergone a profound transformation, shaped by the scholarly discourse that has evolved over the decades.

Cecil Sharp's collections regarded folk songs as valuable representations of national heritage, emphasising their preservation due to their perceived role in expressing the "authentic" national spirit [274]. However, Béla Bartók, through his detailed study in Hungarian Folk Music [13], challenged this somewhat romanticised perspective, presenting folk songs as complex and sophisticated structures that not only inspired contemporary music but also demonstrated their merit for scholarly attention, surpassing their value as mere cultural artefact.

In 1960s, Bruno Nettl broadened the scope, as he underscored their role in the social processes and the ongoing cultural transmission, moving away from the notion of static musical pieces to vibrant elements of communal interaction [216].

Still in the 60s, the aforementioned Lomax [180] advanced the notion of folk songs as deep-seated expressions of cultural identity, offering a novel lens through his *cantometrics* methodology, which correlates song styles with cultural traits. This marked a shift from viewing folk songs merely as musical expressions to recognising them as intricate narratives intertwined with cultural practices.

Later on, Ruth Finnegan [83], further integrated the study of folk songs as the dynamic field of verbal arts, portraying them as living components of oral traditions that actively shape and are shaped by the narratives of community life.

Lastly, David Atkinson and Steve Roud [9], critiqued earlier conceptions and presented folk songs as continually evolving entities. Their research highlighted how these songs are not merely relics but are actively influenced by changing social contexts and media interactions, thereby enriching our understanding of their adaptability and enduring relevance. Similar thought was later also adopted by Rice [249].

These and other scholarly thoughts have collectively transformed the understanding of folk songs from static historical artefacts to viewing them as dynamic cultural processes that both reflect and influence, but most importantly, are embedded in the social practices.

Contrary to the common belief that a folk song emanates from communal creation, it is often the work of an individual who crafts it spontaneously, adopted and repeatedly adapted by the community [155]. The content of the songs is made up of everyday speech and familiar folk song formats, blending reality with imaginative ideas on certain subjects, habits, figures, and similar. The portrayal of historical events in folk songs, hence not a direct reflection but rather interpreted through the lens of the people, adhering to the stylistic and expressive nuances of folk speech.

Melodies support the structure of lyrics, generally serving more as a carrier for the words than as a musical illustration of the text. The alignment of the mood of the lyrics with the melody is not typically a concern; it is regular for the melody or lyrics to be interchangeable among different songs as long as they fit the metric and rhythmic frameworks. And even then, minor adaptations are often made to accommodate such (ex)changes [155].

A significant distinction highlighted between "art" and "folk" songs is that the melodies of the latter are not commonly transmitted through music notation, while the lyrics may have existed in printed form. The ownership—and indeed, copyright—of these songs does not concern the creators [155]. This does not imply the absence of an "original author", but rather, due to the songs being adopted and transformed by the people, that the concept of authorship and rights are not prioritised and practiced within these types of music things, at least not in the same sense as this is practiced in the frames of artist (or composer) centered music (for example, refer to [270]).

The transmission of folk songs across generations fosters a traditional yet non-mandatory character. The life of a folk song lies between enduring tradition and evolving genres, forms, and melodies. This dynamic is driven by the community's creative engagement, both in preserving accepted forms and in innovating new ones [155].

Folk songs in Slovenia were formally classified as such after the Second World War. Previously, they were commonly referred to as "national songs" or "songs of the nation" (Slovenian: *narodna pesem*). The national emphasis gradually gave way to the notion of the "folk" or the "people". However, both terms, "national" and "folk", remain in use to this day.

In our case, I primarily refer to the type of folk song commonly recognised as a ballad or narrative song, with our main reference being [93]. However, it is important to note that much of what I describe before and in the following sections also applies to a broader range of folk songs, as well as the practices involved in their collection and preservation. Additionally, numerous folk song collections from different Slovenian regions have been compiled simultaneously, with ballads being just one among many. To elaborate further on this specific type of folk song, I will explain the terminology and outline the most common features that define a folk song as a ballad.

3.1.2 What is a Folk Song Ballad?

A folk song ballad[1], also referred to as a narrative folk song, is one of the forms of traditional folk music that tells a story through its lyrics. Originating in Europe, folk ballads have persisted through the centuries, maintaining their relevance in communities, where oral tradition remains strong [319]. There is a range of diversity in form and content, which reflects the ballads' evolution over time as being transmitted verbally across generations, leading to numerous variations and adaptations [319].

The role of a ballad is conveying cultural values and historical events, serving as a means of preserving and communicating communal histories, societal norms, and collective experiences. Apart from that, they also recount legends, fairy tales,

[1] For a deeper study on this topic, see also [252].

family faiths, and anecdotes, showcasing a diverse range of subjects from romance to tragedy and heroism [319, 87]. The form is usually concise, focusing on a pivotal moment of action or conflict. Characters are sketched with minimal detail, with their motivations and backgrounds often implied rather than explicitly described. This brevity and directness contribute to the dramatic impact of the ballad, often enhanced by incremental repetition and, in slightly newer examples, refrains [87].

The structure of a ballad is typically strophic, meaning it is divided into verses or stanzas that follow a consistent, repetitive rhyme and meter scheme. This form facilitates recollection and oral transmission, which was especially crucial in cultures with limited literacy or whose processes mainly relied on non-textual communication [87]. However, ballads have evolved over time, adapting to the social and cultural contexts of different regions. For instance, while British and American ballads are typically rhymed and strophic, other variations such as the Russian *byliny* and Balkan ballads are unrhymed and unstrophic.

In 1966, Slovenian researchers officially adopted a common definition that "a ballad is a song that tells a dramatically emphasised story [153]". Additionally, Golež Kaučič summarised various definitions and genre classifications in international folklore studies, emphasising that "a ballad is defined through genre and tradition" [130], as well as through previously mentioned oral transmission. Furthermore, as found by the cited authors above, ballads dramatically recount the story of an event and its outcome. They are embedded in the process of variant creation, maintaining a dynamic relationship with the context, fulfilling specific functions within individual communities [130].

Donald Knight Wilgus [319] introduced a systematic method for categorising traditional narrative songs, particularly ballads. This type-indexing approach classifies ballads based on their narrative content, themes, and structural descriptors. It considers common themes such as love and betrayal, and formal elements like rhyme schemes and stanza forms. This indexing method not only organised ballads for easier study and comparison but also situated them within their cultural and historical contexts, providing a comprehensive tool for researchers in folklore, ethnology and musicology. A similar approach was taken for the collection and curation of Slovenian folk song ballads related to this thesis [93].

3.1.3 What is "Slovenian"?

Here, I will be concise. Instead of delving into a historical discourse on the shifting borders of Slovenian territories, I will focus on defining what is considered Slovenian within the context of the collection this thesis incorporated. This approach ensures clarity in understanding the cultural and geographical parameters that frame the collection's contents, but does not provide the complexities of extensive historical detail.

The perception of what should be collected as a *Slovenian* folk song, including the folk song ballads in question, changed over time. Slovenia declared its independence in 1991, and has territory-wise belonged to Habsburg monarchy, The Kingdom of Yugoslavia, FPR Yugoslavia (Federal People's Republic of Yugoslavia), and SFR Yugoslavia (Socialist Federal Republic of Yugoslavia), to name the most recent. It was geographically, politically and otherwise structurally reshaped several times, and with that, the thought of what is deemed as Slovenian, changed as well.

In the context of the collection, the classification labelled as "Slovenian" was fundamentally based on the Slovenian ethnic territory, eschewing contemporary national borders. It exclusively included songs in the Slovenian language (encompassing dialects), and separately categorised variants found within Slovenian territory that were sung in Serbian or Croatian [93].

The aggregation of songs was meticulously organised and analysed within the framework of regions, such as Upper Carniola, Inner Carniola, Lower Carniola, Styria, and Prekmurje. Predominantly recorded in rural settlements, these songs and their variants capture a snapshot of cultural evolution from the 19th century to today—some regions have experienced significant urbanisation, while others have markedly stagnated and depopulated [93] [2] (see an example of displaying appearances in different regions for a song type *Nevesta detomorilka* / *Infanticide bride* in Figure 3.1).

3.2 Corpus introduction

Our collection features 402 Slovenian folk song ballads, sourced from the archives of the Ethnomusicology Institute ZRC SAZU, compiled over various time periods (Section 3.2.1). For the purpose of this thesis, we are releasing digitised scores, lyrics, annotations, and available recordings. Currently, only the first verse of each ballad is included in the digitised collection, although it is important to emphasise that these ballads traditionally contain multiple verses[3]. Each melody-lyric pair has already been categorised into thematic groups based on the lyrics [93], and further enriched with detailed metadata and annotations.

One of the key contributions of this project has been curating and modelling both the previously available and newly created data. This involved not only combining information from various sources, but also re-organising and, as the new information keep coming to light, re-adapting a significant amount of annotations to fit the new framework. Additionally, we developed a robust system that allowed us to add new annotations to the existing data. Annotations are structured into two main categories: the first describes the song as a whole (focused on the entirety of the first verse), while the second details individual melodic phrases within each song.

[2] A comprehensive geographical explanation of the regional annotations with regional and historically-informed maps can be found in the cited source.

[3] The full lyrics (all that an individual or a group of singers performed) are available in the physical archives of the Ethnomusicology Institute ZRC SAZU and in [93].

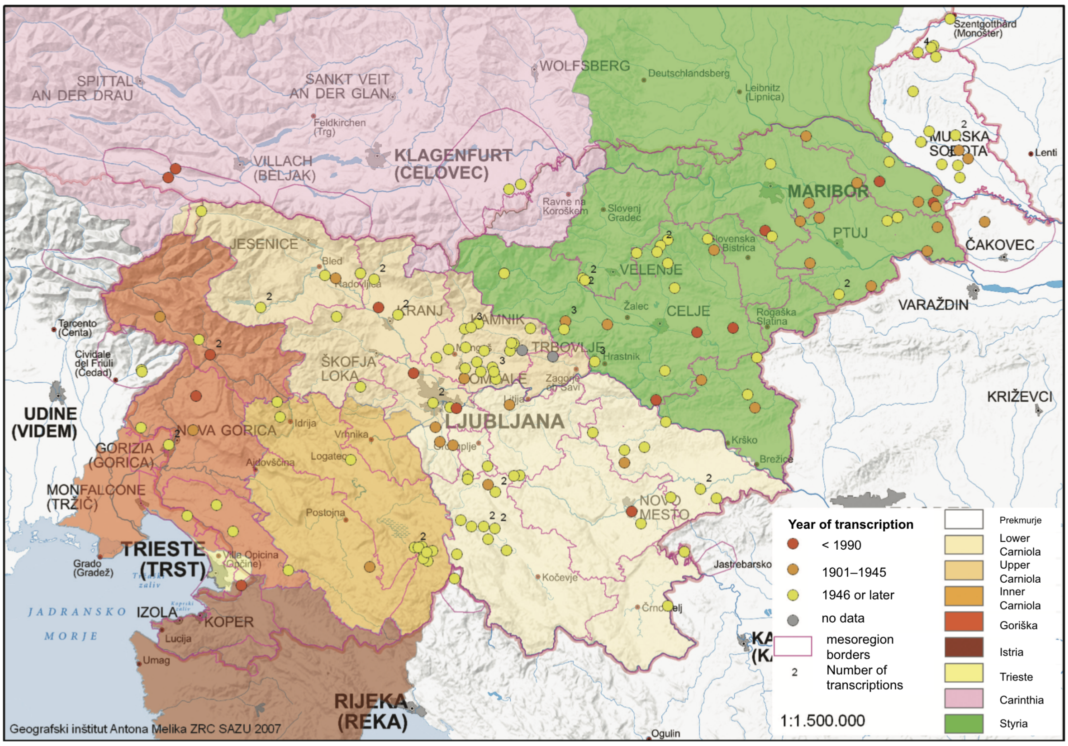


FIGURE 3.1: The figure displays an example from *Slovenske ljudske pesmi 5* [129]. It indicates the appearance spots of the song type *Nevesta detomorilka / Infanticide Bride* in different Slovenian regions with approximate dates. The legend was translated. The original can be found in the cited source.

This section will introduced how the dataset is organised and what kinds of annotations it entails, which will be the basis for the provided the statistical information. But first, it will focus on introducing the collection (and conservation) practices, that had an (in)direct influence on the construction of the observed corpus, its digitisation process and analysis.

3.2.1 Collection practices in Slovenia

The materials from the utilised dataset in this thesis, akin to those delineated earlier, exhibit considerable diversity in terms of their provenance, content, context, and other pertinent variables. The sketches, notes and transcriptions, along with subsequent recordings, span from the early 19th to the late 20th century. Throughout this time period, discernible shifts occurred in the methodologies and technologies employed for data acquisition, inevitably impacting the present state of these materials. Despite undergoing multiple instances of aggregation within similar repositories, all of these materials did not inherently possess substantial commonalities from their inception.

The Pre-Recordings Era

Chronologically, the collection comprises songs from a period before sound recording devices, mainly, only lyrics manuscripts from various smaller collections and notes of ethnographers, linguists, travelers, and similar (for example, Emil Korytko (1813–1839), Oroslav Caf (1814–1874), Anton Breznik (1881–1944), Matija Majar Ziljski (1809–1892), ...).

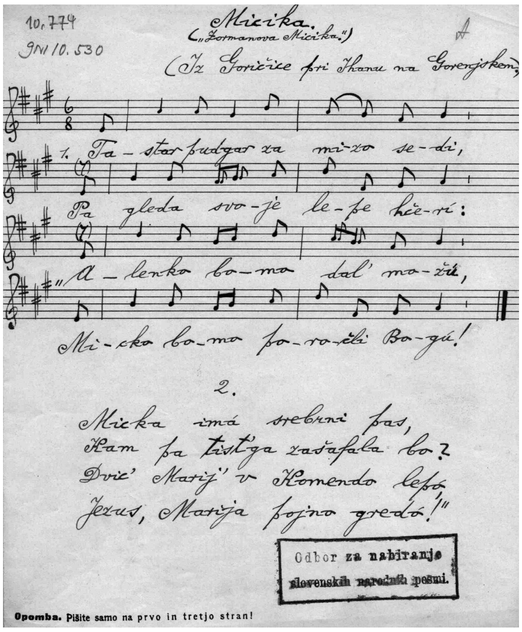
However, many of these were included in a monumental collection by Karel Štrekelj (1859-1912), and later Joža Glonar (1885-1946), who in 1895-1923 edited numerous volumes of Slovenian folk songs titled *Slovenske narodne pesmi*. Glonar was the first to sort them into thematic sections, where each song was assigned a number, and first provided rules for lyrics transcriptions.

Štrekelj and Glonar opted for a simplified dialectal transcription corresponding to the abilities of the transcribers, and the general public. The basic dialectal system is in use until this day. Early collectors, lacking transcription skills, or due to the collection manners and socio-political goals of the time, did not pay great attention to melodic transcriptions, which were, hence, rarely considered. Glonar, with a more detailed definition of a folk song, completed the collection process by publishing a total of 8686 songs and 1944 units of "supplement material", adding introductory verses or stanzas, and what was not considered an authentic material beforehand. Each song included the recording location in parentheses, while other information was in footnotes (recorder's name, location, sometimes date and time, source, textual variations from unpublished versions). Despite several uncertainties (what qualifies as a folk song, missing or inconsistent metadata), he encouraged forming systematic collection regulations.

Earliest transcriptions of melodies deemed to be uncertain as well [150]. Those of Franjo Kuhač (1834-1911), are mainly taken from the editions published in his collections for voice and piano accompaniment (Kuhač 1878-1881), while Stanko Vraz (1810-1851) only annotated sketches of the melodies. The few included in our collection, due to their unreliability (i.e., not matching the notations of the texts metrically, rhythmically, and being melodically unusually transcribed), share little with the rest of the materials and thus cannot be properly compared.

In 1906, when the Austrian government financed a large project of collecting all folk songs (including melodies) of all Austrian lands, *The Committee for the Collection of Slovene National Songs with Their Melodies* was established. The collectors, largely amateurs, were summoned by a public call (OSNP 1907). They were given a set of strict regulations [330], where it was especially stressed that the songs collected should be transcribed as sung originally and not to be freely harmonised, nevertheless, a practice many collectors exercised (Franc Kramar and Ciril Pregelj [94]). It was not allowed to interfere with the lyric, song structure, and not to transcribe the tunes that resembled the ones which were previously collected. The latter, seeming sensible at the time, unfortunately, disabled the tracing minor changes in song evolution. This project was supposed to be based on a systematic foundation, but the inability to record the fieldwork, as well as the loose relationship of collectors towards the regulations, made transcriptions unreliable. The editors of the collection had no supplementary material for re-evaluation of the, commonly, intuitively collected data. Moreover, today's user has to rely on both.

As the project concerned many different nations under a single Habsburg monarchy, it involved many socio-cultural, political, and methodological compromises. These included establishing what was to be collected, considering song popularity ("What constitutes a folk song?"), genre (ballads, work songs, lullabies, folk theatre, ...), "nationality" or the language of the material (German, Austrian, Slovenian, ...), and song sorting (by lyrics, dialects, regions, ...) [213]. Many melodic transcriptions by these transcribers seem simplified and/or inaccurately transcribed. For instance, collectors mostly transcribed individual singers, not groups, leaving out otherwise very rich polyphonic singing tradition [142]. It was not only the lack of skills but conversely, also the thorough Western music education that interfered with the transcription processes, favoring aesthetically more pleasing "Western-sounding" tune harmonisations over the folk practices, hence such transcriptions cannot be considered as the "exact" reflection of the folk song of the time [154, 142]. Numerous publications of folk songs from the OSNP campaign underwent a series of redactions, but the folk family ballads generally remained unedited (one example of the original transcription from 1910 by Franc Kramar can be observed in Figure 3.2).



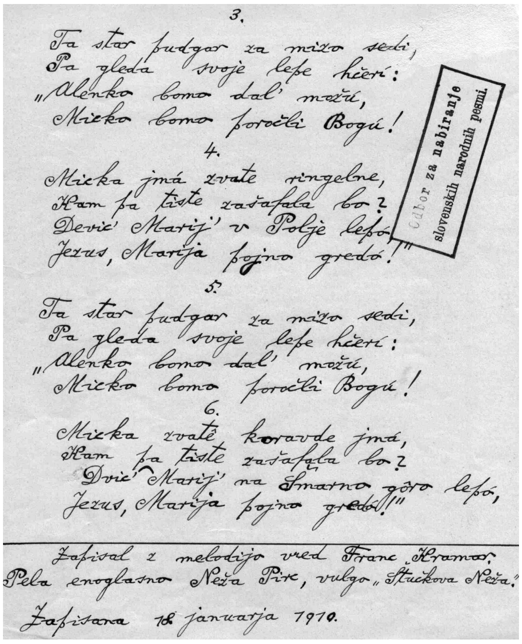


FIGURE 3.2: Transcription notes of the song *Ta star purgar za mizo sedi* (titled as *"Micika"* in the notes), as transcribed by Franc Kramar. The song can be found in SLP 5 [93] as song type 237 (*Oče določa usodo hčere / A father determines his daughter's fate*) variant 3 (refer to the example on Dezrann). This transcription, dated January 18th, 1910, includes musical notation and the first, second, fourth, fifth, and sixth verses of the lyrics. The stamp on both pages indicates that the transcription was prepared for the OSNP collection of folk songs, of which the Slovenian contribution was curated by the *Committee for the Collection of Slovene National Songs with Their Melodies* (Section 3.2.1).

Phonograph Recordings Era

By the 1920s, folk song researchers aimed towards collecting both, notated melodies and recordings, however, the road towards these practices was long. It was Štrekelj, who "planned extensive and systematic recordings of folk songs and aimed to create an archival sound collection of recordings on wax cylinders". He prepared detailed instructions for recording archiving and usage [160]. During that period, extensive debates arose whether recordings should supplement or replace manual symbolic transcriptions. These discussions primarily centred on preserving the "authentic" folk song form [330], but also on resolving the accuracy of polyphonic annotations and other transcription issues in music, for which the Western notation may be insufficient [160]. Initially, melodies or lyrics emerged independently, and subsequently, as the research field evolved, these elements converged with additional metadata information [213]. With technological progress, sound recording soon became feasible. However, the committee responsible for purchasing such devices initially failed to understand the use of recordings beyond a "substitute" for lacking transcription skills. Consequently, Štrekelj was not successful in convincing the Slovenian committees to purchase audio equipment in his lifetime, resulting in his work being confined to symbolic music transcriptions [161, 160].

In 1914, Matija Murko (1861–1952) managed to acquire the first phonograph, which Juro Adlešič (1884–1968) first used in that spring to record folk songs from White Carniola [160]. The appearance of the phonograph brought about "a new methodological approach to documenting and studying folk songs" [160]. From then on, researchers did no longer have to complete their transcribing in real-time, and informants did not need to repeat the song, as recordings allowed playback, enabling better transcriptions based on a single performance rather than approximate live repetitions.

Researchers could shift their focus to performance styles, voice characteristics, precise tempo, and non-tempered melodies, and even emotions and expressions. This progress sparked new debates, like those of Stanko Vurnik (1898–1932), Valens Vodušek (1912–1989), Zmaga Kumer (1924–2008), and Julijan Strajnar (1936) [142]. Through re-listening, re-recording, converting phonograph cylinders, and using new -er tape recording devices after 1995 [159], they gained new insights of the collected folk songs.

Nevertheless, aesthetic, moral, and political influences of that time continued to shape decision-making processes, determining which songs should be collected as "Slovenian". If they were deemed not "authentic" enough, the songs were excluded from the collection and further research. It was the fusion of the concept of "authenticity" and cultural nationalism [19] that was present in the methodology of collecting folk songs in the Slovenian space.

The developing recording media, however, had a great control over the quantity, length and quality of information, and the amount of (acoustic) data (e.g., less capable medium forced one to favour shorter over longer examples, to be selective of what to record, whether to interrupt the performance), and to cut out conversations with the singers. The more mundane the media and the evolved the fieldwork practice (see some photographs of fieldwork in Slovenia in Figure 3.3), the more subject‘s attitude towards the presence of technology and researchers changed, which influenced the performance as well.

With persistent fieldwork and methodology evolution, researchers strongly advocated for a comparative approach. Vodušek streamed towards incorporating elements like context, phonetics, and lyrics. Following the global trends [205, 179], Vodušek pushed beyond material preservation practices for two reasons. First, he considered music a non-static phenomenon, and second, he claimed that to define Slovenian folk song, one must compare it against presumably non-Slovenian material [308]. At the time, several "musical" issues persisted with the recordings [178], but on a larger scale, similar thinking triggered changes in material collection, conservation, and methodology.

FIGURE 3.3: Two different photographs of Marko Terseglav's (1947-) fieldwork from (**Up**) 1975 and (**Down**) 1987 (Source: Etnofon).

Digital era

Recording processes eventually developed, not only in the scope of field recording, but also in terms of transcribing, reproducibility and accessibility. Gramophone records, then CDs, later digital formats such as .mp3, .wav, digital archives, YouTube, Spotify, and other audio-visual platforms, provoked the re-thinking of the conservation practices and research, in general. The need for the digital required the physical materials to be consistently (re)organised. Hence, Drago Kunej and Rebeka Kunej systematised the protocols for collecting (recording and archiving) [162], analysing and digitising music material [289].

There have been several projects that have made folk songs digitally available, such as radio broadcasts (RTV - internal archive, podcast *Slovenska zemlja v pesmi in besedi*), a YouTube channel of the Ethnomusicology Institute ZRC SAZU and other institutes, as well as scientific projects, such as Etnofletno, Klik v domovino, Folk Music Heritage, CD and book editions, and the latest ongoing project Etnofon (2023), a digital collection of all the institute's past and present sound publications. Some of these recently digitised materials have also been computationally annotated and analysed [291, 32]. Following the latest physical edition of these songs [93], [291] focused on analysing the conceptual structure and themes within the lyrics, for which the musical contents were first digitised and computationally processed. More recently, [32], as part of this thesis, concentrated on further digitally curating the collection and making it accessible to a wider audience. In addition to digitising the pre-existing musical materials, this work incorporated metadata and musical annotations of the melodic content, encompassing the songs and their corresponding melodic phrases, and developed an analytical model for pattern discovery within these materials.

3.2.2 Current Dataset Organisation and Annotations

We are introducing the new and the first digitally accessible version of the 402 folk song ballads from Slovenian regions, previously collected, curated and digitised by several members of the Ethnomusicology Institute ZRC SAZU and their external colleagues. In this section, we will discuss on the dataset's structure is divided into metadata on the entire corpus, melodic song phrases and their corresponding descriptors. The descriptors are divided into non-music descriptors (metadata), lyrics (verse structure) and music descriptors.

The structuring of the *non-music* descriptors was adopted from previously mentioned collections and archives that gathered, organised and curated the sources. Some of these non-music descriptors or metadata were adapted to fit the computational system, as well as some, where sensible, were also translated to English (more precisely the names of regions, type title, and the first line of the first verse).

The *lyric* descriptors were also partially borrowed from the systems behind the Slovenian folk song collection and were further curated by current members of the

Ethnomusicology Institute ZRC SAZU. These include the first verse, as well as metric and rhyme verse structure.

Lastly, there are *music* descriptors. These consider several high-level elements of music, such as contour, phrase, "scale", and time signature, among others. Below, I will introduce one by one, and then, some statistical data on introduced descriptors with preliminary analyses.

Metadata

The songs in our dataset have been previously organised into 36 distinct topic types, each having from one to as many as 103 variants. They were transcribed or recorded in 22 different regions, with the majority in Styria, Upper Carniola, and Lower Carniola (Table 3.2). The transcriptions and/or recordings with identifiable years of origin span across 68 different years. The earliest year traces back to 1819, while the most recent transcription was done in 1995 (Table 3.2). Each type was assigned a topic title, but the context of the individual song was further determined by the beginning of the first verse of each song.

The majority of songs in the dataset were performed by solo female singers (Table 3.2 and Figure 3.4), followed by solo male singers. Klobčar [136], for instance, investigates the prominence of female singers as the primary bearers of folk songs, especially the ballads, proposing that this shift occurs after the songs lose their original social significance and men lose interest. That same study also draws the connection between women singers and these songs within the context of identity-building during the national process.

Given that our collection primarily consists of monophonic tunes, the representation of group singing is considerably lower. While other collections of Slovenian folk song ballads do include polyphonic (or more commonly homophonic) arrangements, the majority of recorded or transcribed ballad songs are monophonic, even though many of these were originally sung homophonically. Therefore, the prevalence of monophonic tunes is not necessarily indicative of their performance style in spontaneous settings but rather a consequence of the collection methodology (more detail on this is provided in 3.2.1)

Nonetheless, the thesis focuses on monophonic melodies. Consequently, the analysis of musical descriptors presented here will be based on the information gathered from monophonic material, as this, among other, facilitates a more consistent computerised comparison.

Music Descriptors

Music descriptors (Table 3.4) can pertain to either the content of the music or represent the music content itself. Below, I provide explanations for each descriptor separately. These were primarily extracted computationally and refined through manual editing and redaction.

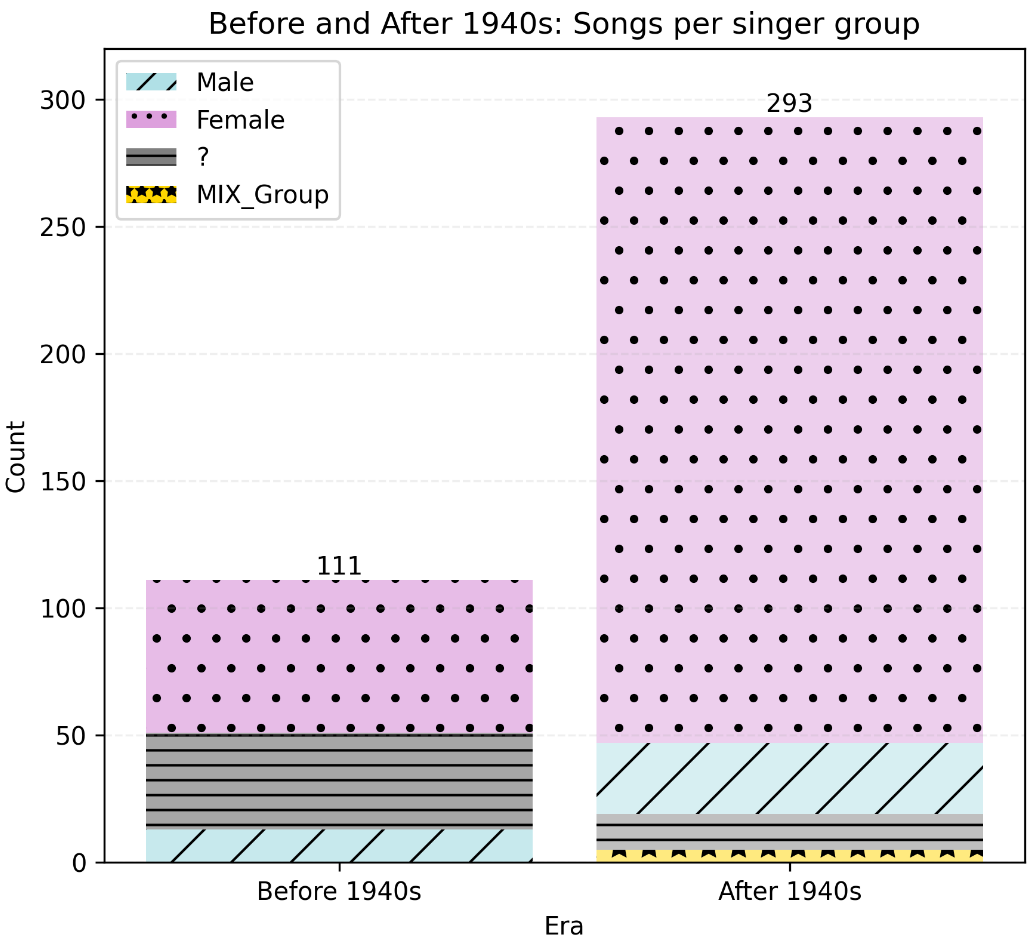


FIGURE 3.4: The count of male (solo/group), female (solo/group), mixed group and unknown singers in the dataset of *Slovenian folk song ballads* the eras before and after 1940s.

Time signature(s). These are noted in the order of their first appearance in whole songs. For individual phrases, the relevant time signature for each phrase is labelled independently, regardless of the song's overall time signatures.

Upbeat. This data specifies the number of quarter notes in an upbeat measure at the start of a tune (for example, 1.0 = 1 quarter note).

A number of music note units. This represents the total count of note units in a score, regardless of their duration. For example, both a quarter note and an eighth note are counted as 1 unit each.

The number of measures. This descriptor indicates the number of all measures in a score representation of individual songs, including upbeats.

Range. The entire song's range is expressed in two ways: first, as a melodic interval (for instance, m7 for minor 7th, M6 for major 6th, P8 for perfect 8th, and so on), and

Descriptor	Categories / Explanation	Example
ID	Derives from the printed collection and follows the sequence: type, variant, additional marker	*256.2.A2*
Type	One type title/enumeration per song	*Mačeha in sirota / The Stepmother and Her Stepchild*
Variant	The first verse of the song	*Pšenička na polju že zori / Wheat in the field is already ripening*
Region	Regional information on where the song was transcribed or recorded	*Dolenjska / Lower Carniola*
Transcriber or Collector	The person who collected or transcribed the song	*Etnomusicology Institute ZRC SAZU (Zmaga Kumer)*
Year	Year of transcription/recording	*1963*
Singer(s) Name	The person who performed the song	*Frančiška Lavrič*
Singer(s) Type	Type is based on singer's gender, e.g., female or male solo, female or male group, mixed group	*F_Solo (female solo)*

TABLE 3.1: Non-music descriptors or metadata of *Slovenian Folk Song Ballads* corpus, indicating descriptor type, the categories or explanation of that descriptor followed by an example.

Song Types (35)	Regions (22)	Transcription Year (67)	Singer Type (6)
286 (103)	Styria (111)	1957 (37)	Female Solo (292)
252 (82)	Upper Carniola (86)	1970 (31)	Male Solo (39)
287 (45)	Lower Carniola (52)	1908 (21)	Unknown (52)
256 (44)	Raba (HU) (35)	1907 (16)	Female Group (14)
267 (28)	Prekmurje (34)	1961 (16)	Mixed Group (5)

TABLE 3.2: The most frequent song types, regions, transcription years, and types of singers, with the respective number of categories and occurrences in the *Slovenian Folk Song Ballads* corpus shown in parentheses.

second, as an integer representing semitones (M6 for major 6th = 9, P8 for perfect 8th = 12, and so on).

Pitch mean. The pitch mean of the entire song is expressed in an average MIDI value, with up to 2 decimals.

Pitch direction. This descriptor indicates the relationship between the first and last notes of the song. Annotations include "+" for ascending, "-" for descending, and "=" for the song starting and ending on the same pitch.

Pitch variety or scale size. It quantifies the number of distinct pitch classes present in the score, disregarding octaves. It assigns a numerical value without specifying a particular scale name.

Tone set. This information complements pitch variety by listing distinct pitch classes as letters without octaves, such as 5: E, C, A, B, and D, indicating the unique pitch classes present in the score.

Descriptor	**Categories / Explanation**	**Example**
ID	See Table 3.1	*256.2.A2*
Verse Structure (meter)	A number or set of numbers indicating the number of syllables in a phrase.	*6*
Verse Structure (rhyme)	An alphabetic character (starting from M) indicating the rhyme pattern of the final word in a phrase, plus two types of refrain: full (R), where an entire verse line is a refrain, or half (r), where part of a verse line is a refrain. Refrains generally consist of non-lexical vocables, such as "tralala" (286.23), "jupajde" (156.16.A3), or repeated words (273.7).	*MNOPOP*

TABLE 3.3: Lyrics descriptors of the *Slovenian Folk Song Ballads* corpus, indicating the descriptor type, the categories or explanations of that descriptor, followed by an example.

Descriptor	**Categories / Explanation**	**Example**
ID	See Table 3.1	*256.2.A2*
Time signature	Time signature(s) of the song in order of first appearance.	*6/8*
Upbeat	Upbeat expressed in quarter notes.	*Example provided in the full dataset.*
Note count	Total number of notes in the song.	*37*
Measure count	Total number of measures in the song.	*9*
Range	Melodic range expressed as an interval and in semitones.	*m9 (13)*
Pitch mean	Mean pitch of a song in MIDI notation.	*65.94*
Pitch direction	Relationship between the first and final pitch, indicating ascending, descending, or equal.	*Ascending*
Tone set	Number and set of unique pitch classes in a song.	*7: G, A, B, C, D, E, F#*
Leading Tone	Indicates the presence or absence of the leading tone or f# in respect to the score's transpositions to G.	*YES*
Phrase number	Number of melodic phrases, enumerated in order of appearance.	*6*
Phrase labels	Alphabetic characters indicating melodic relationships among phrases. The same letter indicates an identical, transposed, or partially adapted melody (considering lyrics, range, or singing mistakes). Major differences are marked with a new letter.	*AABABA*
Contours	The melodic shape of individual phrases in a tune.	↗, ↘↗, ↗, ↘↗, ↗, ↘↗

TABLE 3.4: Music descriptors of the *Slovenian Folk Song Ballads* corpus, indicating descriptor type, the categories or explanation of that descriptor, followed by an example.

Leading tone. To enhance the specification of each song's tonal space, the annotation indicates the presence or absence of a leading tone. Considering the dataset, where all tunes were previously transposed to G, the leading tone refers to F#. This detail further determines whether the song can be thought of as sung in major tonality or not.

The number of phrases. This data provides a numerical annotation for the number of phrase units in each song. In the context of individual phrases, it labels each phrase with its ordinal number within the song, rather than the cumulative count of all phrases. Phrases were determined by combining the knowledge of no verse structure, punctuation marks in lyrics, and pauses in a tune.

Phrase labels. These alphabetical annotations denote melodic deviations among phrases in each song. The initial phrase is designated as "A," and subsequent phrases are labelled alphabetically, if their melody differs. If a phrase repeats the melodic material, it is assigned the same letter as the first appearance. This descriptor was simplified from [32] (Figure 3.5).

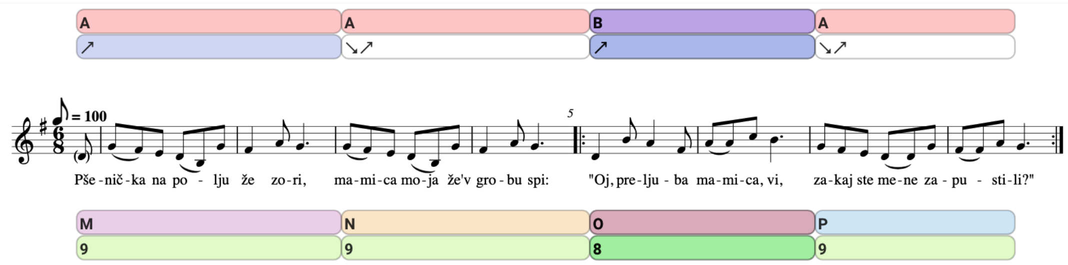


FIGURE 3.5: The song *Pšenička na polju že zori / The wheat in the field is already ripening* (type 256.A2, variant 2) with labels (up to bottom) of melodic phrase structure (AABABA), contours (↗, ↘↗, ↗, ↘↗, ↗, ↘↗), and rhyme (MNOPOP) and syllable (998989) verse structures. This example is a screenshot of an interactive score and annotations visualisation on Dezrann. (see section 3.4).

Contours. These were computed for phrases only by the principle of Huron's melodic arches (Sections 2.1 and 2.3). The computation involves analyzing the first and last MIDI values of the phrase and calculating the mean of all the values in between, provided there are more than 3 notes in the phrase. The annotations currently encompass all 9 Huron's types of melodic contours or arches (Figure 3.6).

Melodic sequence. These were converted to alphanumeric sequences, representing note names and octaves, capturing detailed pitch information (for example, A4 = approx. 440 Hz). In some of the analyses, the octave information was omitted, and for the rest, the sequences were further converted into enumerated semitone interval successions. This conversion simplifies the representation of the musical data, focusing solely on the pitch relationships between notes without considering the octaves.

Verse - metric structure. These are numerical annotations, labelling the metric structure of the verse or several syllables in each phrase (see the upper bottom labels in Figure 3.5). We acknowledge two types of syllabic annotations, the "general", which describes the general idea of all verses, and "specific", which annotates the specific phrase of the published first verse.

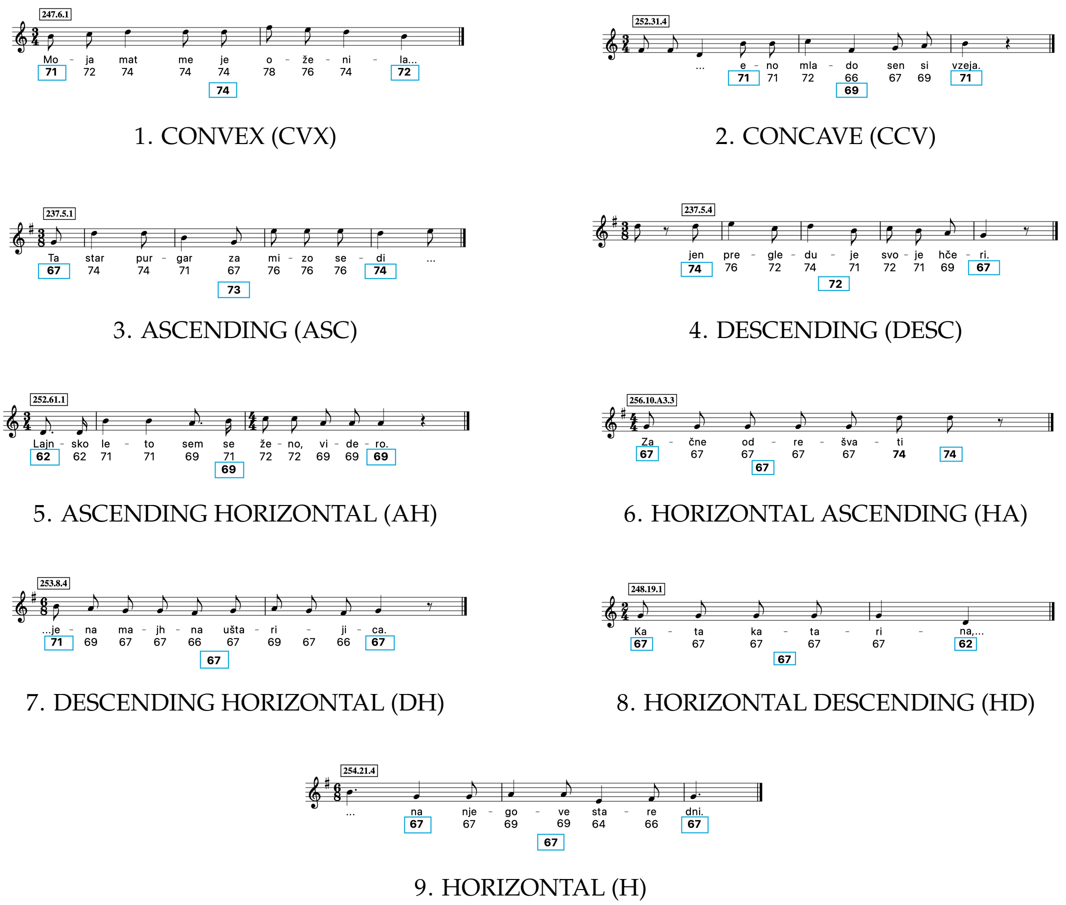


FIGURE 3.6: The 9 contour types by Huron [120] computed with MIDI values, where the first and the last MIDI value stand in place of the first and the last pitch, while the middle value describes an average MIDI value of all intermediate pitches.

Verse - rhyme structure. We took the rhyme structure annotations from the large collection of Slovenian folk song ballads. The annotations follow the alphabetical order, starting with the letter M, except R, which refers to refrain and r, which refers to a partial refrain within a larger structure [157, 93] (Figure 3.5). For this corpus, the two largest types (ID 252 and 286) were re-evaluated by members of Ethnomusicology Institute ZRC SAZU.

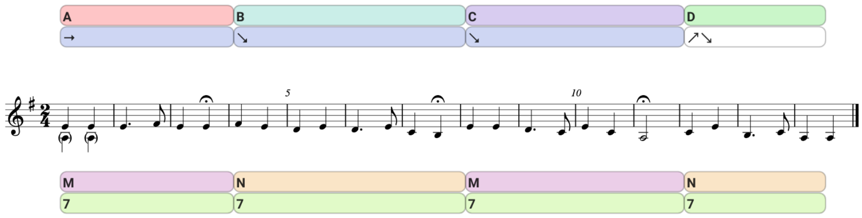
3.3 Statistical Overview

This section focuses on statistical information on 402 songs of the introduced corpus and the added descriptors, introduced in the previous section. The statistical analysis includes the majority of the music descriptors, such as scale, range, intervals, and structural labels of both lyrics and melody, as well as specific correlations between non-music and music descriptors, and melodic sequences.

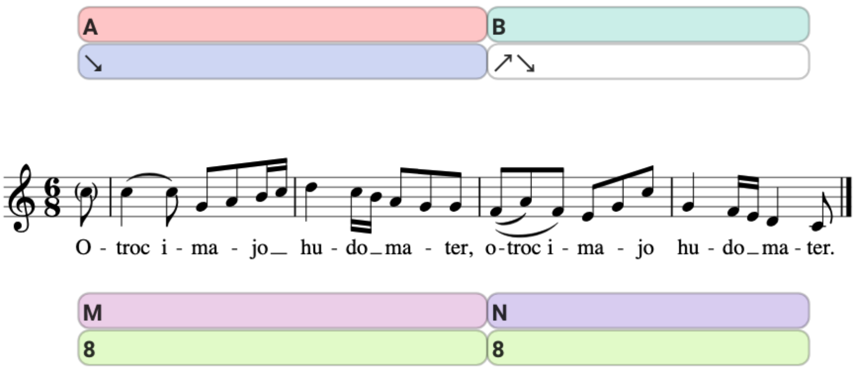
Starting with time signatures, the songs were typically sung in 2/4, 3/4, 4/4 or a mix of the first two. Melodic range usually falls somewhere between a minor seventh and a perfect octave. Extremely narrow melodies, covering a range of less than a perfect fifth, are exceptionally rare. Generally, songs tend to conclude on a higher tone, rather than an equal or descending one (to the starting tone), most of which are a third (B), first (G) or fifth (D) degree. Ten examples end on different tones. The melodies of these deviate from the resolution on the tonic chord's notes, often due to errors such as incorrect notation (Vraz's melody sketches like 248.1; example 3.71 in Figure 3.7), lack of transposition to G-major (259.3; example 3.72 in Figure 3.7), or inaccuracies in singing as indicated by the collector's notes (257.6, 286.100). Another possibility is that the song follows an older mode outside the major-minor melodic systems prevalent in the periphery of present-day Slovenia (283.6, 248.20, . . . ; examples 3.73 and 3.74 in Figure 3.7).

Most melodies are composed of approximately six or seven different pitch classes; however, examples of "few-tone" melodies (pentatonic, tetratonic, and tritonic) also exist. Among the 25 different combinations of pitch classes, more than half derive from the (G-)major scale or (G-)major scale without the third degree (Table 3.5).

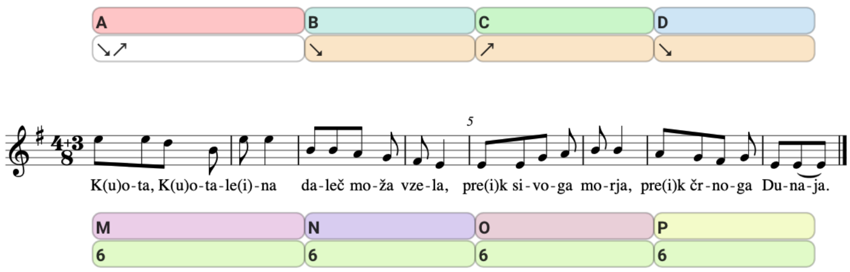
Some descriptors are worth observing in direct relation to phrases (Tables 3.6 and 3.7). There are seven different sizes of phrase structure, ranging from two to eight, with the most common being a structure of four phrases, which are mostly composed of contrasting melodic material. In terms of labels, ABCD is the most represented descriptor sequence of a song in the corpus, followed by AB and ABAB. Lyric-wise, the metric structure most commonly consists eight-seven syllables, followed by six-syllable verses. We note that this is a generalised structure, which tends to be often inconsistent among verses in a song. The most common rhyme type is MNMN, followed by four contrasting parts, MNOP and two contrasting parts MN.



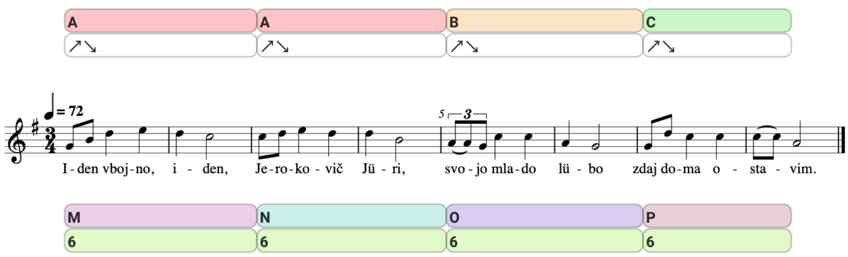
1. 248.1 (*Untitled*)



2. 259.3 ("*Otroci imajo hudo mater / The children have a strict mother*")



3. 283.6 ("*Kota, Kotalena*")



4. 248.2 ("*Iden v bojno, iden, Jerokovič Jüri / I go to war, I go, Jerokovič Jüri*")

FIGURE 3.7: Song examples with resolution "abnormalities" in respect to "tonic" (G): 3.71 (approximate sketch by Stanko Vraz), 3.72 (not transposed), and 3.73 and 3.74 (not in major-minor modes).

TS	R	PD	PC (number)	PC (toneset)
3/4 (142)	P8 (101)	↗ (179)	6 (182)	G, A, B, C, D, E, F# (145)
2/4 (94)	m7 (99)	→ (116)	7 (148)	G, A, B, C, D, F# (92)
4/4 (84)	M6 (66)	↘ (108)	5 (65)	A, B, C, D, E, F# (41)
3/8 (23)	M9 (32)	-	4 (4)	G, A, B, C, D, E (36)
6/8 (19)	P5 (27)	-	8 (2)	A, B, C, D, E (20)
3/4+4/4 (12)	m9 (21)	-	3 (1)	G, A, B, C, D (20)
7/8 (5)	m6 (20)	-	2 (1)	G, A, B, D, E, F# (7)
2/4+3/4 (4)	M7 (13)	-	-	G, B, C, D, E, F# (6)
3/4+2/4 (3)	m10 (6)	-	-	B, C, D, E, F# (6)
5/8 (3)	M10 (6)	-	-	G, A, B, D, F# (5)

TABLE 3.5: Metadata on time signature (TS), range (R), pitch direction (PD), and pitch classes (tone set) (PC) in the dataset of *Slovenian Folk Song Ballads*.

Phrase Count	PL (Melody)	SL	PL (Verse)
4 (256)	ABCD (133)	8-7 (236)	MNMN (106)
2 (84)	AB (72)	6 (49)	MNOP (78)
6 (29)	ABAB (36)	7 (41)	MN (66)
3 (23)	AABC (21)	6-5 (29)	? (39)
5 (6)	ABCB (13)	8 (24)	MNNO (26)
8 (5)	ABCDCD (10)	10-9 (9)	MNOPOP (16)
-	ABBC (12)	10 (8)	Unsegmented text (11)
-	AA (12)	Heterometric (3)	MN? (6)
-	ABC (11)	9 (2)	MNN (5)
-	AABA (10)	? (2)	MNRN (5)

TABLE 3.6: Analysis of phrase counts per song, phrase label (PL) combinations for both melody and lyrics, and the syllabic structure of lyric verses (SL) in the *Slovenian Folk Song Ballads* dataset.

3.3.1 On Year-Related Correlations

These are the correlations between various musical descriptors and the year of transcription. While numerous studies investigate how tone sets, range, and length may evolve with the age of a song, this study is constrained by data availability, which limits a more comprehensive exploration of the era from which a transcription of a variant originated. Since only the year of transcription is known, our findings provide an overview of descriptors across "older" (before or during the 1940s) and "newer" (after the 1940s) eras. The largest number of older songs were collected in Styria, while the largest number of newer songs come from Upper Carniola (Figure 3.9). We recognise that a singer in a specific year in the newer era might have performed songs from both categories as well as there have been many more similar variants collected from one region, while others remain underrepresented. This subtlety adds complexity to our analysis, making it difficult to distinguish songs solely based on the provided year.

Contour Combinations	**First Phrase (9)**	**Last Phrase (9)**
↗↘ (16)	↗ (120)	↗↘ (178)
↗, ↗↘ (12)	↗↘ (96)	↘ (129)
↗, ↗→, ↗↘, ↗↘ (12)	↘ (64)	→↘ (43)
↗, ↗↘, ↗, ↗↘ (10)	↘↗ (58)	↘↗ (18)
↘↗, ↗↘, ↗↘, ↘ (10)	↗→ (24)	↗→ (15)
↘, ↘ (10)	→↘ (15)	↗ (9)
↗↘, ↘ (6)	→↗ (11)	↘→ (8)
↗, ↗, ↗→, ↗↘ (6)	→ (11)	→↗ (2)
↗↘, ↗↘, ↗↘, ↗↘ (6)	↘→ (4)	→ (1)
↗, ↗→, ↗→, ↗↘ (6)	-	-

TABLE 3.7: Analysis of contour combinations, as well as for the first and last phrases of individual songs from the *Slovenian Folk Song Ballads* dataset.

3.3.2 On Transcriber-Related Correlations

When speaking of the eras, we must also consider transcriber's style or bias (Section 3.2.1), as we are actually comparing transcriptions of that year, and not the origin year of the actual song. We took the most common transcribers before and after the 1940s (Figure 3.8), and compared several descriptors. While most of these are equally (un)common for the two groups or cannot be further inspected due to the unbalanced data (45 for older versus 262 examples for newer transcriptions), the smallest tone set, and the metric verse structure of eight syllables almost exclusively appears in older examples (Figure 3.10), both of which confirm some of Vodušek's manual observations [308]. Older tunes were also more commonly transcribed in 6/8 time signature, had a wider range (m9, m10), a larger phrase number (eight), and rarely consisted of more than three contrasting melodic parts (ABCB, ABC, ABAB). Similar to "year-related" correlations, there is a level of unreliability of older transcriptions, unbalanced data and the fact that some tunes have many very similar variations from the same era (Section 3.3.4).

3.3.3 On Region-Related Correlations

Further information can be retrieved about the differentiation among the most represented regions. However, in some regions, many variants have been collected for the same song, whereas in others, only one or two were recorded, resulting in unbalanced data. For the purpose of the analysis, five of the most frequent regions (Table 3.2) are included in a brief descriptor analysis.

3.3.4 On Music Descriptors

The metric structure of verse was found to be one of the most common descriptors when differentiating between eras, regions, or types. This is expected for several reasons. Firstly, songs tend to adapt to verse rather than vice versa (Sections 3.1.2 and 3.2). Secondly, each type shares most of the content of the first verse. Thirdly,

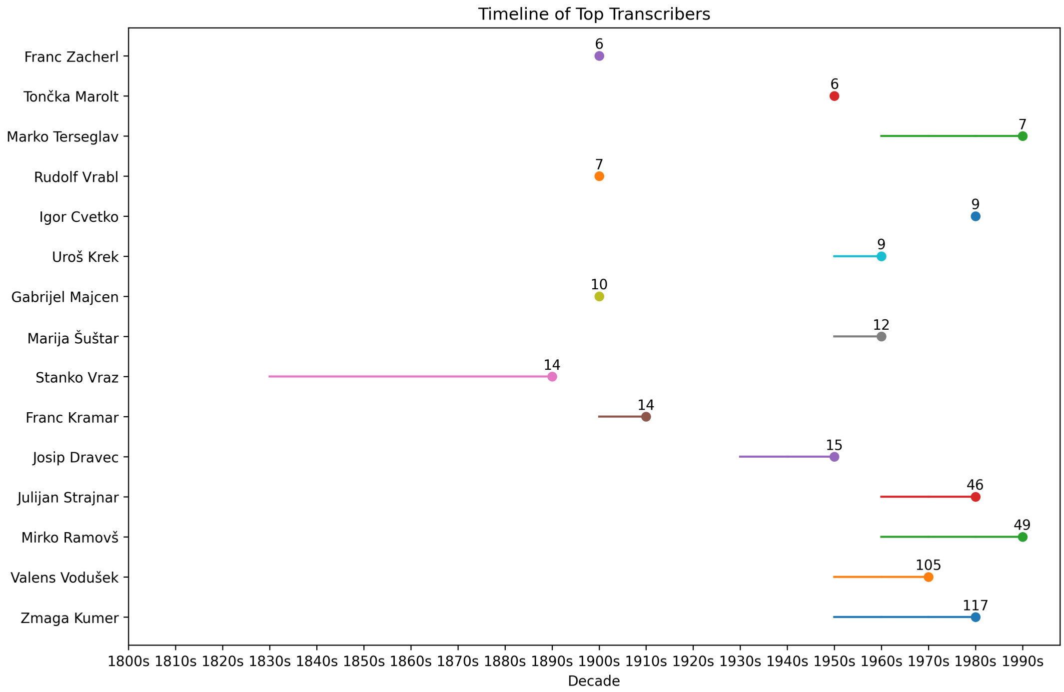


FIGURE 3.8: Most represented transcribers, with the number of transcribed songs and the years of transcriptions.

some types are more popular in certain regions, leading to the dominance of one verse type over another in those regions.

Mixed metric structrures (8-7, 6-5, 10-9, ...) account for 69% of the corpus compared to simpler structures. As anticipated by Vodušek, in correlation with the transcription era, the metric structures of newer songs are more commonly mixed, than in older, where simpler structure is more common (Figure 3.10). The opposite is true for rhyme, for which more older cases correspond with MNOP, while younger songs tend to be structured repetitively as MNMN, MN, or MNNO (Figure 3.11). No region was particularly notable in the analysis of lyric structure (suggested further reading [296]).

Similar to lyric metric structures, mixed time signatures are more commonly found in newer than in older transcriptions (Figure 3.12). The latter were generally transcribed in either 2/4, 3/4, or 4/4, or the combination of at most two of those. The 6/4 or 5/4 are only found in older songs. We found that, in relation to region, most songs in 2/4 are found in Raba (about 80% of all songs from region), followed by the ones from Prekmurje (about 45%), and Styria (about 30%). The most common meter for Upper Carniola is 3/4 (about 35%), and for Lower Carniola 4/4 (about 38%). All of there structures highly depend on collection practices. As described in Section 3.2.1 the more contemporary, the more data, meaning that the mixed time signatures, for example, could be a result of transcription style rather than the actual age of the song.

However, that is not entirely true for contours, as they provide a coarse representation of melodic sequences and do not rely of absolute accuracy of the transcription.

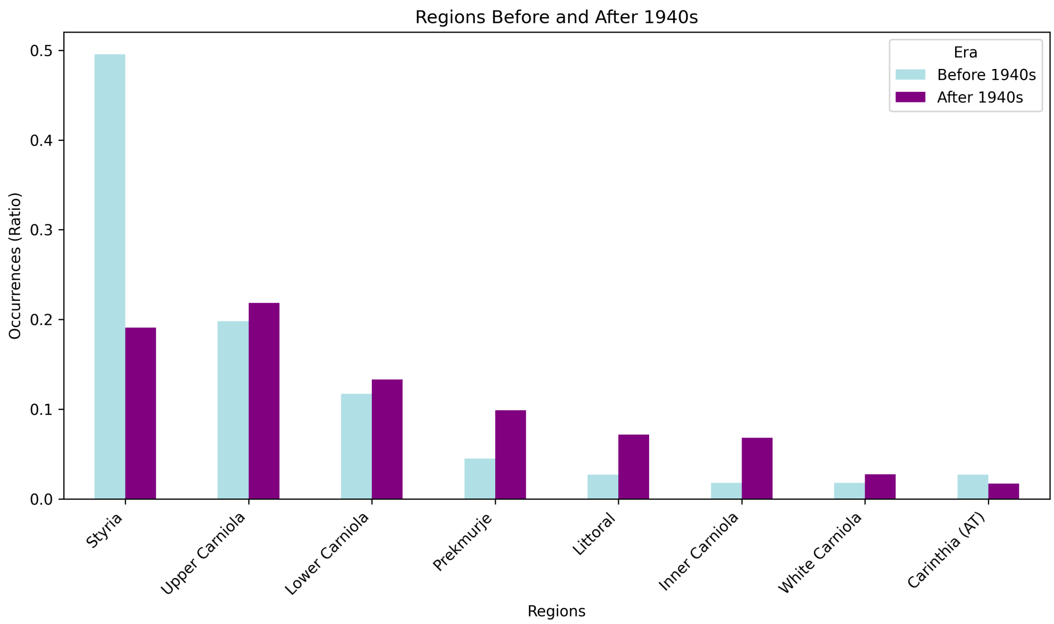


FIGURE 3.9: Songs before/after 1940s per region (not all regions are shown).

Second, contours are more significantly influenced by phrase position than other descriptors. For instance, the most prevalent contour is convex (down-up-down), more frequently found in the last phrase than the first, contrasting with other contour types (Table 3.7). Notably, descending arches are more common in the last phrase, while ascending arches predominantly occur in the first phrases. While it is hard to claim the overall relevance of these results, the trend is well visible. First phrases tend to generally go up, while last phrases go down [120].

For the middle phrases, the results are more mixed. These results are valid for all five regions regarding the last phrase (descending or convex contour in over 50% cases per region), but not consistent for the first. Ascending contour is most represented in Upper (about 50%) and Lower Carniola (about 40%), concave (about 50%) and ascending (about 20%) in Raba, convex (about 40%) and concave (about 20%) in Prekmurje, and interestingly, descending (about 30%) and ascending (about 30%) in Styria. We found no direct connection between contours and the era of transcription.

The correlations between year and tone set did not show any large differences. It was, however noticeable, that the full G-major scale is most common in the regions of Raba and Prekmurje (about 60%). In the other three, G-major scale is present at about 40%. The smallest tone sets are found in Prekmurje (three and two), while the largest are found in Lower Carniola (eight). Most songs in all regions include six or seven pitch classes.

3.3.5 Melodic Sequence Analysis

Melodic phrases typically consist of about eight tones, with each pair of subsequent tones a minor second apart, and within a range of a perfect fifth. The majority of

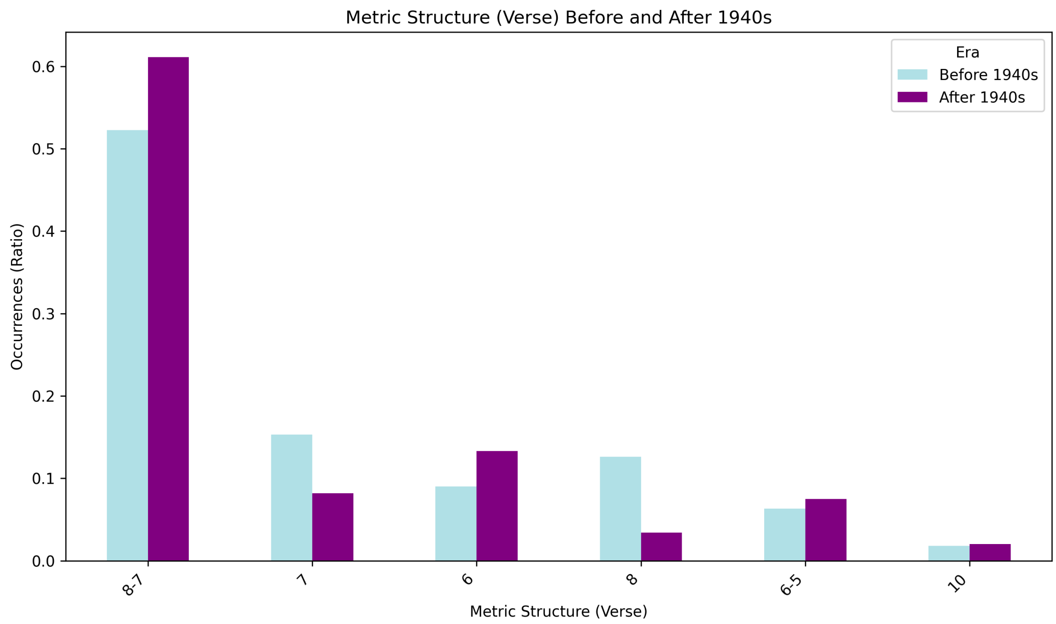


FIGURE 3.10: Most frequent (more than 1%) metric structure of lyrics before/in or after 1940s.

melodies are syllabic. To explore melodic relationships between phrases among types and variants, we tested the most frequent melodic sequences from the two most common types, 252 and 286.

Sub-sequences, consisting of the first and last four tones of a phrase, rarely occur in other types or in the same position of the same phrase, or correspond to the most frequent combinations. This varies among songs. For instance, the f#-a-d-c usually appears in the first position of the third phrase of type 286. In opposite, some melodic material is shared between types 252 and 277 (Figure 3.13), but no more than two subsequent phrases (out of four). One of the most frequent short patterns, found in types 286 and 252, and in the full dataset, is a concluding "cadence" of b-a-g. Entire melodies within our collection are seldom transferable, but may be found other types, contexts, or folk song genres beyond this particular collection. Additional digitised collections are needed for further exploration.

It is important to note that melodic transcriptions themselves are approximations, so exploring such melodies with absolute tone values without metadata may not be the most fruitful approach. A good example for this is the case of "few-tone" melodies. Ethnomusicological research describes these as being of older historical origins than those, with more than 5 tones. They were preserved in the song tradition that was collected in the 20th and 21st centuries, mainly in ritual songs and the musical tradition of peripheral and cross-border regions. As ballads did not commonly belong to ritual ceremonies, I will stress the second. The musical characteristics of peripheral or cross-border regions, unlike those of central Slovenia, more often show an intertwining with the musical characteristics of the folk music of neighboring cultures, and in some cases these regions are (were) culturally more

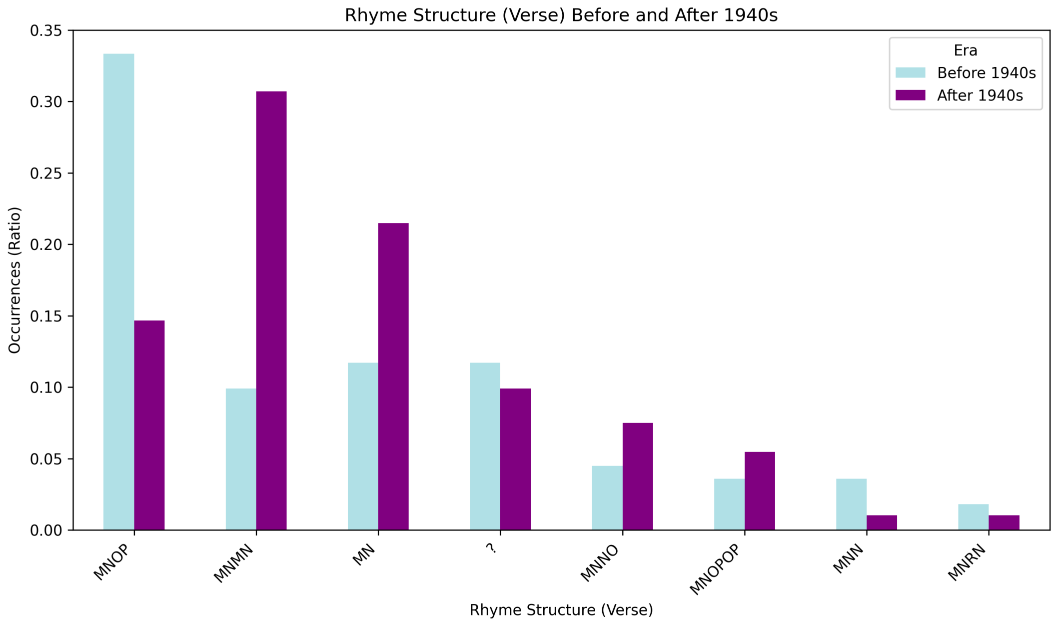


FIGURE 3.11: Most frequent (more than 1%) rhyme types in songs before/in or after 1940s.

isolated from the musical influences of other practices. In our corpus, we identified a tone set for each song and examined those, containing three to five tones, i.e. tritone to pentatonic melodies. The tri-tone set was found in only one song, the four-tone set in two, and the five-tone set in 47. The following examples show how additional field material (metadata, sound recording, etc.) shed light on the processes that need to be considered when using the dataset for computational music analysis.

First, a tri-tone song could be a mere coincidence or error in the data acquisition process, or, since an example of the song is sung by a singer from Raba, nowadays a Hungarian region, where the Slovene minority lives, its "few-tone" structure is the result of contact with neighboring cultures. It could also display the preservation of older layers of tradition in the rather isolated spaces. However, the audio recording and metadata reveal that the singer sang another verse of the song and that was not recorded by the recorders. The second stanza is melodically different and more reminiscent of a rhythmic rendition of counting.

Second, certain pentatonic melodies (252.21/A3) exhibit similarities with heptatonic melodies (252.24/A3). These melodies come from separate regions, one from Prekmurje and the other from Upper Carniola. Nonetheless, it is plausible that the transcription of either melody could be flawed, oversimplified, or fail to account for singing errors, which often become apparent only upon hearing subsequent verses. In another pentatonic example, 257.6 the transcriptor's note reveals that the singer sang only a fragment of the song, meaning the full melody was not captured.

Lastly, pentatonic melodies with a gradual movement and a narrow range of song type 256/A3 come from the same melodic base, but they are recorded in very different landscapes. Two cases in Notranjska, two in Styria, one in Lower Carniola and one in Prekmurje. The comment of ethnomusicologist Urša Šivica that refers

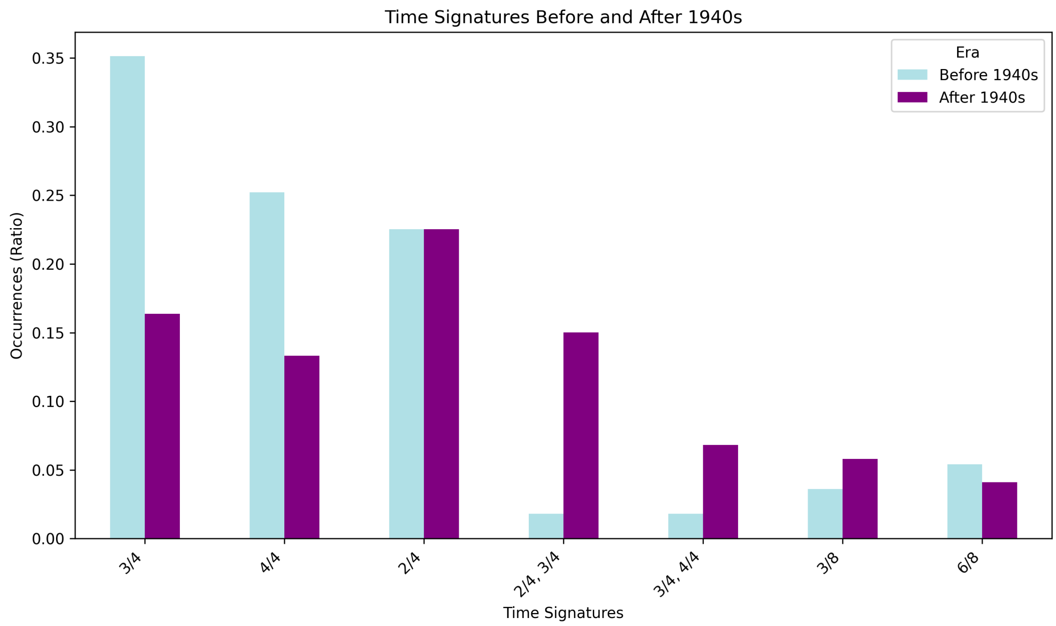


FIGURE 3.12: The most frequent time signatures (more than 1%) in songs before/in or after 1940s.

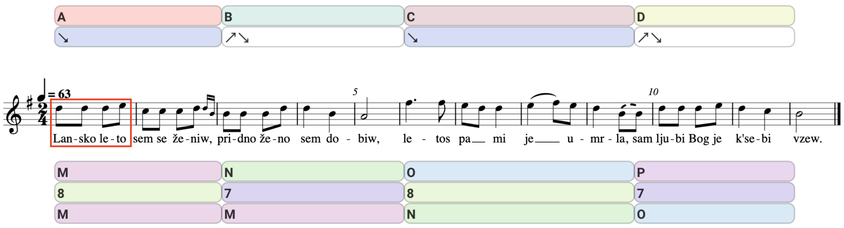
to examples of polyphony, reveals, that this melody "began to appear more consistently in the second half of the 20th century, namely in a wide ethnic territory (data for Carinthia and Littoral are not available). After 1956, as many as 33 melodies out of 39 recorded belong to the mentioned melody" [94].

3.4 Data Release and Dezrann Integration: Structure and Code

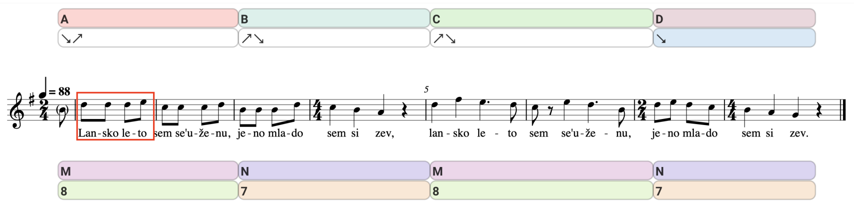
Having outlined the background, structure, and representation of the various contents, I will now explain how all of this information is integrated into the Dezrann platform. Additionally, I will provide an overview of the available data, specifying the associated licenses and where the data can be accessed.

3.4.1 Dataset Release

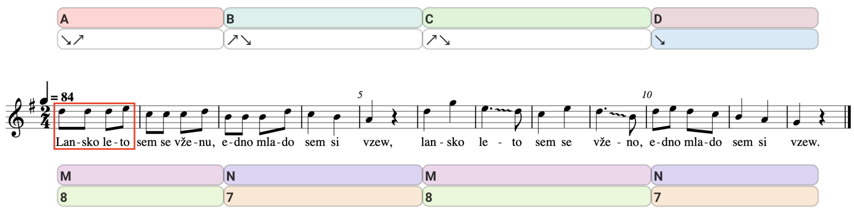
The dataset (scores in .musicXML, recordings in .mp3, metadata, and annotations in .csv/.json files) is released under the open CC BY-NC-SA 4.0 license and available for download throughAlgomus Data Repository and git. We also release the songs on the Dezrann open-source platform, where all annotations for each song are visualised, and, when audio files are available, synchronised to the scores. The Dezrann visualisations of the corpus are set to default with the initial preassigned labels, which can be altered, deleted or added by the user. If the original predicts minor melodic variations for one or more verses, these are, for now, excluded, but can be found in other editions [93]. The whole digitised corpus is also being assigned a permanent DOI.



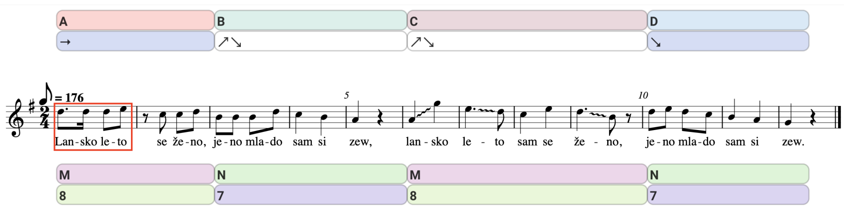
1. 252.113 (*Lansko leto sem se ženiw / Last year I got married*)



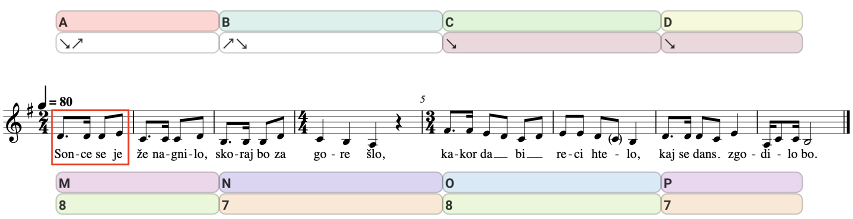
2. 252.121 (*Lansko leto sem se uženu / same as above*)



3. 252.125 (*Lansko leto sem se vženu / same as above*)



4. 252.129 (*Lansko leto sem se ženo / same as above*)



5. 277.19 (*Sonce se je že nagnilo / The sun has already set*)

FIGURE 3.13: The initial melodic phrase (pattern *d-d-d-e*), characteristic of type 252, is replicated with notable similarity (transposed an octave lower) in the identical position (initial) of the same phrase number (one) within the variant of type 277. The figure features a selection of labeled type 252 variants 113, 121, 125, and 129, juxtaposed with a congruent variant 19 derived from type 277.

3.4.2 The Dezrann Interface

Dezrann. The Dezrann platform hosts ten curated corpora, which collectively include over 1500 music pieces and more than 35000 analytical labels. These corpora are represented in various forms of data, such as scores, audio files, and video content, as well as detailed annotations and metadata. The platform integrates different representations of music, allowing users to view and interact with scores (using Lilypond/Verovio), waveforms, and spectrograms, all synchronized with corresponding audio/video content. Corpora included in Dezrann are carefully selected and curated, as well as have individual ratings regarding the quality of each individual element (audio, score, corpus, musical-time, and synchronisation). Some corpora, like the Weimar Jazz Database and Mozart Piano Sonatas, contain specific musicological or analytical annotations, while others focus on a diverse range of musical traditions, expanding the scope of research beyond the Western classical canon (see [89, 12]).

Slovenian Folk Song Ballads. One of such corpora is the Slovenian folk song ballads[4]. The corpus on Dezrann consists of 402 scores with annotations, of which 22 also include the recordings, synchronised to the displayed first verse. The digitised dataset includes a brief introduction to the contents with references and a list of scores with a search bar (Figure 3.14). Currently, each song is described by region, type, first verse "title", singer's name, reference to the pyhisical collection [93] and year of transcription/recording.

Music-wise, it provides a quick overview of pitch classes, meter and metric structure of the verse, as well as provides information on the types of data available (for instance, a music note symbol for score, heaphones symbol for recording, and a crown for annotations). More metadata can be retrieved by moving one's mouse over the (i) symbol next to the verse's "title".

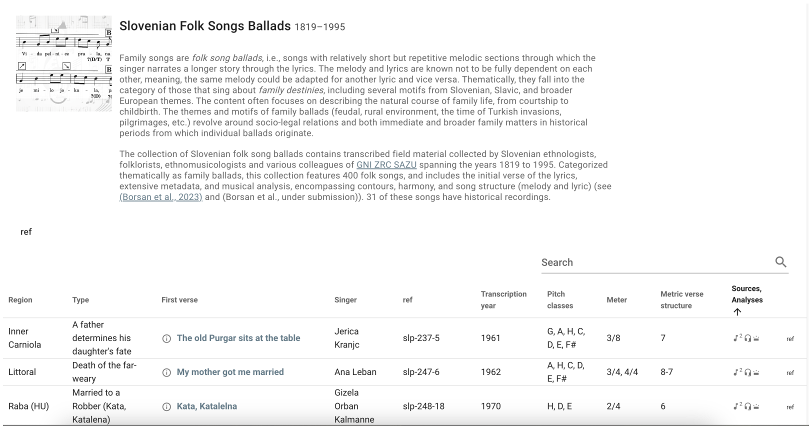
Next, a visalisaton of individual score is available with the annotation of phrase's melodic and verse's metric structure, contours and syllables per verse. These can be deleted, edited, or added by individual users. The score can be displayed in both, symbolic music notation and as wave form (individually or simultaneously) (Figure 3.15).

While the Dezrann interface is available in 7 languages [12], most information on the said corpus is available in two languages, Slovenian and English, with short corpus summaries in other languages (French, Italian, Greek, German, and Croatian).

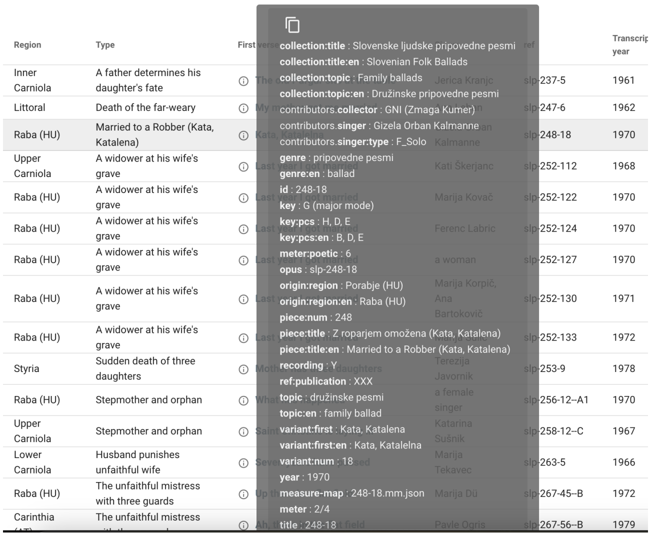
3.4.3 The Integration

To accommodate Dezrann's formats and interface, an integration script was necessary to consolidate the scores, recordings, and annotations, which were previously

[4] https://www.dezrann.net/explore/slovenian-folk-songs, accessed on 4th October 2024.



(A) Front page of the *Slovenian Folk Song Ballads* dataset.



(B) More information/metadata on an individual song.

FIGURE 3.14: The figure shows the front page of the *Slovenian Folk Song Ballads* dataset on the platform Dezrann. The top image displays the main interface, while the bottom image shows the additional information a user can retrieve by hovering over the (i) symbol.

stored in separate files. To incorporate those (and potential changes) to the platform, we wrote a script that enables easier conversion.

Scores and Recordings

This is not the first corpus that the Dezrann platform has integrated; therefore, an existing script was utilised for this purpose. All scores are accessible via Git, and it is precisely these files that are operational on Dezrann. The same applies to recordings. The Dezrann interface facilitates direct synchronisation of uploaded recordings with their corresponding scores. It generates a so-called .dez file, which enables the system to maintain synchronisation of the score based on user input, or even allows switching between multiple synchronisation sources when several recordings are available for a single score, although this is not applicable to the dataset in question.

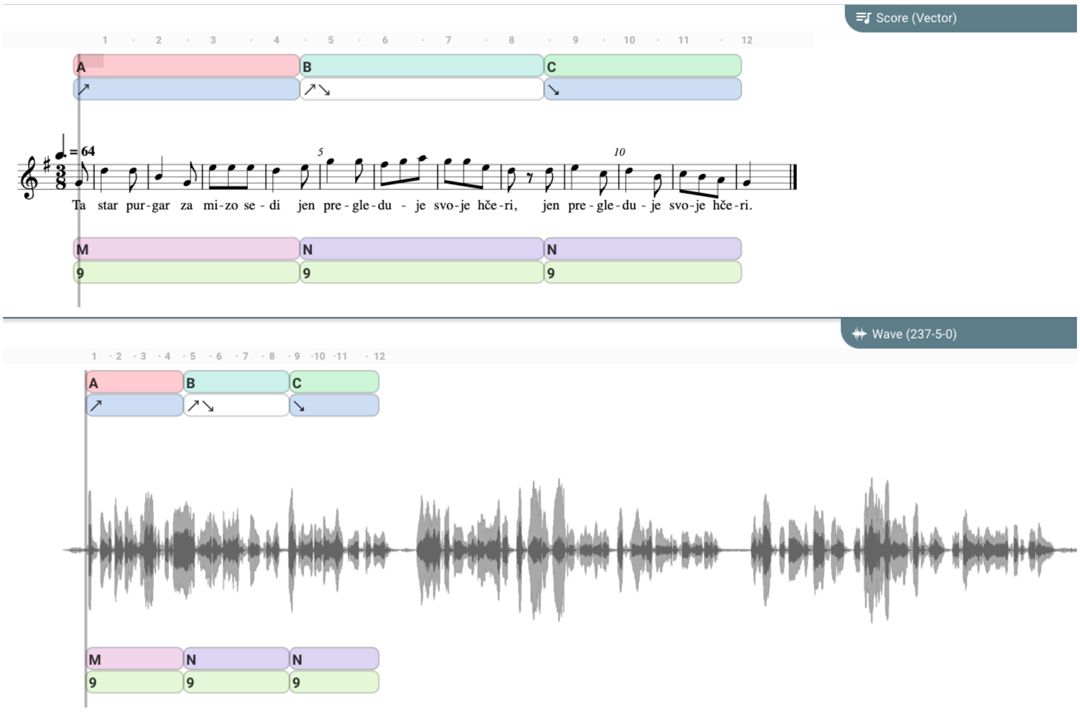


FIGURE 3.15: The figure displays the Dezrann visualisation of the score *Ta star purgar za mizo sedi / The old purgar sits at the table* (slp-237-5). It shows both symbolic notational and waveform representation with annotations.

Metadata and Annotations

The corpus was initially equipped with all the necessary data, but it did not fully align with the default metadata display settings on Dezrann. As the first folk song corpus to be integrated into the platform, certain adjustments were required. At first, the scores had not been annotated within Dezrann but were instead maintained in an Excel file, which necessitated some integration efforts. However, due to the flexibility of the Dezrann framework, these were straightforward to implement. New categories were created for annotations, and since both the annotations and scores were stored in a spreadsheet, the implementation process for both was quite similar.

We also modified field names, such as region, year of transcription, and tone set, while removing irrelevant fields like opus number, composer, as these songs did not have these. Additionally, I advocated for the ability to download metadata, recognising that ethnomusicological research relies as much, if not more, on this data as it does on the musical materials themselves. The integration script worked with two types of files—song-centred and phrase-centred annotations—depending on the required output.

Metadata. The first was used to extract metadata, such that it first began with the acquisition of Slovenian folk song metadata in relation to musical scores with a common ID from an Excel spreadsheet. This spreadsheet contained comprehensive details on each song, including types, variants, and musical notations which were crucial for the subsequent processing steps. To handle and transform the musical data effectively, two Python functions were developed.

First, `sort_pitch_classes` with a string input, which was designed to sort musical pitch classes into a predefined musical order, enhancing the consistency of data presentation. The function parses the input string to extract pitch classes sorts them according to their musical order from "C" to "B", and then outputs a sorted comma-separated string. Second, `slovenian_toneset` considered the regional variations in musical notation, and replaced the pitch class "Bb" with "B" and "B" with "H", aligning with Slovenian musical notation practices. This adjustment was essential for when Dezrann is used in Slovenian language as opposed to the English version.

Each song in the dataset was iteratively processed to generate a unique identifier which coincides with the ID with the corresponding score, and standardise pitch notation using the `sort_pitch_classes function`.

A comprehensive metadata dictionary was compiled post-processing, which was then serialised into a JSON file (slp-metadata.json), to facilitate integration with the Dezrann platform. The output includes some of the pre-assigned data, such as contributors, editors, collection number, and others, while also incorporating information on type "titles" in both Slovenian and English, and musical data, such as the first line of the variant, poetic meter, region of origin, pitch classes (processed for both standard and Slovenian notation), and year of transcription (for visualisation of the resulting file see Figure 3.14). When applicable, it also included recording data.

Annotations. The script related to annotations processes musical data a phrase-centred spreadsheet containing various attributes of Slovenian folk songs.

As the script iterates through each record of the DataFrame, it consolidates fields such as type, variant, and additional marks into a unique identifier for each piece, which correspond with the IDs found in both metadata annotations as well as score and sycnhronisation files. For every piece, the script generates multiple labels concerning musical contours, phrases, and verses (all as described in Section 3.2.2), storing them in a structured dictionary. One of the functions, `contour_conversion`, is employed to translate descriptive musical contour labels (for example, CVX, CCV) into more friendly and recognisable symbols (for instance, ↗↘, ↘↗).

All labels are subsequently serialised into JSON format and saved in files named after each song's unique ID. Moreover, the script manages the dataset's filenames, particularly amending naming discrepancies and eliminating placeholders stemming from missing data.

Some considered parameters like upbeat and similar are not of a main priority in our corpus analyses, but they provide a foundational basis for synchronising labels and visualising the score. Despite the presence of this information, along with measure numbers and phrase boundaries, there were still errors in the scores due to inconsistent or changing meters. To address this, we implemented the Measure Map methodology [99] to align the two sources effectively.

Building on our understanding of the context, structure, trends, and integration of the dataset, the next chapter will explore how patterns from this corpus were extracted and introduce the contributed methodologies supporting four distinct pattern matching tasks.

Chapter 4

Pattern Matching Task

Statement: Some parts of this chapter were expanded from the published paper at ISMIR 2023 [32], however the majority of the following content is new.

4.1 Introduction

In the previous chapters, I introduced foundational concepts from ethnomusicological practices, along with digital analysis methodologies pertinent to the materials informing this thesis as well as contextualised the corpus of Slovenian folk song ballads within its newly digitised framework, elucidating its structure and annotations. This chapter will progress to detailing the development and structuring of data and algorithms for pattern matching tasks (the rightmost part of Figure 4.1 and Figure 4.2). The initial version of key algorithms was introduced in [32]; however, other methods have since been added to extend the possibilities of resolving different pattern matching problems. We will detail on both, older and newer approaches.

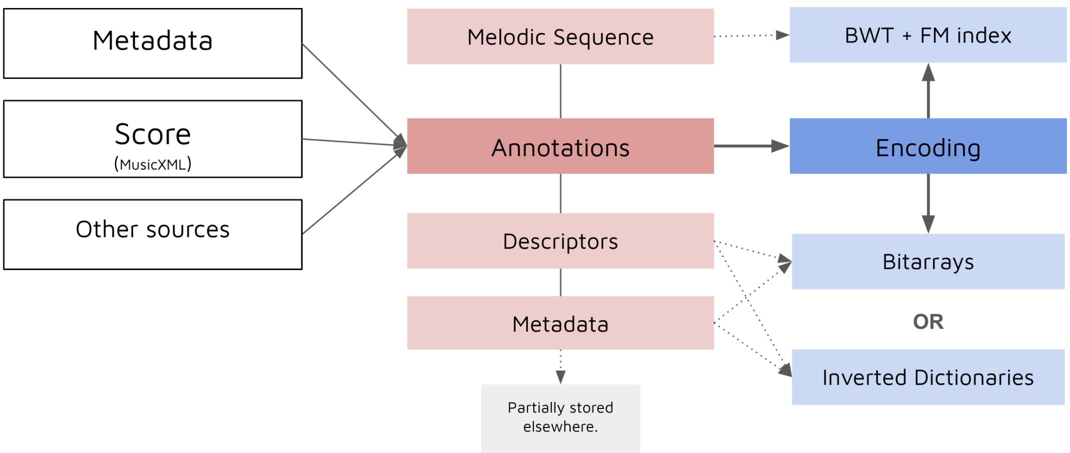


FIGURE 4.1: The process from physical scores to digitisation and annotation, followed by encoding for pattern matching tasks.

The methods are categorised into three general groups: *melodic sequence*, *descriptor*, and *mixed* pattern matching. Before delving into each of these, as shown in the leftmost part of Figure 4.2, we will address the encoding of the data previously discussed in Section 3.2, particularly the encoding that directly depends on the pattern matching methodologies.

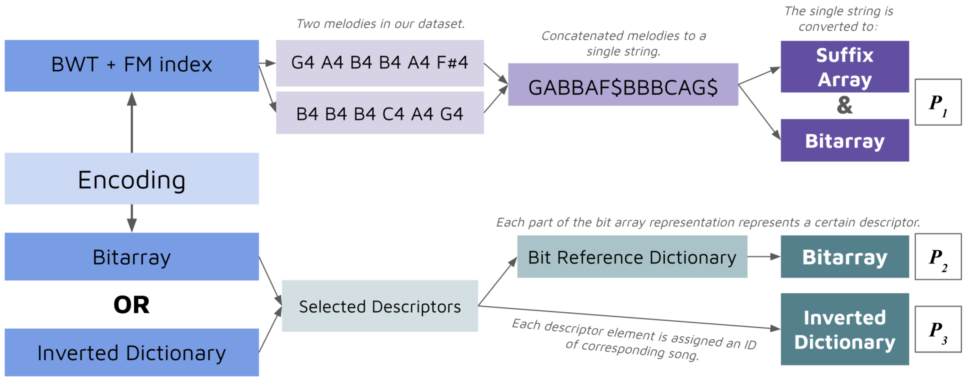


FIGURE 4.2: The process of data encoding of our dataset. (**Top**) First, there are the melodic sequences (in our case, phrases), which are stripped of octaves, sharps and flats, and then, together with the rest of the sequences in the dataset, concatenated to a single suffix array and bit vector. (**Bottom**) Each of these sequences has a number of descriptors, for which we introduce two approaches. The first out of two descriptor set matching algorithms converts descriptors of each phrase to a single bitarray, while the second one exercises an inverted approach, assigning IDs of phrases to each of the possible descriptor elements.

The pattern matching algorithms address 4 general problems (P_n):

1. P_1 (melodic pattern matching) considers a short melodic sequence, such as "G4 A4 B4". The algorithm identifies every occurrence of that sequence within the provided dataset, which, in our cases, encompasses all melodic phrases of the Slovenian folk song ballads collection along with their associated details.

2. P_2 (descriptor set matching) omits the melodic queries and instead considers one to several descriptors with zero to at most one associated element, such as an ascending contour and a 3/4 time signature. It searches for all song phrases in the dataset that correspond to the selected descriptors with their associated elements. The method also allows ignoring certain descriptors or interchanging between different combinations of descriptors.

3. P_3 (multiple descriptor set matching) focuses on descriptors as well. However, here, multiple descriptor elements can be associated with a single selected descriptor, meaning there can now be zero, one, or *several* descriptor elements associated with descriptor category. For example, an ascending or horizontally ascending contour, and 3/4 or 6/8 time signature. Similar to P_2, it searches for all song phrases that correspond to any (but at least one) of the provided descriptor elements per descriptor.

4. P_4 (mixed pattern matching) focuses on combining melodic and descriptor pattern matching. It considers both, melodic sequence (for instance, "G4 A4 B4" (P_1)), with corresponding descriptors, such as ascending contour and a 3/4 time signature (P_2). Or, in turn, pairs melodic sequence query with a descriptor query (P_3), such as ascending or horizontally ascending contour and a 3/4 or 6/8 time signature. In both cases, the pattern is considered to be a

match when both melodic and descriptor problems are resolved, meaning, the melodic sequence and descriptors must match one or more phrases in the provided dataset.

In the following sections, I will first describe the fundamental data representations and their relationships, and later, move on to detailing the individual pattern matching tasks highlighted above.

4.1.1 Definitions

We consider a dataset $D = \{(p_1, d_1), (p_2, d_2), \ldots, (p_i, d_i), \ldots, (p_n, d_n)\}$, where n is the number of phrases in a set, and i is the index or identifier, which, in our case, is the song phrase ID. Each entry is split into a sequence p_i and its corresponding descriptors d_i.

Pitch sequence. A pitch sequence p_i in D is an alphabetic representation of an individual phrase of individual song, regardless of their octave, sharps and flats, and rhythmic values[1]. For example, a first phrase of a song displayed in Figure 4.4, "G4 B4 D5 D5 E5 D5 D5 B4 A4 G4", would be represented as $p_{239.10.A.1}$ = "gbddeddbag".

Descriptors. Each phrase has a corresponding list of descriptors which consists of a selection of descriptors $Categories = (\Delta^1, \Delta^2, \Delta^3, \ldots, \Delta^t, \ldots, \Delta^m)$, where m is a number of a descriptor, and Δ^t is a *specific descriptor*, such as contour, time signature, or similar. Each Δ^t must have at least one value belonging to the finite set of values $V(\Delta^t)$. For instance, let us consider 5 descriptors from our dataset[2]:

$$\left(\Delta^1, \Delta^2, \Delta^3, \Delta^4, \Delta^5\right) = (\mathrm{POS}, \mathrm{LBL}, \mathrm{CONT}, \mathrm{SYLB}, \mathrm{TS})$$

where:

$$\begin{aligned}
V(\Delta^1) &= \{\mathrm{First}, \mathrm{Middle}, \mathrm{Last}\},\\
V(\Delta^2) &= \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}\},\\
V(\Delta^3) &= \{\nearrow, \searrow, \nearrow\rightarrow, \searrow\rightarrow, \searrow\rightarrow, \nearrow\searrow, \searrow\nearrow, \rightarrow\searrow, \rightarrow\nearrow, \rightarrow\},\\
V(\Delta^4) &= \{10, 9, 8, 7, 6, 5\},\\
V(\Delta^5) &= \{2/4, 3/4, 3/8, 6/8, 4/4\}.
\end{aligned}$$

For P_2, a *single descriptor sequence* d is a list consisting of a single element for each $V(\Delta^t)$, defined as $d = (d^1, \ldots, d^t, \ldots, d^m)$, where each $d^t \in V(\Delta^t) \cup \{\star\}$. The special

[1] This step was facilitated as the majority of songs in the dataset were previously transposed to a G-centred tonal space, which minimised the impact of sharps and flats on the analysis. The exclusion of octaves was similarly justified, as contours, scales, and ranges were precomputed and included among the descriptors, providing a more precise understanding of melodic shape and range for focused analysis.

[2] POS = phrase position, LBL = phrase label, CONT = contour, SYLB = verse syllables, TS = time signature. If indicated in the process, the descriptors can extend to any metadata information, including regions, titles, singers, and similar.

value $\star$ indicates that the descriptor is to be ignored. For example, a sequence may be $d = (\text{First}, \text{A}, \nearrow, 7, 2/4)$, where we consider all of the descriptors $V(\Delta^t)$, meaning POS, LBL, CONT, SYLB, and TS. If some descriptors are wished to be ignored, such as $d = (\text{First}, \star, \nearrow, \star, 2/4)$, the $\star$ informs the algorithm to exclude LBL ($V(\Delta^2)$) and SYLB ($V(\Delta^4)$) from the pattern matching task.

Conversely, P_3 addresses *multiple descriptor sequences* $md = (md^1, md^2, \ldots, md^m)$, where $md^t \subset V(\Delta^t)$ denotes a set of descriptors, potentially containing multiple descriptors per category. For example, md = ({First}, {A}, {$\nearrow$, $\searrow$}, {6, 7, 8}, {2/4, 4/4}), or if we want to omit a descriptor, md = ({First}, $\star$, {$\nearrow$, $\searrow$}, {6, 7, 8}, {2/4, 4/4}), where $\star$ is the empty set.

Dataset. Each phrase p_i of D is associated with a *descriptor sequence* $d_i = (d_i^1, \ldots, d_i^m)$ and $d_i^t \in V(\Delta^t)$ for $1 \leq t \leq m$.

Given a first song phrase of song 239.10.A (Figure 4.4), the dataset entry consists of melodic sequence $p_{239.10.A.1}$ = "gbddeddbag" and corresponding descriptors $d_{239.10.A.1}$ = (First, A, $\nearrow\searrow$, 8, 3/4). An overview of the entire data encoding is presented in Figure 4.3. This will be elaborated upon in detail as each individual problem is addressed.

4.1.2 Problem Statement

P_1: **Melodic pattern matching** Given a melody p, find every i, where the melodic sequence p is a substring of p_i. For example, if p = "bac", it should be found as a match with $p_{239.10.A.1}$ at position 1 (Figure 4.4).

P_2: **Descriptor set matching** Given one descriptor sequence d, find every i, such that $d_i = d$. For instance, if d = (First, $\nearrow\searrow$, A, M, 8), it will be matched with the descriptors of $p_{239.10.A.1}$, hence the match will equal to 239.10.A.1. However, if d = (Last, $\searrow\rightarrow$, B, $\star$, 8), then $p_{239.10.A.2}$ will be found as a match (Figure 4.4).

P_3: **Multiple descriptor set matching** Given a list of descriptor sets $md = (md^1, \ldots, md^m)$, find every i, such that $d_i = (d_i^1, d_i^t, \ldots, d_i^m)$ and $d_i \in md^t$ or $md_t = \star$.

For example, if md = $(\{\nearrow\searrow\}, \{A, B\}, \{M\}, \{8\})$, then descriptors of $p_{239.10.A.2}$ will be matched, while if md = $(\{\nearrow\searrow, \rightarrow\searrow\}, \{A, B\}, \{M, N\}, \{8, 9\})$, then both, descriptors of $p_{239.10.A.1}$ and $p_{239.10.A.2}$ will be matched with md (Figure 4.4). The final matches are expressed as a set of song IDs as integers i.

P_4: **Mixed pattern matching (melody + descriptors)** Given (p, d) or (p, md), find every i, such that both p_i and d_i match the query. For example, if p = "gbd" and md = $(\{\nearrow\searrow, \rightarrow\searrow\}, \{A, B\}, \{M, N\}, \{8, 9\})$, $p_{239.10.A.1}$ is a match, but if p = "bac", then it is matched with $p_{239.10.A.2}$ (Figure 4.4).

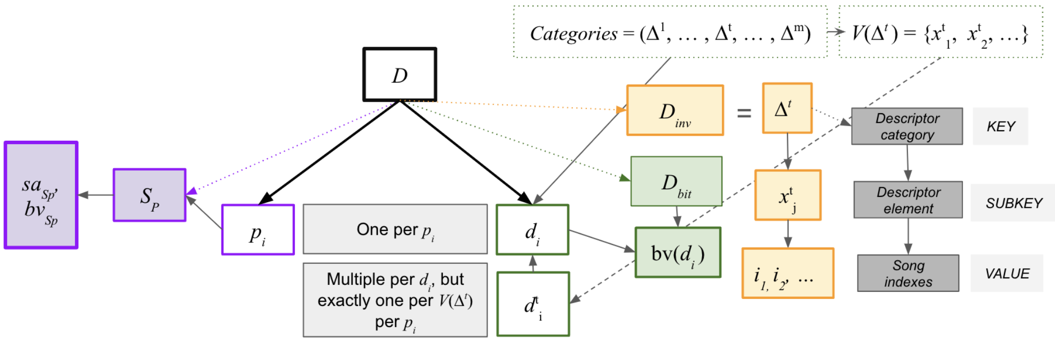


FIGURE 4.3: The dataset encoding scheme for all, *melodic sequence*, and two types of *descriptors*. D is a dataset, which holds pairs of sequences p_i and a corresponding descriptors d_i. Within the latter, there are several descriptor elements d_i^t per descriptor Δ^t. These belong to a set of possible descriptor values $V(\Delta^t)$, where Δ^t is a pre-selected descriptor category. (**Left**) The purple (darkest) squares show S_p which is a concatenated string of all p_i in D. The latter is described by two additional elements, a suffix array sa_{S_p} and a bit vector bv_{S_p}. (**Right**) The orange (light) colored squares in the center describe the data structure D_{inv} of descriptors. This structure is composed of keys Δ^t and a subset of keys x_j^t where the values are the corresponding dataset entry IDs (i). The green (darker) colored squares on the left represent an alternative to that method, where each d_i is converted to a bit array $\text{bv}(d_i)$ and stored in D_{bit}.

4.2 Melodic Sequence Pattern Matching (P_1)

4.2.1 Problem Definition

Given a query melody p and a dataset D composed of a set of sequences and corresponding descriptors $D = \{(p_1, d_1), (p_2, d_2), \ldots, (p_i, d_i), \ldots, (p_n, d_n)\}$, the objective is to efficiently find all occurrences of p in D, where p is a substring of p_i.

4.2.2 Melodic String Encoding

The melodic string is encoded in two stages. First, the spaces, sharps, and flats are removed from each sequence, which in our case is a melodic phrase, and a $ symbol is added to the end of each one. For instance, a melodic phrase string "G4 B4 A4 B4 F#4 A4", as described above, is transformed into the prepared phrase string $p_1 =$ "gbabfa$"

All p_i are concatenated into a single string $S_p = p_1 p_2 \cdots p_i \cdots p_n$, with each phrase p_i ending with $.

Now a single string, S_p is further transformed into one suffix array sa_{S_p} and one bit array bv_{S_p}. The suffix array sorts the suffixes lexicographically. Simultaneously, the bv_{S_p} marks positions within the string: 0 for non-$ characters and 1 for $, aiding in identifying phrase boundaries and lengths within the suffix array. The $bv_{S_p} = b_1 \ldots b_{|S_p|}$ is defined as $b_i = 1$ if $S_p[i] = \$$, otherwise $b_i = 0$.

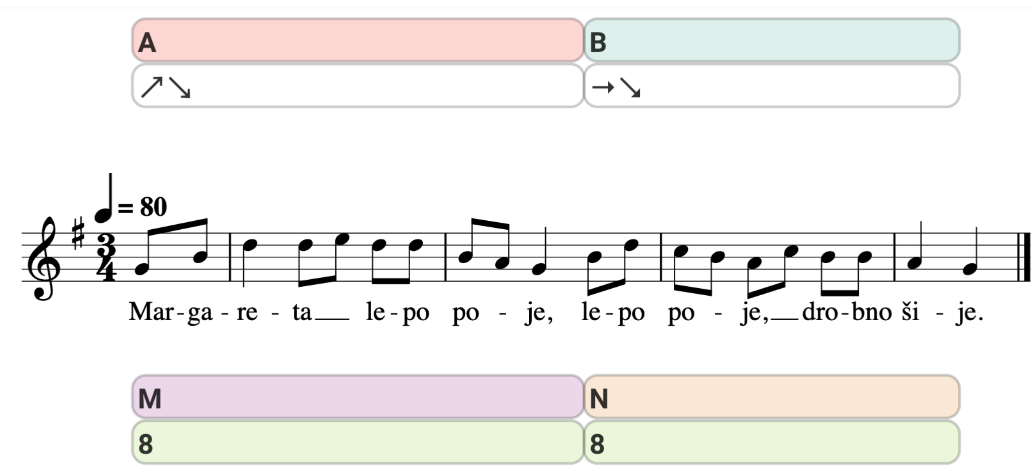


FIGURE 4.4: The song *Margareta lepo poje / Margareta sings beautifully* (type 239/A, variant 10) with 2 phrases, labeled (**Top to Bottom**) with melodic phrase structure (A, B), contours (↗↘, →↘), verse structure in terms of a rhyme (M, N) and syllables (8, 8). This example is a screenshot of an interactive score and annotations visualisation on Dezrann.

4.2.3 Melodic Pattern Matching Task

The Algorithm 1, which considers a melodic query p and a dataset D first retrieves the phrases and their exact starting positions from a suffix array in linear time, and then filters these matches with bitwise operators. An index data structure such as a compressed suffix array is computed and stored to retrieve all occurrences of a pitch sequence.

When a melodic query p is matched at position k in S_p, the corresponding identifier i—which, in our specific case, is the ID of each phrase of each song in the dataset, and its position in p_i are retrieved using the pre-computed bv_{S_p}. This process utilises the functions $\text{rank}_1(bv_{S_p}, k)$ and $\text{select}_1(bv_{S_p}, k)$, which determine, respectively, the number of occurrences of 1 in the prefix of length k of bv_{S_p}, and the index of the k-th 1 in bv_{S_p}.

Hence, we retrieve set of of tuples i, j, such that the query p occurs in phrase p_i at position j within p_i with $i = \text{rank}_1(bv_{S_p}, k)$, and $j = k - \text{select}_1(bv_{S_p}, i)$.

Retrieving the phrases and positions of a query, meaning a pitch sequence p of length m, is done in time $O(m + occ)$, where occ is the number of occurrences of p in S_p provided that rank and select operations are performed in constant time.

4.2.4 Example

Consider the melodic query p = "bac" and a dataset D represented by sa_{S_p} and bv_{S_p}, of which all p_i were converted to a single string S_p = "gbddeddbag$bdc**bac**bbag$" and indexed by sa_{S_p} and bv_{S_p}. The encoding to p, sa_{S_p}, and bv_{S_p}, returns:

Algorithm 1 Melody Match Algorithm

Input: Melodic query p, dataset D, suffix array sa_{S_p}, bit vector bv_{S_p}

Output: A set of of tuples (i, j), such that p occurs in phrase p_i of D at the position j.

Use the suffix array sa_{S_p} to locate all positions k where the query p matches a substring in S_p

Initialise matches $\leftarrow \emptyset$

for each matched position k in S_p **do**

  $i \leftarrow \text{rank}_1(bv_{S_p}, k)$

  $j \leftarrow k - \text{select}_1(bv_{S_p}, i)$

  matches $\leftarrow$ matches $\cup \{(i, j)\}$

end for

return matches

$$
\begin{aligned}
S_p &= \text{“gbddeddbag\$bdc}\textbf{bac}\text{bbag\$”}, \\
sa_{S_p} &= [23, 22, 11, 16, 20, 9, \mathbf{15}, 19, 8, 18, 12, 2, 14, 17, 7, 13, 6, 3, 4, 5, 21, 10, 1], \\
bv_{S_p} &= [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \mathbf{1}, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1].
\end{aligned}
$$

- **Suffix Array Search**: To find the starting positions of the query p in the concatenated string S_p, we use the suffix array (refer to Section 2.4.1 and Table 2.1 on how suffix array is constructed). In this example, the result is $k = 15$, indicating that the pattern "bac" begins at position 15 in S_p.

- **Bit Vector Mapping**: Next, we use the bv_{S_p} to determine which phrase contains the substring. We apply $\text{rank}_1(bv_{S_p}, k)$, which tells us how many **1**s there are up to position $k = 15$ in the bv_{S_p}. In this case, the result is $i = 1$, indicating that the pattern first occurs in the second substring (or, in our instance, song phrase) of S_p. Using $\text{select}_1(bv_{S_p}, i)$, we find the first **1** in the bv_{S_p} at position 11.

- **Final Position Calculation**: To determine the exact position within the second substring, we use the formula $j = k - \text{select}_1(bv_{S_p}, i)$, where $k = 15$ (from the suffix array) and $\text{select}_1(bv_{S_p}, 1) = 11$. This gives $j = 4$, meaning the query p starts at the 4th position in the second substring of S_p.

The first (and only) tuple we obtain in *matches* (returned by Algorithm 1) is (4, 2) can then be converted to a phrase identifier i (of p_i), and an exact phrase position. In our case, this would mean that the searched query was found at position 4 of phrase 2.

4.3 Descriptor Pattern Matching

This encoding and pattern matching methodology was initially introduced as a single-descriptor pattern matching in [32], meaning only one descriptor element per descriptor category was considered per matching task (P_2). Subsequently, a second methodology was developed to accommodate multiple descriptor elements per descriptor, prompting to adopt a slightly modified approach. This latter development necessitated a restructuring of the encoding phase, impacting both the encoding of the dataset and the potential queries, as well as initiated a new pattern matching algorithm. Both methods will be defined, explained, and will include examples of implementation in the following subsections.

4.3.1 Problem Definition

Descriptor set matching with bitwise operators. Given a descriptor query d, where $d = (d^1, d^2, d^3, \ldots, d^m)$, and each element d^t for $1 \leq t \leq m$ is a descriptor chosen from a specific pre-selected descriptor category Δ^t, or is ignored ($\star$), we aim to identify all instances, the query d matches the descriptors of i.

Multiple descriptor set matching with inverted dictionary. Given a descriptor query $md = (md^1, md^2, \ldots, md^t, \ldots, md^m)$, where each $md^t \subset V(\Delta^t)$ we aim to identify all i in dataset D such that $d_i = (d^i_1, d^i_2, ..., d^i_t, ...d^i_m)$, matches at least one d^i_t per Δ^t.

4.3.2 Descriptor Set Matching with Bitwise Operators (P_2)

First, I detail one of the two approaches for descriptor pattern matching, that converts the dataset into bits, and then, performs the single-query pattern matching as described in the subsection above.

Dataset encoding. Let us consider dataset D, where d_i are the descriptors of p_i.

The latter is, in this case, turned into a sequence of bits $\mathrm{bv}(d_i)$ per d_i (Algorithm 2), such that it first considers the selected descriptor categories of interest, and then constructs a reference dictionary R of which the keys are all selected descriptor categories Δ^t, and their values are all possible associated descriptor elements (Figure 4.5).

Each descriptor $V(\Delta^t)$ can be represented by b^t bits, with $b^t = \lceil \log_2 |(\Delta^t)| \rceil$. Here, each value $d^t_i \in V(\Delta^t)$ is associated with a bitarray bv^t_i. Each descriptor sequence d_i is then stored as a bitarray $\mathrm{bv}(d_i)$.

Each section of bits in a bitarray $\mathrm{bv}(d_i)$ is allocated for a specific descriptor $\mathrm{bv}(d^t_i)$, where each descriptor's bit length may vary. The bitarray can be represented as:

$$\mathrm{bv}(d_i) = \mathrm{bv}(d^1_i) \ldots \mathrm{bv}(d^m_i)$$

Each $\mathrm{bv}(d^t_i)$ in the bitarray represents a different descriptor. or example, consider the descriptor sequence $d_i = (\text{First}, \text{A}, \nearrow\searrow, 4/4)$, where the bitarray $\mathrm{bv}(d_i)$ = 00 0 1111 000 ‘‘00’’ represents the phrase position (“First”), ‘‘0’’ represents the phrase label ("A"), ''`1111`'' represents the contour type ("↗↘"), and ''`010`'' represents the time signature ("4/4") (refer to Figure 4.5).

The exact number of bits allocated to each unique descriptor element and their order within $\mathrm{bv}(d_i)$ are determined by a bit reference dictionary R, retrieved from the list of selected descriptor *Categories* and those, available in D. The length of each bit segment, and consequently the total size of the bitarray $\mathrm{bv}(d_i)$, depends on the diversity of descriptor elements within Δ^t.

Algorithm 2 Dataset and Query Conversion, and Reference Dictionary Creation

Input: descriptor sequence d, dataset D, *Categories*
Output: A set of integers i, such that d_i matches d
Initialize matches $\leftarrow \emptyset$
$R \leftarrow \mathit{Categories}, D$
$\pi(d) \leftarrow \mathrm{bv}(d), R$
$D_{\mathrm{bit}} \leftarrow D\ \ R$
$\mu \leftarrow \mathit{Categories}$
return $(\pi(d), D_{\mathrm{bit}}, \mu(d))$

Algorithm 3 Descriptor Set Matching with Bitwise Operators

Input: $\pi(d)$, D_{bit}, $\mu(d)$
Output: matches, a set of integers i, such that $\mathrm{bv}(d_i)$ matches $\pi(d)$
Initialize matches $\leftarrow \emptyset$
for $\mathrm{bv}(d_i)$ in D_{bit} **do**
    $\mathit{masked\_query} \leftarrow \pi(d)$ AND $\mu(d)$
    $\mathit{masked\_descriptors} \leftarrow \mathrm{bv}(d_i)$ AND $\mu(d)$
    $\mathit{xor\_result} \leftarrow$ (masked_query XOR masked_descriptors)
    $\mathit{result} \leftarrow$ all bits of $\mathit{xor\_result}$
    if all bits of $\mathit{result} \rightarrow 0$ **then**
        matches $\leftarrow$ matches $\cup\, i$
    end if
end for
return matches

Pattern matching task. Reflecting on the problem definition presented at the beginning of this section, we aim to identify similar entries within a dataset based on their descriptors. Given a descriptor query d, and a converted dataset D stored as D_{bit} (refer to Figure 4.3), the method searches for all instances where bit representation $\pi(d)$ of the query matches $\mathrm{bv}(d_i)$, optionally applying a mask $\mu(d)$ (see definition below and Figure 4.5), a bitarray of the same length as $\pi(d)$, and considering only the bits of $\pi(d)$ in positions where mask bits equal to 1 (Algorithm 3).

As indicated in the encoding section, each descriptor Δ^t is represented by $\lceil \log_2 |V(\Delta^t)| \rceil$ bits, each value x_j^t in $V(\Delta^t)$ being associated with a specific bitarray. Each descriptor sequence $d_i = (d_i^1, d_i^2, d_i^3, \ldots, d_i^t, \ldots, d_i^m)$ is stored as a concatenation of a bitarray of selected descriptors $bv(d_i) = (bv(d_i^1), bv(d_i^2), bv(d_i^3), \ldots, bv(d_i^t), \ldots, bv(d_i^m))$

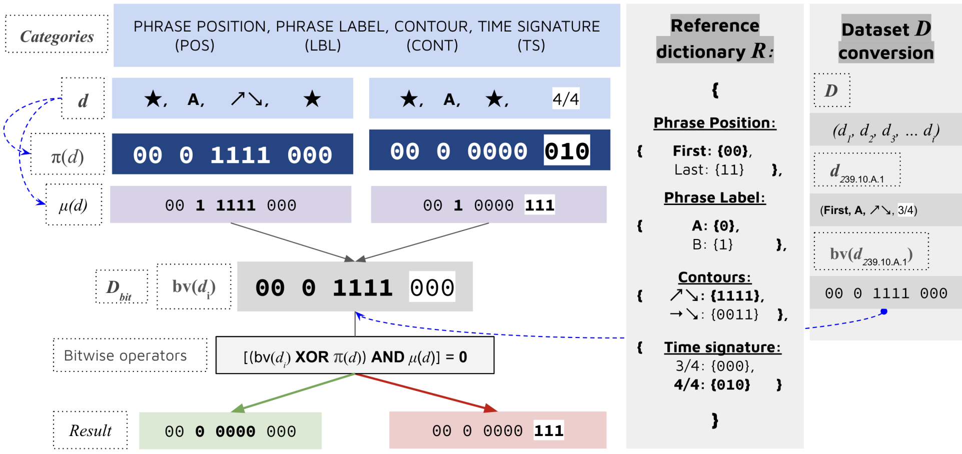


FIGURE 4.5: Descriptor pattern matching with bitwise operators. The method compares the query bitarray $\pi(d)$ against each converted dataset entry $\text{bv}(d_i)$. It also applies a mask $\mu(d)$ that determines which *Categories* are to be compared and which to be left out. We provide two examples, in each of which two out of four possible descriptors are considered: **Left** example considers phrase label (A) and contour (↗↘), while the **Right** includes the phrase label and time signature. The bitarrays for both the query and dataset entries are generated using (**Far right**) the pre-computed reference dictionary R, which is used for converting (follow the blue arrows) both the dataset (Dataset D conversion on the right) and query into bitarrays. (**Bottom**) A green arrow and a square indicate a match was found, while a red arrow and a square point out a mismatch (with a mismatched element highlighted in a white square).

(see Algorithm 2 for data encoding, and Algorithm 3 for single descriptor set matching).

A descriptor query d, where $d = (d^1, d^2, \dots, d^t, \dots, d^m)$, is associated with two bit masks $\mu(d) = \mu^1 \dots \mu^m$ and $\pi(d) = \pi^1 \dots \pi^m$, where

$$\begin{cases} \mu^t = \pi^t = 0\dots0 \\ \text{if } d^t = \star, \\ \text{where } \star \text{ represents the descriptors we wish to ignore,} \\ \mu^t = 1\dots1 \text{ and } \pi^t = \text{bv}(d^t) \\ \text{otherwise.} \end{cases}$$

Checking whether a descriptor d matches a descriptor d_i is done in $O(1)$ time, provided that the bitarray fit in one machine word, with the following procedure:

$$[(\text{bv}(d_i) \text{ XOR } \pi(d)) \text{ AND } \mu(d)] = 0.$$

Example. If we consider our dataset D of phrase entries from Slovenian folk song ballads with accompanying metadata, the initial step for each row is to identify the extracted phrase melody sequence and its corresponding metadata. An identifier i is then constructed for each entry, or in our case, each song phrase. When comparing descriptors alone, we proceed to encode the selected query and subsequently engage in pattern matching.

Given descriptor query d = (F, A, ↗↘, 4/4), then its default associated masks are $\mu(d)$= 1111111111 and $\pi(d)$ = 0001111010 (see reference dictionary R in Figure 4.5).

Now, let us consider a D where $d_{239.10.A.1}$ = (First, A, ↗↘, 3/4). Its associated bitarray is $\mathrm{bv}(d_{239.10.A.1})$ = 0001111000. Since $[\mathrm{bv}(d_{239.10.A.1})\mathsf{XOR}\pi(d)\mathsf{AND}\mu(d)]$ = 000000000, d matches $d_{239.10.A.1}$, hence i=239.10.A.1 would be returned as a result in the set of matches.

For example, if $d = (\star, A, ↗↘, \star)$, where $\star$ represents a descriptor we wish to ignore, then

$$\begin{aligned}
\mu(d) &= 00 \cdot 1 \cdot 1111 \cdot 000 \\
\mathrm{bv}(d_{239.10.A.1}) &= 00 \cdot 0 \cdot 1111 \cdot 000 \\
\pi(d) &= 00 \cdot 0 \cdot 1111 \cdot 101 \\
\hline
\mathrm{bv}(d_{239.10.A.1})\ \mathsf{XOR}\ \pi(d) &= 00 \cdot 0 \cdot 0000 \cdot 010 \\
\mathrm{bv}(d_{239.10.A.1})\ \mathsf{XOR}\ \pi d\ \ \mathsf{AND}\ \ \mu(d) &= 00 \cdot 0 \cdot 0000 \cdot 000
\end{aligned}$$

However, if, for instance, $d = (\star, A, \star, 4/4)$, the result of

$$\mathrm{bv}(d_{239.10.A.1})\ \mathsf{XOR}\ \pi(d)\ \ \mathsf{AND}\ \ \mu(d) = 0000000111$$

This leads to a mismatch, even though the phrase label (A) matches, because the time signature is also considered. In this case, our query specifies a 4/4 time signature, whereas $d_{239.10.A.1}$ has a time signature of 3/4 (Figure 4.4 for the song example, and Figure 4.5 for visualisation example).

In the example provided, $d_{239.10.A.1}$ represents an example in our dataset, a specific song phrase that is to be compared against our query d. The mask applied in this context is $\mu(d)$ (as defined in the paragraph titled *Pattern matching task* and shown in Figure 4.5). The comparison process involves matching $\mathrm{bv}(d_{239.10.A.1})$ using the XOR, as well as AND bitwise operators. This is due to the mask that is taken into account, ensuring that any potential mismatches at the initial and final descriptor elements of $\pi(d)$ (the bitarray representation of the query d) are disregarded ($\star$), thereby enabling the two examples to align (refer to the bits highlighted in blue).

4.3.3 Multiple Descriptor Set Matching with Inverted Dictionary (P_3)

Second descriptor matching technique considers multiple descriptor entry, meaning, one can select zero, one, or multiple values per descriptor category.

Dataset encoding. In contrast to the previous approach, this method requires encoding D into a single inverted dictionary of dictionaries, $\mathrm{D}_{\mathrm{inv}}$. This structure is constructed by first including all selected descriptors, Δ^t, as the primary keys, with their corresponding values, $V(\Delta^t)$. Each key x_j^t of each main key Δ^t represents an entry in the dictionary Δ^t and is linked to its associated values, the identifier i. Each element in d_i that corresponds to x_j^t is thus mapped within $\mathrm{D_{inv}}$ (Figure 4.3).

For example, consider the descriptor POS (representing positions within a sequence). The dictionary $\mathrm{D_{inv}}$ might look like this: POS: {F: {ID1, ID3}, M: {ID4}, L: {ID2, ID5}}. Here, F (First), M (Middle), and L (Last) are keys under the descriptor POS, each associated with a set of identifiers that indicate which songs or sequences correspond to these positions. This inverted structure allows for efficient retrieval of sequences without requiring further data conversion, unlike the bitwise operators approach, which necessitates encoding before pattern matching (Section 4.1.1).

Let us consider three of our song phrases as an example (see visualisations of each in Figure 4.6):

$D = \{(p_{239.8.A.1}, d_{239.8.A.1}), (p_{252.76.1}, d_{252.76.1}), (p_{279.2.B.1}, d_{279.2.B.1})\}$ with selected

$Categories$ = (POS, LBL, CONT), where:

$d_{239.8.A.1}$ = (F, A, $\nearrow\searrow$)

$d_{252.76.1}$ = (F, A, $\nearrow\searrow$)

$d_{279.2.B.1}$ = (F, A, $\searrow\rightarrow$)

Then, the $\mathrm{D_{inv}}$ of D is constructed as:

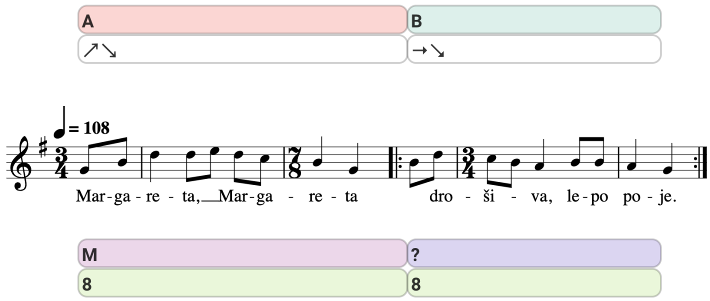
$$\mathrm{D_{inv}} = \left\{ \begin{array}{l} \mathsf{POS} : \left\{ \begin{array}{l} \mathrm{F} : \{239.8.A.1, 252.76.1, 279.2.B.1, \ldots\}, \\ \mathrm{M} : \{\}, \\ \mathrm{L} : \{\} \end{array} \right\}, \\ \mathsf{CONT} : \left\{ \begin{array}{l} \nearrow\searrow : \{239.8.A.1, 252.76.1, \ldots\}, \\ \searrow\rightarrow : \{279.2.B.1, \ldots\} \end{array} \right\}, \\ \ldots \end{array} \right\}$$

Pattern matching task. The Algorithm 4 provides a method for P_3, which aims to consider multiple possible descriptor elements of the same descriptor Δ^t in a query md. It operates as follows:

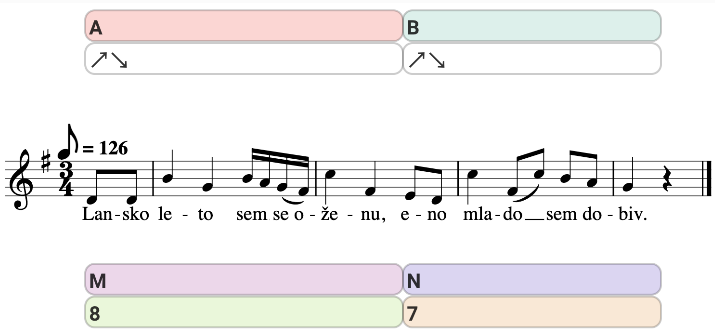
- Let dataset D represent our dataset, containing melodies and a corresponding list of descriptors.
- Let $\mathrm{D_{inv}}$ denote the inverted dictionary constructed from D where Δ^t is a key in $\mathrm{D_{inv}}$, and x_j^t is a key of Δ^t, of which the corresponding values are identifiers i. Each element d_i^t corresponds to the descriptor x_j^t in $\mathrm{D_{inv}}$. Avoiding the descriptor is expressed with a $\star$.

Given a query

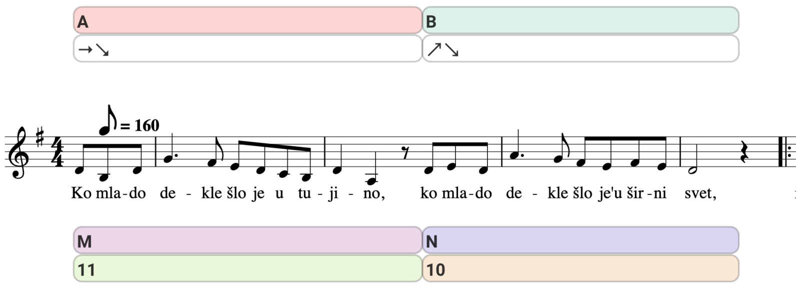
$$md = (\{x_1^1, x_2^1, x_3^1\}, \{\star\}, \{x_1^2\}, \{x_1^3, x_2^3\})$$



(A) *Margareta, Margareta* (239.8.A)



(B) *Lansko leto sem se oženu / Last year I got married* (252.76)



(C) *Ko mlado dekle šlo je u tujino / When a young maiden went abroad* (279.2.B)

FIGURE 4.6: Three annotated examples of songs from the collection of Slovenian folk song ballads on the platform Dezrann.

The algorithm first checks whether the descriptor is found in selected *Categories* = (Δ^1, Δ^2, Δ^3), for example (POS, LBL, CONT), and then finds all instances of p_i, where the Δ^t and x_j^t match one of the selected descriptor elements in *md* query, for example, $\{\{F\}, \{A\}, \{\nearrow\searrow\}\}$.

Let Δ^t be an individual key, and let x_j^t be a of each Δ^t. If Δ^t is associated with multiple x_j^t, then matches for x_j^t are stored as a separate set of identifiers $S(x_j^t)$.

To establish a single identifier set $S(\Delta^t)$ per descriptor Δ^t, we compute the union of all possible sets $S(x_j^t)$.

Given two sets $A(\Delta^t)$ and $B(\Delta^t)$, each containing identifiers, the method always computes their union, denoted $A(\Delta^t) \cup B(\Delta^t)$. This results in a single set that includes all distinct elements present in either $A(\Delta^t)$, $B(\Delta^t)$, or both.

The process is repeated for each Δ^t in a query md, such that we acquire exactly one set per considered Δ^t.

If a query contains multiple $V(\Delta^t)$, resulting in multiple sets of identifiers i, then the intersection of all these sets is performed. This ensures that the final set of matches includes only the IDs i, which have at least one x_j^t for each Δ^t in common.

Algorithm 4 Descriptor Set Matching with Inverted Dictionary

```
Input: A multiple descriptor query md, a dictionary D_inv
Output: A set of IDs i, where md matches with d_i of p_i
Initialise matches ← ∅
for descriptor_set in md do
    identifiers ← ∅
    if descriptor_set ≠ ⋆ then
        for value in descriptor_set do
            identifiers ← identifiers ∪ D_inv[descriptor_set][value]
        end for
    else
        identifiers ← D_inv[descriptor_set][value]
    end if
    matches ← matches ∩ identifiers
end for
return matches
```

Implementation example. Let us revisit Figure 4.6 and consider the following example:

$$Categories = (\mathsf{POS}, \mathsf{LBL}, \mathsf{CONT})$$

a query

$$md = (\{\mathrm{F}\}, \{\mathrm{A}\}, \{\nearrow\searrow, \searrow, \nearrow\})$$

and suppose that

$$D = \{(p_{239.8.A.1}, d_{239.8.A.1}), (p_{252.76.1}, d_{252.76.1}), (p_{279.2.B.1}, d_{279.2.B.1})\} \text{ (Figure 4.6).}$$

We initialise pattern matching by finding the matching IDs from the dataset that align with F for POS, resulting in:

$$\text{pos\_ids} = \{239.8.A.1, 252.76.1, 279.2.B.1\}$$

We repeat the process for the LBL descriptor A, and again identify the IDs:

$$\text{label\_ids} = \{239.8.A.1, 252.76.1, 279.2.B.1\}$$

from the dataset.

As CONT descriptor query contains more than one descriptor element, we perform a union over the set $\{\nearrow\searrow, \searrow, \nearrow\}$, resulting in

$$\text{cont\_ids} = \{239.8.A.1, 252.76.1\}$$

Finally, by computing the intersection of these sets,

$$\text{matches} = \text{pos\_ids} \cap \text{lbl\_ids} \cap \text{cont\_ids} = \{239.8.A.1, 252.76.1\}\,,$$

we obtain the final set of matches. This procedure utilises union operations to gather potential matches from each descriptor category and employs intersection operations to narrow down to IDs common across all criteria, ensuring that the final set of matches

$$\{239.8.A.1, 252.76.1\}$$

satisfies all the conditions specified in the query, leaving out the song phrase, of which the contour (it being $\rightarrow\searrow$) does not match with any of those, found in our query md.

4.4 Mixed Pattern Matching (P_4)

This section provides a concise overview of the methods previously discussed, demonstrating how they are integrated to achieve mixed pattern matching. By "mixed", we mean the simultaneous matching of melodies and descriptors. Specifically, given a dataset D, we aim to find all matches for two types of combined queries: p with d to match melodic sequences and single descriptors using bitwise operators, or p with md to match melodic sequences along with multiple descriptor elements per descriptor using inverted dictionaries.

Pattern matching task.

- **Melody Matching:** Initially, the algorithm attempts to find a match for the provided melody (Section 4.2). Upon identifying a match, it proceeds to the descriptor matching phase.
- **Descriptor Matching:** Descriptors are then matched using the selected method, either by utilising bitwise operations (Section 4.3.2) or the inverted dictionaries (Section 4.3.3).
- **No Melody Match:** In the absence of a melody match, or subsequently a descriptor match, the algorithm moves on to the next entry in the dataset.

Result: The algorithm returns all matches with their identifiers, along with the p_i and queried pattern's positions within the phrase, and listing the matched descriptors d_i of that particular entry.

Implementation example. We will once again refer to Figure 4.6 and assume our dataset $D = \{(p_{239.8.A.1}, d_{239.8.A.1}), (p_{252.76.1}, d_{252.76.1}), (p_{279.2.B.1}, d_{279.2.B.1})\}$. Given selected descriptor categories, *Categories* $= (\texttt{LBL}, \texttt{CONT})$, and a mixed query, we have:

- p = "ddb" and $d = (\text{A}, \nearrow\searrow)$, which corresponds to song phrase index 252.76.1.
- p = "bd" and $md = (\{\text{A}\}, \{\nearrow\searrow, \rightarrow\searrow\})$, which correspond to song phrase indices 239.8.A.1 and 279.2.B.1.

4.5 Conclusion

This chapter introduced four distinct pattern matching problems: P_1 (melodic sequence), P_2, P_3, and P_4. All of these propose a separate algorithm (with an exception of P_4, which is a culmination of P_1 and P_2 or P_3), which was methodologically explained.

The effectiveness and utility of these (and other) pattern matching techniques vary significantly depending on the type of music, the specific nature of the input, and the diverse research questions posed. This necessitates adaptable methodologies. Whilst P_1 addresses an efficient comparison of melodic sequences, it is insufficient, for example, for music that has developed outside of the Western classical music tradition or similar notational music systems, such as verbally transmitted music. Thus, our method, especially with P_2 and P_3, by incorporating descriptor pattern matching, offers a framework that extends beyond conventional approaches to include more context-sensitive and diverse musical interpretations. The descriptor methods thereby complement the sequential queries effectively.

Optionally, some descriptors are possible to be integrated into P_1, where the results can be further refined, using the information on the phrase number and the exact position. Adding these to our query ensures, that only the matches satisfying both the position within the phrase and phrase number criteria are returned by finding the intersection of the filtered results. Furthermore, although these methodologies have been primarily applied to music, they may also be extendable to non-musical queries.

The following chapters will follow these conclusions by providing an extensive evaluation of the introduced methods, focusing technical performance, as well as testing the methods on two case studies. The first case study will focus on the main corpus, specifically Slovenian folk song ballads, and the second case study will examine an existing digitised collection of children's songs.

Chapter 5

Implementation and Evaluation

When reviewing and comparing contributions, particularly the more "technical" ones within the scope of MIR, it is evident that they almost invariably include some form of evaluation. Given the wide range of music tasks and queries considered in this field, one is prompted to reflect on what exactly is being assessed, what these assessments truly reveal about their contributions, and who finds these results valuable.

Interdisciplinary research, such as MIR, presents considerable challenges in terms of evaluation. As noted by Downie [68], a general issue is that "there has been no way for research teams to scientifically compare and contrast their various approaches." Several authors have emphasised [69, 68, 84, 301, 125, 75, 293] the need for a more holistic approach that encompasses all aspects of the algorithms, including their interactions with the materials used and the diverse profiles of the individuals involved.

On one hand, [69, 68] call for adopting ready-made evaluation methods from similar fields, such as text algorithmics. However, they also frequently argue that music information tends to be more (or too) complex[1] or unsuitable for a direct transfer of those methods to MIR. The authors also emphasise the difficulty in finding and/or storing large testing collections that cover all types of notation, audio, annotations, and different music genres. As mentioned in Section 2.3, there are numerous ways to link music data, making it difficult to evaluate them collectively, as they are often very distinct from one another. Additionally, they acknowledge, that there are endless possibilities for music queries, many of which do not originate from real-world examples, leading to potentially misleading evaluations[2].

Next, Downie discusses the concepts of relevance, precision, and recall. While these are suitable in particular cases, they cannot replace a comprehensive evaluation, as the definitions of what is relevant, precise, and so forth, are far from being objectively answerable in all instances. To this end, Downie hopes that, by making the query records as data-rich as possible, a "reasonable person" standard could

[1] There is no uniform understanding of what music is. It encompasses several components such as rhythm, melody, text, and praxis, among others, all of which can be easily varied, and yet represent a very similar content. Music also comes in many formats and encodings, making the processes of methodology and its evaluation even more challenging.

[2] A similar view is also put forward by [125] and [293].

emerge as the criterion for judging the relevance of returned items. That is, there should be enough information contained within the query records that reasonable persons would concur as to whether a given returned item satisfied the intention of the query. The validity of the "reasonable person" assumption would, of course, be subject to empirical verification [68].

Lastly, he mentions issues of accessibility and intellectual property laws. While it is true that much more data is available now compared to 2004, questions regarding ownership and publication rights remain persistent challenges in developing consistent, well-informed collections of music.

Since then, a line of comprehensive ideas on how to build these frameworks for evaluating such algorithms, especially in terms of computational music co-creativity, has been succinctly proposed. One such example is [125], who introduces the concept of the four Ps or perspectives. She argues that, in their case, computational creativity should be assessed beyond the commonly narrow focus, which only considers the *process*[3] and *product*[4]. In her view, computational creativity, being a human-related process, should also account for the *person or producer*[5] and the *press* or *environment*[6]. All four Ps contribute to the novelty and value of the proposed method, meaning that discussing only a few aspects (or Ps) limits the range of areas and disciplines that may find these ideas valuable [125].

A similar caution regarding how we evaluate, or rather validate, methods and their experiments was proposed by [293]. They reflect on validity[7], which assesses the legitimacy of the conclusions drawn from an experiment or study. In the context of MIR research, they draw on the four types of validity classification by [273][8], which are essential for ensuring the robustness, reliability, and generalisability of systems across diverse datasets and real-world scenarios. The authors indicate that, "[a]n experiment does not possess 'truth value'. Validity is a property of a conclusion made given evidence collected from an experiment."

The components of an experiment—units, treatments, design, observations, and setting—have major consequences for the validity of conclusions drawn from it, whether it is statistical conclusion validity, internal validity, construct validity, or external validity" [293]. This is crucial because, as has already been previously stressed by [301], "reaching wrong conclusions from evaluation experiments may not only hamper the proper development of our field, but also make us follow completely wrong research directions."

Building on the discussed matters above and upgrading the Figure 4 from [301] (Figure 5.1), the most important considerations in our evaluation is to: 1. clearly

[3]"What the creative individual does to be creative" [125].

[4]"What is produced as a result of the creative process" [125].

[5]The individual agent [human or machine] that is creative." [125].

[6]"The environment in which the creativity is situated" [125].

[7]Validity addresses whether the findings are credible and reflect the real-world phenomena they intend to represent, ensuring that the results are not merely artifacts of the specific experimental setup but can be generalised to broader contexts.

[8]These concepts are found in [301] as well.

communicate what is being evaluated and the reasons for it, 2. evaluate from multiple perspectives, and 3. understand the limitations of the testing corpus and methodology (state where evaluation is corpus-specific rather than a generalised result), which will be based on real-world data. Evaluating pattern matching is not a grateful task as, evident from the discussion as well as Chapter 2, different (ethno)music (ological) studies rarely agree on what a pattern is and what is the most meaningful result of such approach.

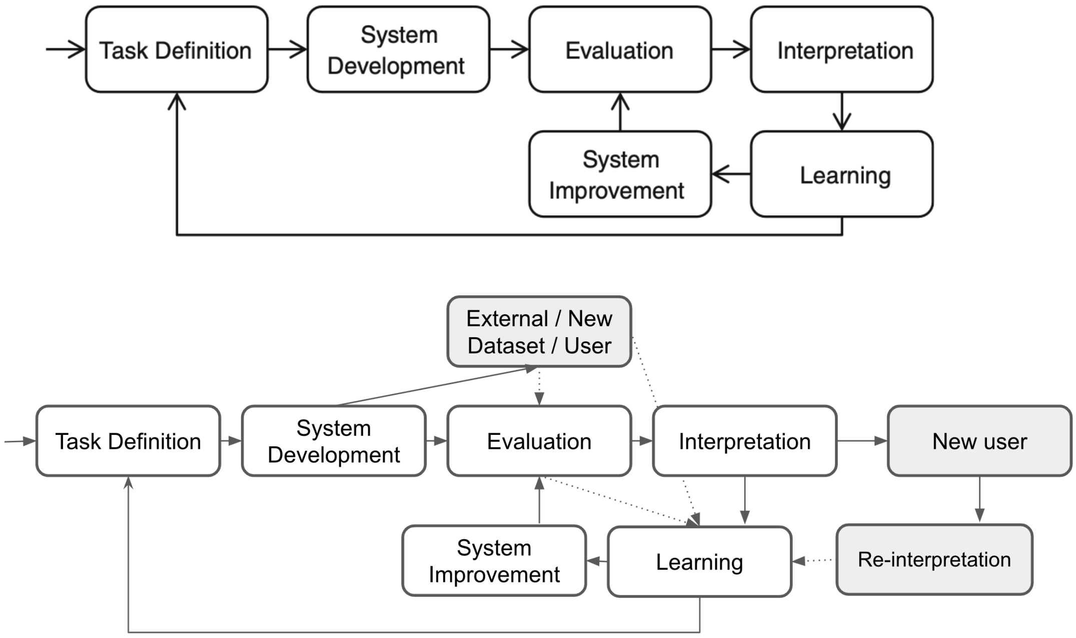


FIGURE 5.1: MIR process from task creation to evaluation. **(Up)** Figure 4 from [301]. **(Down)** Suggestion for an update of that figure with the inclusion of new user/data interaction with the existing task.

This chapter, hence, is dedicated to a comprehensive evaluation of the algorithms and findings presented in this thesis. We begin by examining the most technical aspects, such as algorithm performance and the accuracy metrics associated with specific classification problems. This examination is further enriched by a discussion of the metrics results, aiming to go beyond mere numerical data to critically assess whether it is even possible to design or formulate meaningful patterns for such studies, and to understand what the accuracy of these patterns truly represents. Furthermore, we investigate whether employing more "complex" patterns can improve these outcomes.

To address the first, we begin by evaluating the execution time of all phases, namely *dataset conversion*, *descriptor pattern matching*, and *melodic pattern matching* (P_1), using the Slovenian folk song ballads corpus (SLP). The descriptor approaches are divided into two categories, assessing both the *bitarray* (P_2) and *inverted* (P_3) methods.

Following this technical evaluation, we assess the functionality in both, technical, (e.g., accuracy/metrics on a classification problem) and circumstantial evaluation of the algorithms through two distinct case studies. The first case study focuses on the SLP corpus, examining the tune families within the most represented song type in the collection. The second case study utilises the corpus from a GMGM project (*Glasba mladih po 1945 in glasbena mladina Slovenije / Youth music since 1945 and Jeunesse Musicale*), a collection of children's songs, to evaluate the suitability of this corpus for children in comparison to the SLP corpus, which was not curated specifically for a younger audience, centering the problem around melodic parameters, such as range and intervals. Each case study involves different corpora, with varying levels of metadata and annotations, and applies different algorithms.

Finally, we synthesise the findings from all three sections to provide a comprehensive discussion on the evaluation of the proposed approaches, extending the analysis beyond mere statistics. It is important to note that, by excluding evaluations from various user profiles, the assessment does not necessarily reflect the general performance of the algorithms, but rather their effectiveness within the defined scope. Given that this algorithm is intended to interact with diverse research queries, descriptors, and datasets, we will discuss why an algorithm that performs well in our specific context may perform poorly with another dataset or query, and vice versa.

5.1 Performance Evaluation

5.1.1 Implementation

P_1: **Melodic Sequence.** Melodic sequence pattern matching was implemented using the Python library `pysdsl`. Specifically, we utilised `pysdsl.SuffixArrayBitcompressed` and `pysdsl.BitVector` for string encoding. The pattern matching process employed the `rank` and `select` functions to search for specific patterns within the encoded strings.

P_2: **Descriptors.** For descriptors conversions and matching in P_2, the Python library `bitarray` was used. Both datasets and queries were converted into bit arrays. Pattern matching was performed using Python's `XOR` and `AND` operators, with and without masks, to identify matches between dataset and query descriptor sets.

P_3: **Descriptors.** Descriptor matching for P_3 was implemented using set operations, specifically `union` and `intersection`, to evaluate the relationships between different descriptors. Additional operations were performed using standard Python functions related to dictionary manipulation.

Testing Environment. All testing was performed in Jupyter Notebook on a MacBook Air (M2, 2022) with 16 GB of RAM.

5.1.2 Performance

Here, we provide a detailed evaluation of the time and memory metrics (Figure 5.2) for various code implementations, noted above. The time metrics represent the average execution duration for each loop iteration, reported in microseconds (μs) or milliseconds (ms), along with their respective standard deviations. These measurements assess the consistency and performance efficiency of each implementation. The memory metrics, reported in mebibytes (MiB), include the peak memory usage and the incremental memory consumption during the execution of the program. These metrics explain the memory efficiency and potential overhead introduced by different implementations, as noted above.

Data Conversion. First, we evaluated the data conversion methods (Table 5.1). Both methods were tested using the same input: the first method converts the descriptors into bitarrays (P_2), while the second method stores them in inverted dictionaries (P_3). When applying mixed pattern matching, the the melodic input and output remains unchanged whether it is paired with P_2 or P_3. In terms of melody, the dataset is stored as a single sa_{S_p} and bv_{S_p} for the melody, and bitarrays (P_2) or $\mathrm{D}_{\mathrm{inv}}$ descriptor dictionary (P_3) for the corresponding descriptors. The P_2 approach also creates and stores a bit reference table in addition to the two, which is used to convert the potential descriptor query to a suitable format. Although the differences between the two conversion methods for descriptors are minor, the evaluation favours the P_3 approach, which is both more time- and space-efficient.

Melodic Pattern Matching (P_1). As previously mentioned, there is only one method for melodic pattern matching (as described in Section 4.2). For the evaluation, the input melodic pattern query was set to $p = bag$. We searched for all instances of this sequence within the corpus of Slovenian folk songs and retrieved 362 instances. Additionally, we applied two parameters, that technically derive from the descriptors, considering only the sequences which start in the first position of a phrase, regardless of the phrase number, which filtered the results down to 20 instances. Both methods were evaluated through multiple iterations to determine the mean or peak performance, as shown in Table 5.3.

Descriptors Set Matching (P_2 & P_3). Next, we compare two types of descriptor pattern matching. The P_2 method allows the input query with the selected descriptors of time signature, phrase label, contour and verse syllables, to provide only a single descriptor element x_j^t per descriptor Δ^t. In this case, the query is composed of:

$$d = (3/4, A, \nearrow\searrow, 4)$$

In contrast, the P_3 method can match several descriptor elements d_j^t per descriptor category Δ^t. If we take the same descriptors as above, then the input query is:

$$md = \left(\left\{ \frac{3}{4} \right\}, \{A\}, \{\nearrow\searrow, \searrow\}, \{1,4\} \right)$$

While the query format for the method of P_3 fits as is, the one of P_2 requires referencing the pre-constructed dictionary of bits to convert our query to bitarray, before proceeding with the match. Additional time is required for each label combination we wish to test. If we wish to compare different descriptor elements with P_2, it must execute separate functions for each descriptor element combination (with always allowing one per descriptor). Due to conversion and smaller flexibility in terms of adding descriptors, as shown in Table 5.2, the method of P_2 is evaluated as being less efficient in terms of both time and memory. We also tested mixed pattern matching, adding melodic query and matching to both (Table 5.4).

Metric	Problem	Mean	Standard Deviation (±)
Time	P_2 **(Ref. Dict)**	167 µs	30.2 µs
	P_2 **(Dataset)**	271 ms	467 µs
	P_3	48.5 ms	1.1 ms
Peak Memory	P_2 **(Ref. Dict)**	225.95 MiB	-
	P_2 **(Dataset)**	227.12 MiB	-
	P_3	225.94 MiB	-

TABLE 5.1: Data conversion evaluation (7 runs, 10 loops each). P_3 involves a single step, whereas P_2 requires two distinct steps: first, creating a reference dictionary, and then converting the dataset (see Algorithm 2). These two steps should therefore be considered together for a comprehensive comparison.

Metric	Problem	Mean	Standard Deviation (±)
Time	P_2 **(Query Conversion)**	2.64 µs	615 ns
	P_2 **(Descriptor Matching)**	866 µs	16.8 µs
	P_3	674 µs	170 µs
Peak Memory	P_2 **(Query Conversion)**	227.97 MiB	-
	P_2 **(Descriptor Matching)**	227.97 MiB	-
	P_3	240.16 MiB	-

TABLE 5.2: Descriptor evaluation for P_2 and P_3 (7 runs, 10 loops each). P_3 is executed in a single step, while P_2 involves two steps: first, converting the query to bits (see Algorithm 2), followed by the descriptor set matching (see Algorithm 3). The two steps should be regarded together.

5.2 Case Study: Slovenian Folk Song Ballads

Statement: This case study is a slightly edited version that was published paper [32].

Problem	Metric	Mean	Standard Deviation (±)
Melodic Pattern Matching	**Time**	365 µs	40.4 ns
	Peak Memory	227.31 MiB	-
Filtering	**Time**	44.5 µs	328 µs
	Peak Memory	227.31 MiB	-

TABLE 5.3: Melodic pattern matching evaluation (7 runs, 10 loops each). In addition to melodic pattern matching, a filtering function was included and evaluated, which accounts for both the position of the pattern within a phrase and the phrase's position within the song.

Metric	Problem	Mean	Standard Deviation (±)
Time	P_2	914 µs	310 µs
	P_3	1.13 ms	414 µs
Peak Memory	P_2	192.67 MiB	-
	P_3	192.66 MiB	-

TABLE 5.4: Mixed pattern matching evaluation for P_2 and P_3 (7 runs, 10 loops each). The evaluations for each P are best understood in the previously provided evaluations (separately for melody and descriptors), which detail the steps more clearly.

We conducted the first case study using 103 monophonic variants of type 286 — *Nevesta detomorilka*, a song well-established in the European folk song tradition, particularly noted for its widely known theme of an infanticidal bride. A subtype of this song in our corpus, labelled 286.T1, consists of 34 tunes, which have been manually classified as a subtype due to their melodic similarity. These tunes frequently exhibit similar melodic patterns at the beginning or end of specific phrases, such as the *fad* as a middle phrase starting pattern or the *bag* as a final phrase ending pattern (Figure 5.2).

We developed combined melody and/or descriptor queries, which assumed the main features of the specific phrases of subtype 286.T1. These queries were then evaluated as a binary classification problem: for instance, can we accurately identify the 34 phrases of 286.T1 and distinguish them exclusively from other phrases based on melodic features and its descriptors?

Some notes on the dataset. This case study [32], was conducted using the first version of the algorithm and annotated dataset[9]. The primary differences from the current version are as follows:

1. the dataset was later slightly extended with additional phrases after re-exami-na-tion with experts from the Institute of Ethnomusicology ZRC SAZU; 2. the phrase labels of the dataset were simplified; 3. the dataset initially included harmonic labels for the beginning and ending segments of phrases, which were later omitted due to their ambiguity, as discussed in previous chapters; and 4. the algorithm for

[9] Algomus Data, http://www.algomus.fr/data, accessed on 4th October 2024.

descriptor pattern matching was originally based entirely on the method using bitwise operators, which precluded the comparison of multiple descriptor elements within a single descriptor category.

Despite these differences, the majority of the objectives that remain relevant today were already apparent in the methodology used at that time.

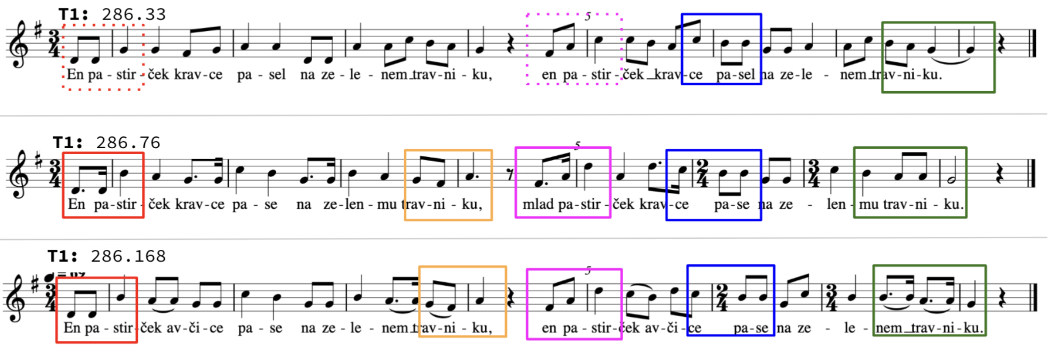


FIGURE 5.2: Three variants (out of 34) of the subtype 286.T1 with similar short melodic patterns in the same phrase positions (in colored squares).

5.2.1 Pattern Matching

Table 5.5 illustrates that simple melody queries consisting of 1 to 3 notes achieve reasonable recall rates (50%–80%), although they exhibit limited precision. By adding additional relevant descriptors, both precision and recall gradually improve, leading to higher F_1 scores.

The *ddb* melody query alone produces 93 matches: 18 are True Positives (TP), but 75 of them are False Positives (FP) unrelated to the first phrase of 286.T1 tunes. Incorporating a phrase position descriptor (F, first) refines the query, while adding relevant contour ($\nearrow$) and starting harmonic information (H_S) of a dominant (D) further enhances specificity. This query (F, $\nearrow$, H_SD) results in only 2 False Positives (FP), achieving a precision of up to 0.88, with minimal sensitivity loss. The *ag* pattern in the last phrase, characterised 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 starting harmonic label (H_S) of a tonic (T) as a stable harmony descriptor proves effective in this context.

However, including too many or irrelevant descriptors can lead to poor results. For instance, the *cbb* pattern is primarily found at the end of the middle phrase. Requiring a stable harmonic framework (ending harmonic label (H_E) as T) for middle phrase endings reduces precision, as it is less common in those positions (Table 5.6). Another interesting instance is the *fad* pattern, occurrences of which are almost evenly split into two contours. If we matched (*fad*, *M*, $\nearrow\searrow$ or $\nearrow\rightarrow$), we would obtain 23 True Positives with a precision of 0.82 and a recall of 0.82. Consequently, the descriptor pattern matching algorithm was, as anticipated, extended to accommodate the matching of a subset of multiple descriptors within the same category, rather than solely relying on one descriptor element per descriptor.

5.2.2 Patterns as Building Blocks

Melody and descriptor patterns have versatile applications beyond classification. In our case, the most effective queries incorporate position descriptors, suggesting that we should consider phrases and their positions in songs as *building blocks* for pattern matching.

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 musical material among tunes and vice versa. For instance, the *ag* pattern in the last phrase, exhibiting a ↗↘ contour and ending with H_E, 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 may provide insight into the extent to which these melodic ideas are unique to (Slovenian) folk songs.

5.3 Case Study: Children's Songs Collection

Youth magazine *Ciciban* had its first issue in 1945. Since then, it was a great deal of Slovenian children's lives. It includes didactic, entertaining and overall educational contents in all literary forms, as well as illustrations and music, and Q&A section, where children can learn more about what they are interested in. It issued works of many established cartoonists, youth writers and poets, as well as composers, and has been a great tool for parents, teachers and other educational workers. The magazine's primary goal is to provide high-quality content for aesthetic education, covering verbal, visual, and musical arts, while also supporting intellectual, physical, and social development [221, 307]. *Ciciban*, originally aimed at pre-school children and those up to ten years old, has shifted its focus over the last 15 years to cater mainly to younger students [307]. The project GMGM, among other tasks, digitised a corpus of 123 music examples[10] (mainly composed, arranged, or curated children's (folk) songs with lyrics) from the earliest issues (1948–1991) of *Ciciban* magazines. For the purpose of testing the algorithm proposed in this thesis, we took the entirety of the examples with existing annotations, as well as added some based on the available data.

[10] Slovenska mladinska in otroška glasba 1945–1991, https://korpusi.musiclab.si/, accessed on 4th October 2024.

Query (melody + descriptors)			**TP**	**FP**	**FN**	**Precision**	**Recall**	**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, H_SD		27	189	7	0.12	0.79	0.22
d	F, $\nearrow$, H_SD		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, H_SD		17	21	17	0.45	0.50	0.47
ddb	F, $\nearrow$, H_SD		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, H_ET		**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, $\nearrow\searrow$, H_ET		23	51	11	0.31	0.68	**0.43**

TABLE 5.5: This table is a part of analyses published in [32]. 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. The assessed descriptors are phrase position (F=first, M=middle, L=last), harmonic labels (Phrase starting on a H_SD=Dominant, or ending on a H_ET=tonic), and contours ($\nearrow$=ascending, $\nearrow\searrow$=convex, $\nearrow\rightarrow$=ascending horizontal).

		T	**D**	**?$_T$**	**?$_D$**	**?**
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 5.6: This table is part of the analysis published in [32]. The starting harmonic H_S and ending harmonic H_E 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 "?$_T$", "?$_D$", and "?", as they can be interpreted as either T or D, influenced by previous and following pitch values or bars, making the exact annotation spot unclear.

5.3.1 Corpus Structure

The GMGM corpus mainly includes music notation with (some) lyrics (see the example in Figure 5.3). The publicly available metadata is currently relatively scarce, making it challenging to extend queries with contextual or accompanying information. In most cases, the title and composer (if applicable) are annotated, along with music descriptors such as time signature, key, ambitus in semitones, number of measures, melodic sequence in MIDI and alphabetic pitch values, as well as numerical relative values describing the melody as an intervallic sequence. All of these descriptors were presumably automatically retrieved, as no manual annotators are specified.

Unlike the corpus of Slovenian folk song ballads, the GMGM corpus is not split into phrases. However, the provided online annotations do include information on rhythmic sequences and the beginning positions of each measure in relation to note (or rest) units for both rhythm and melody.

Due to the difficulty in navigating between breaks in lyrics and/or rhythm, the task of automatically segmenting the melodies was too challenging for our short case study. The only two units one could consider were individual measures or full songs, with the latter being chosen. We later added annotations on the maximum and minimum pitch values of the range, as well as interval mean and median, and re-grouped the information on melodic range and intervals as described below in the following subsection.

FIGURE 5.3: Two examples of the same song, *Zajček zobke brusi / The Rabbit is Sharpening His Teeth*, composed by the Slovenian composer Janez Bitenc and set to the lyrics of Neža Maurer from the youth magazine *Ciciban* (1961–62). (**Top**) A scanned example from the *Ciciban* magazine; (**Bottom**) a digitised transcription from the GMGM collection[288].

5.3.2 Defining a Children's Song: Melodic and Other Key Characteristics

According to Fran[11], a digital repository of Slovenian dictionaries and linguistic resources, the term *ciciban* specifically refers to a pre-school child, or less commonly, to children in the early grades of primary school. As highlighted in the introductory paragraph of this section, the content of the early issues of the magazine, including its musical materials, was primarily aimed exactly at pre-school children and young children up to the age of ten.

Consequently, the question arises: how can we determine whether the materials are genuinely aimed at the intended young audience, how to locate such songs in

[11] Fran, https://fran.si/133/sskj2-slovar-slovenskega-knjiznega-jezika-2/4464617/ciciban, accessed on 4th October 2024.

other repertoire, and how do (and if) these songs differ from those in other collections, such as the Slovenian folk song ballads? More specifically, to what extent can we distinguish the two corpora based on their melodic features. Unlike the questions addressed in the internal analysis of the primary corpus—such as fieldwork bias, tune families, melodic contours, phrase structures, and so forth—the music examples from *Ciciban* prompt an external and education-oriented enquiry. To determine what makes these songs suitable in any form of music education for children as opposed to other corpora, we focused on research in music pedagogy, particularly studies concerning children's vocal range and the intervallic structure of melodic lines. Additionally, we concentrated on studies commonly referenced or conducted in the Slovenian area.

While it is notable that studies on song repertoire in primary music education and earlier stages of a child's interaction with music consider more than just the melodic aspect of songs (such as lyrics, origin, popularity, rhythm, and more), the melodic range and its intervallic features are crucial in selecting songs manageable for young children to perform. In two studies [290, 140], a survey was conducted regarding the importance of various musical features for music education. The majority of teachers who participated in these surveys confirmed that melodic difficulty (in terms of intervallic leaps) and vocal range are important or very important factors (all but two participants in [140] responded this way, as did 94.7% of participants for range and 95.4% for melodic intervals in [290]). A slightly smaller proportion of respondents also considered rhythm to be important. The least important features were the selection of composers and song popularity.

Building on the research conducted by [290, 140], as well as the studies reviewed in one or both of these (for example, [175, 29]), and others (see the literature reviewed by [290, 140]), we propose two general criteria for classifying songs based on their suitability for children: vocal range and melodic structure, which form the basis for the classification task in our evaluation.

Melodic Range. By summarising several studies, we can consider that children up to ages 10-12 can generally be categorised into two groups. The first group consists of preschoolers and sometimes also includes first-grade students, typically aged 4-6, and the second starts around 7 and extends up to 12 year-olds. In general, people can produce up to an octave and a half when singing, varying with the respect of their musical training and physical development [290].

In terms of music perception, by the age of 5, children begin to recognise different pitch registers, and by age 6, they can discriminate between simpler melodic patterns. While intervals and range become gradually perceivable around the age of 6, it is not until around age 9 that the majority of non-musically trained children are capable of developing a sense of tonality and harmony [140].

There is some disagreement regarding the ideal vocal range for children of different ages, however, most sources define the range for younger children (up to 6 or

7 years old) as somewhere between C4 or D4 and A4 to C5. For children aged 8 to 12 years, the range is generally defined as somewhere between A3 to D4 and C5 to E5, and in fewer cases, up to G5 (Figure 5.4).

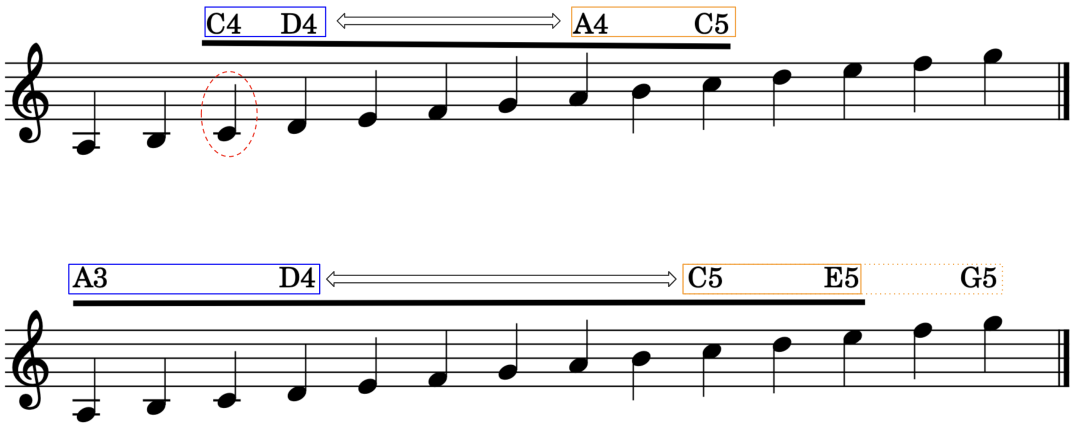


FIGURE 5.4: Approximate melodic ranges for two age groups of children, based on reviewed literature in Section 5.3.2. (**Top**) Melodic range for children up to 6 or 7 years old. While C4 is sometimes challenging, it is generally included in the preferred range for this age group. (**Bottom**) Melodic range for children aged 8 to 12 years old. Notes above E5 are seldom referenced in the literature for this age group.

A comprehensive summary of various studies on children's vocal range can be found in [290, 140], where a table summarising these ranges is provided on page 139 of [175]. The study summarised by Winnifred Adelmann [331, 140] suggests that the range for younger children should be between C4 and A4 (or G#4), noting that very high or very low tones are not ideal. This range is also specified to be appropriate for preschoolers in [184, 259, 287], while C4 to C5 is specified there as an appropriate range for children 6 to 8 years old in the first two, or up to 10 years old in the latter. The same range of C4 to C5 is also mentioned to be the most common for children from 5 to 6 years old in [168], stressing that most children struggle to reach below C4.

In contrast, Bogdana Borota [29] recommends a range of approximately D4 to B4 (or C5 for 6-year-olds) as preferable for younger children.

It is worth noting that while we can discuss an approximate vocal range, some researchers also emphasise the importance of tessitura, which refers to the most comfortable vocal range for the participants being observed. Although we did not explore this aspect further, it may be worth considering in future research.

Finally, it should be stressed that teachers in the (music) education process generally have the ability to transpose or adjust songs if necessary. However, the songs curated for children are intended to be, for the most part, appropriate without such adjustments.

Melodic Intervals. The mentioned studies and their reviewed literature support a preference for pentatonic, major, and minor scales in melodic structure. Before the age of 8, children do not consciously perceive intervals, although most can distinguish between different pitch heights a few years earlier. The perception of actual keys or tonalities typically does not develop until after the age of 8, thus the focus should be put on the relative values, meaning the consecutive intervals.

There are some discrepancies regarding which intervals are easier for children to sing. However, most agree that consonant intervals—particularly major and minor thirds, as well as potentially the perfect fourth, perfect fifth, and octave—should be preferred [290, 332, 140, 331]. The use of seconds is less clear. Although they are the smallest intervals in diatonic scales (besides unison), minor seconds are considered difficult to perform when several appear consecutively in a chromatic order. Outside of this context, as an isolated sequence of two pitches, major and minor seconds are generally considered the easiest intervals for younger children to sing [290, 332]. Easier melodies also tend to include two to three consecutively repeated tones, descend in pitch (especially from the dominant to tonic), use repetitive material, and avoid modulation [290, 332].

5.3.3 Melodic Filters

In this section, I will introduce four melodic filters, grouped into two categories: more restrictive and less restrictive, based on melodic interval jumps and range. These filters were constructed in reference to the literature reviewed in the previous section. The first set consists of highly restrictive filters, tailored to the smallest possible range and intervals appropriate for younger children, while also taking into account the limitations of the corpus. The second set of filters, less restrictive and still easily manageable by slightly older children, allows for a somewhat larger range and broader melodic intervals.

Melodic Filters Definition. I constructed two types of filters: range filters (hereafter, RF1 and RF2) to define the upper and lower pitch boundaries, and interval filters (hereafter, IF1 and IF2) to constrain the intervallic relationships between consecutive pitches in a melodic sequence.

In IF1, I excluded the minor second if it occurred more than twice in succession (i.e., before and after the observed interval) to avoid excessive chromatic movement. While this step could be refined by further developing the methodology for subsequent pitch relations, it remained as described here, given that this refinement was beyond the primary focus of the thesis.

More precisely (Figure 5.6), I defined the range filters (RF), and interval filters (IF), to help to classify whether songs are children-appropriate or not.

- RF1: pitches from D4 to A4.
- RF2: pitches from A3 to C5.

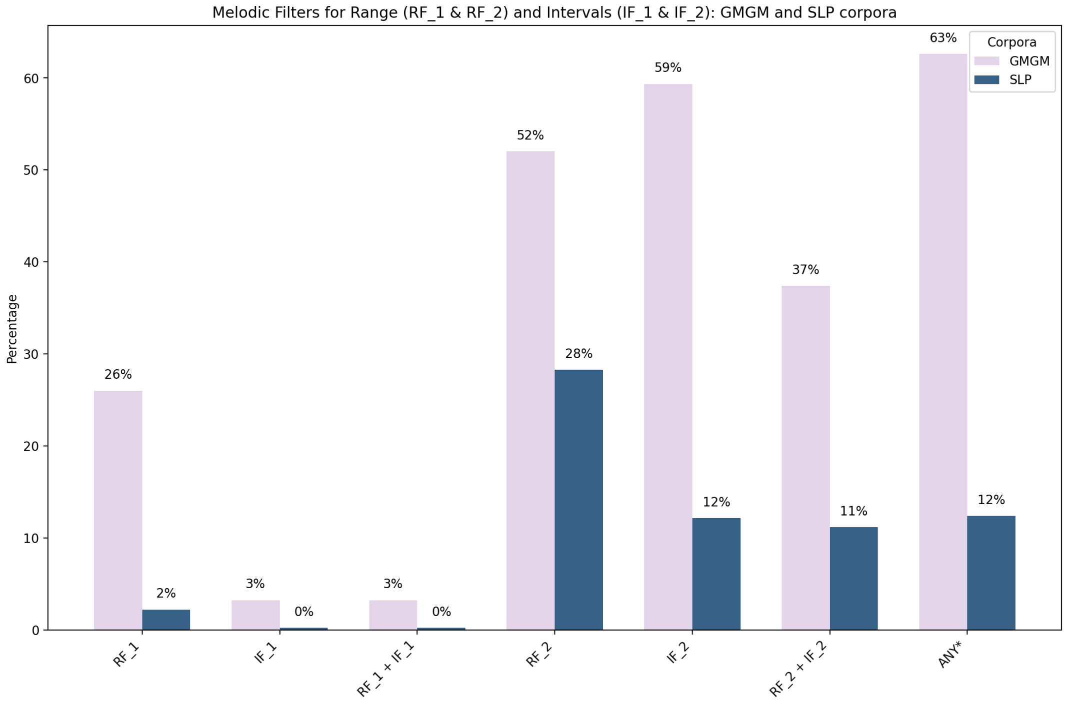


FIGURE 5.5: Statistics on 4 melodic filters applied in various combinations (RF1 and RF2 = range filter 1 and 2, IF1 and IF2 = intervals filter 1 and 2, ANY = at least one out of two per range and one per interval). We compared two corpora: GMGM (light purple bars, songs from the *Ciciban* youth magazine collection) and SLP (dark blue bars, a collection of Slovenian folk song ballads). The results indicate that the children-centered and curated corpus (GMGM) performed significantly better than the general folk song collection (SLP) when applying melodic restraints across all filtering combinations.

- IF1: the allowed intervallic jumps are minor second (m2) and major second (M2), minor third (m3) and major third (M3), and unison (P1).
- IF2: the intervallic jumps should be no larger than perfect fifth (P5).

Other filters. Although we did not address the tessitura for each age group, I incorporated an additional criterion: the overall range in terms of interval regardless of starting and ending pitch (as opposed to the RF, which sets the lower and upper boundaries of the melody, but does not set a strict distance between the two). This was done in consideration of studies suggesting that a smaller range of different pitches generally results in a less challenging song. Using these additional filter, we are able to identify songs that may have a slightly larger upper and/or lower range, but do not extend to, for example, more than an octave, fifth, or similar.

Lastly, we further support an analysis of melodic intervals, focusing on trigrams of consecutive melodic intervals as the basis for sequence pattern matching (as opposed to using absolute values, as in the first case study). Our goal here was to further support the preference for smaller interval sequences over larger jumps. While

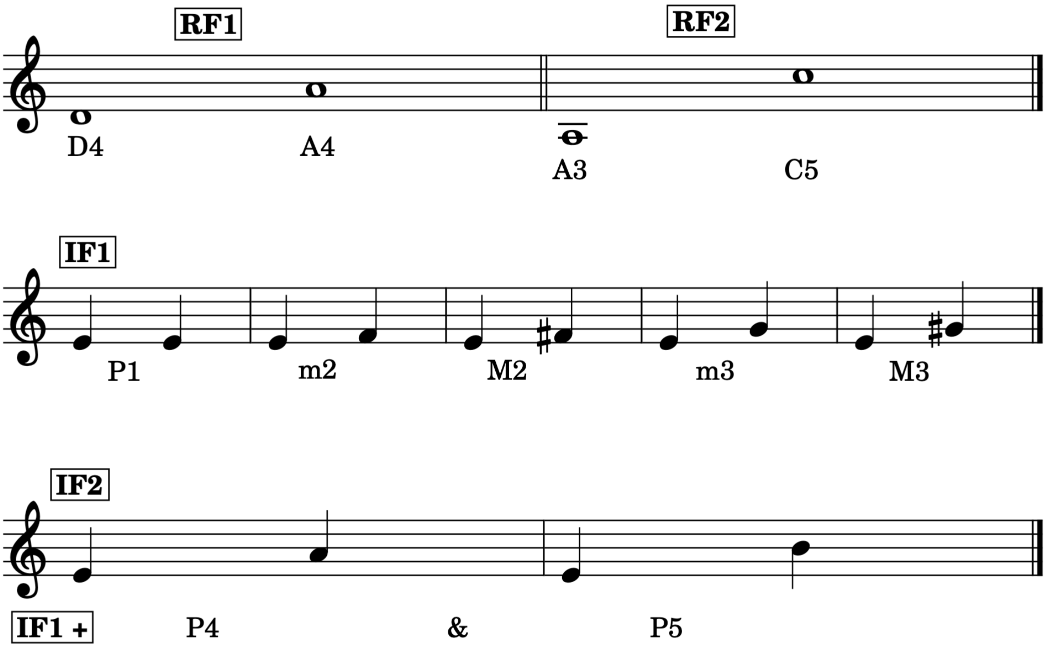


FIGURE 5.6: Visualisation of all four melodic filters, with (**Upper Staff**) RF1 extending from D4 to A4, RF2 extending from A3 to C5, (**Middle Staff**) IF1 allowing subsequent melodic intervals from minor second (m2) (limited repetition to avoid highly chromatic melodies, see Section 5.3.2) to major third (M3), (**Lower Staff**) and IF2 incorporating all intervals of IF1 plus perfect fourth (P4) and perfect fifth (P5).

we tested various sequences, we decided to discuss on only the smallest possible sequences in the evaluation, specifically those consisting of unison (P1), minor second (m2), and major second (M2).

5.3.4 Results and Evaluation

Statistics. We constructed the filter boundaries as previously described and applied them to the existing descriptors of the two corpora. Songs that matched any of the filters were assigned new descriptors based on the corresponding filter groups. All four filters were applied as such to every song, including those from the Slovenian folk song ballads. We then analysed the statistics for both corpora (see Figure 5.5).

The figure shows that more than half of the GMGM songs (63%) comply with RF2 (of which some also fit with RF1), and with IF2 (of which some fit with IF1 as well) in any combination (see ANY in Figure 5.5), whereas this is true for only 12% of the SLP songs. More than half of the GMGM songs comply with either RF2 (52%) or IF2 (59%). Although RF1 is very restrictive, about 26% of the GMGM songs, but only about 2% of the SLP songs comply. Applying IF1 (as well as that combined with RF1) managed to identify only 3% of the GMGM songs and 0.25% of the SLP songs. In all categories, the GMGM corpus better complies with these filters than the SLP corpus, despite the latter, overall, being a collection of relatively simple songs.

RF	IF	Range (MI)	Melody	TP (GMGM)	FP (SLP)	FN (GMGM)	TN (SLP)	Precision	Recall	F_1
X	X	X	P1-P1-P1	119	83	4	320	0.59	0.97	**0.73**
1	X	X	X	32	9	91	394	0.78	0.26	0.39
1	1	X	X	4	1	119	402	0.80	0.03	0.06
1	1, 2	X	X	31	5	92	398	0.86	0.25	0.39
1, 2	1, 2	X	X	77	48	46	355	0.62	0.63	**0.62**
1, 2	1, 2	X	P1-P1-P1	65	11	58	392	0.86	0.53	**0.65**
1, 2	1, 2	X	P1-P1-M2	77	23	46	380	0.77	0.63	**0.69**
1, 2	1, 2	X	P1-M2-M2	103	32	20	371	0.76	0.84	**0.80**
1, 2	1, 2	X	M2-M2-M2	107	91	16	312	0.54	0.87	**0.67**
1, 2	1, 2	X	m2-m2-m2	4	3	119	400	0.57	0.03	**0.06**
1, 2	1, 2	FIFTH OR SIXTH	P1-M2-M2	76	16	47	387	0.83	0.62	**0.71**
X	X	ABOVE EIGHT	X	21	63	102	340	0.25	0.17	0.20

TABLE 5.7: Evaluation of case study 2, where we calculate True Positives (TP), False Positives (FP), False Negatives (FN), True Negatives (TN), precision, recall, and F_1 scores for various combinations of descriptor queries, incorporating two melodic filters for minimum and maximum range (RF: 1 and 2), and melodic intervals (IF: 1 and 2), with additional categories of range expressed in melodic interval (MI), at times combined with short melodic patterns, converted to relative values of melodic intervals between each subsequent tone. The focus is on identifying songs deemed suitable for younger children, as hypothesised in Section 5.3. The primary corpus under investigation is the GMGM corpus, a collection of children's songs sourced from *Ciciban* magazines, which is compared against the SLP corpus (*Slovenian folk song ballads*, see Chapter 3, and Sections 3.2.1 and 5.1).

Results. As noted in the statistical breakdown, the GMGM corpus exhibits a better suited match for the pre-set filters across all predefined parameters, including filters for maximum and minimum range, and intervallic melodic movement.

Table 5.7 presents different sets of queries based on the introduced filters and issues discussed earlier. Unlike the SLP example where we used single descriptor set matching (Section 4.3.2), we applied the multidescriptor sections matching method (Section 4.3.3) and performed melodic pattern matching or mixed pattern matching (Section 4.4) using relative values rather than absolute values for the observed sequences.

I looked at this matter as a classification problem. I tried to classify songs between GMGM and SLP, and see if the 4 pre-set filters (RF1, RF2, IF1, and IF2) are sufficient conditions for classification of the two corpora or not.

We assumed both, GMGM corpus of 123 songs (True Positives, False Negatives), and 402 SLP songs (False Positives, True Negatives) as our "ground truth". In addition to the 4 filters for minimum/maximum range (in pitch) and maximum interval jumps, the classification task also considered range (expressed as an interval between the lowest and highest pitch value, regardless of that value) and short melodic patterns (as interval sequences) (Table 5.7).

First, we applied a sequence of unisons, which resulted in one of the highest F_1 scores (0.73) in our evaluation, retrieving almost all (119 out of 123) GMGM examples. Next, we tested the RF1, and then added IF1. The latter performed particularly poorly with the F_1 score falling to 0.06. As shown in the statistics, it is very rare to find songs with such strict restrictions in either collection. When allowing either of the two filters (applying a query of (P_3) as $md = \{\mathsf{IF1}, \mathsf{IF2}\}, \{\mathsf{RF1}, \mathsf{RF2}\})$, TP increases, as many more GMGM songs are matched; however, this also results in the inclusion of a number of SLP songs, lowering the F_1 score to 0.62 (see Figure 5.5: ANY, and Table 5.7: 5th row).

By separately incorporating melodic sequences of three unisons (P1-P1-P1), two unisons and a major second (P1-P1-M2), one unison and two major seconds (P1-M2-M2), and three major seconds (M2-M2-M2), both precision and recall increased to the point where we could predict the most true positives and the fewest false positives and negatives. This resulted in the highest F_1 score of 0.8 when including the melodic sequence P1-M2-M2. We then took this melodic sequence and added the additional interval range filter of a fifth and a sixth, which also performed well, achieving an F_1 score of 0.71.

I confirmed that chromatic movement of the melody (m2-m2-m2...) is very unlikely to be considered attractive or simple for singing, as both corpora performed poorly, resulting in the lowest F_1 score of 0.06. Additionally, we confirmed that a range (Range (MI)) above an octave is very unlikely to be linked to children's songs, with an F_1 score of 0.2.

5.3.5 Conclusion

While the descriptor set matching with multiple descriptors was sufficient to demonstrate that the GMGM corpus aligns better with the constraints of children-appropriate songs, incorporating mixed pattern matching with relative melodic sequences enhanced our understanding of the differences between the corpora, thereby improving the evaluation of classification task.

However, we undertook a rather challenging task, as both corpora are intended for non-professional singers, resulting in many similarities, particularly in terms of melodic limitations[12]. We acknowledge that these parameters alone are insufficient to fully determine the appropriateness of songs for children. Future research could explore identifying specific repetitive patterns in melody or rhythm that are used to teach different pitch heights, keys, or rhythmic patterns, as well as observing the tune length and structure, and examining the topics and structures of the lyrics.

Furthermore, extracting additional descriptors from the provided musical elements could allow for a more comprehensive analysis of the songs. Better documentation of the songs, along with real-life feedback, such as children's and teachers' responses to the material, would help address this issue more thoroughly. If we had access to more information or related corpora at the time of this study, we could have compared different children's songbooks or examined the same collection across different eras, among other possibilities.

Overall, despite evaluating two corpora with such similar material, the GMGM corpus proved to be a relatively good example of how to compose appropriate children's songs. The corpus itself has proven useful and could potentially serve as

[12]This presumably shows the influence and invaluable role of data curation. In the case of SLP, it aimed to preserve musical heritage, while GMGM's curation focused on carefully selecting songs for children's music education and entertainment.

valuable training data for various machine learning and AI tasks related to children's music education. Given its strong alignment with the melodic scope commonly found in music pedagogy research, it could also be effectively utilised for further educational purposes.

5.4 Discussion

We tested the algorithms presented in Chapter 4 on two distinct corpora. Initially, we tested the melodic pattern matching and descriptor set matching with bitwise operator, by applying the melodic and descriptor pattern matching techniques to Slovenian folk song ballads. Our findings indicate that employing slightly longer melodic patterns and incorporating additional descriptors enhanced precision, particularly when attempting to distinguish between different tune types within and across various folk song categories. We concluded that, although descriptor and melodic patterns can be individually identified in different songs, the position of our fundamental structural parts, meaning phrases, plays a pivotal role in recognising the overall pattern of a song. This is especially evident in relation to contour sequences and the harmonic function patterns at the beginning and end of phrases.

Secondly, we posed a distinctly different research question, grounded in pedagogical, didactic, and educational research on music perception and performance in young children. We focused on the most common features identified in these studies, particularly those related to melody. Our investigation aimed to determine whether the GMGM corpus, which presumably comprises only songs intended for children, aligns with these criteria. Furthermore, we sought to establish whether these standards could be applied to identify suitable songs in other corpora, such as Slovenian folk song ballads. Although we acknowledge that appropriate songs for children encompass other musical parameters, our findings confirm that the GMGM corpus adheres to the extended criteria for essential children's songs, particularly regarding melodic structure and related descriptors. Moreover, it can be seen as a solid foundation for identifying similar features in more general corpora, or used to automatically generate suitable music examples for children. Finally, by utilising a corpus not directly related to our primary research focus, we demonstrated the versatility of the proposed pattern matching methodology.

In both cases, we showed the intention of including as much information on and about music as possible, and hence, joined the research that advocates for extended music studies. While it is interesting and insightful to observe music notation, the format itself has many pitfalls and shortcomings, and with it, the corresponding methodology for its exploration. In order to evaluate these algorithms, neither time-space performance statistics nor focused case studies suffice. We acknowledge that first, these algorithms, in its current form, cannot be autonomously utilised by any researcher with modest to none computational skills. We also understand the shortcoming of the algorithms focusing on only notational parameter of the songs, completely bypassing audio-visual aspects that complements it.

Finally, it is worth noting that, although we aimed to address reasonable and insightful research questions, there remains significant potential in incorporating additional contextual information both, in the scope of existing methodology, and as its possible future methodological extension. This would enable the algorithm to consider a more nuanced socio-historical context, including the usage, transcription, and archiving of the materials, as well as the discourse surrounding these songs, their perceived value (thus, the socio-economic angle) and the roles of and relationships between performers and the audience.

To our knowledge, no such comprehensive and inclusive analysis has been undertaken in the field of music pattern matching or music research more broadly. Therefore, this evaluation should serve as an invitation to move beyond the confines of music feature pattern matching, notation, or isolated parameters (rhythm, melody, harmony, and similar). Addressing these broader aspects may open the door to more robust theoretical frameworks, leading to clearer insights into the future direction and purpose of both MIR and musicological studies. In the following and concluding chapter, we will expand on the latter, touching the indispensible dynamics between the two fields, problematising and evaluating MIR's position in musicology and vice versa.

Chapter 6

Discussion: MIR and (Ethno)musicology

Statement: This chapter extends on thoughts developed in two published contributions [30] and [31].

The thesis provided detailed review of knowledge and methodological progress at the nexus of computation and (ethno)musicology as well as introduced a computational approach for music pattern analysis. In order to understand its place in the MIR community, I dedicated some time to observe the relationships between the two sides in interaction, meaning (ethno)musicology and MIR. The findings were discussed in two published contributions, first the introduction and editing of a cross-disciplinary special issue of Musicological Annual, centred around the topics of computational musicology [31], and in conjunction with other authors, the paper published at ISMIR 2023 [30].

The latter explored the dynamics between MIR and musicology through a case study, which focused on the accessibility and usability of MIR productions for musicological research. It involved an analysis of ten years' worth of papers from the International Society for Music Information Retrieval (ISMIR) spanning from 2012 to 2021. More than 1000 citations of ISMIR papers were reviewed, revealing that only 51 subsequent works published in musicological venues drew from the findings of 28 ISMIR papers. The final results indicated that most contributions from ISMIR rarely reach musicology or humanities.

Since then, the topic was once again extended in order to try to identify and address the obstacles that lead to discordance between these closely related disciplines.

6.1 Introduction

The emergence of new technologies inevitably brings about changes. This trend is not only prevalent in modern times but has been apparent since the inception of any form of tools. The first significant change, marking a fundamental difference between humans and animals, occurred with Homo sapiens, who was the first to fully transform the core technique of human motor behavior: tools ceased to be directly attached to the human body, instead introducing a detachment that created a discontinuity between body parts and tools (such as clothing, pliers, hammers, and so on) [173]. Advancing further, the detached tool facilitated detached movement, meaning that physical actions to execute a task were replaced by the tools themselves (for instance, reaching the target using a bow instead of the hand). Eventually, detached machines and memory emerged (e.g., clocks, and later, computers). And lastly, with the emergence of AI, the borders between human and technology shifted again, which prompted numerous physical and cognitive adaptations.

New tools lead to individual adaptations, and have induced numerous changes in social (or socio-economical) structures. For instance, the invention of writing and broader literacy partially transitioned reliance from faith and the wisdom of the local elder townsmen to verifiable facts and science [97]. Similarly, the gradual decentralization of the human workforce in production processes can be seen as machines extending beyond human capabilities, rather than the opposite [277].

(Music) research. In the field of research, the introduction of new tools often stimulates increased scholarly activity and the expansion of potentialities [277]. Such developments can result in the establishment of new disciplines, the emergence of novel fields of study, the replacement of existing disciplines, or, at the very least, a significant transformation of their methodologies. Music studies have experienced precisely this trajectory, evolving from the creation of musical instruments and notational systems to the advancement of recording technologies and digital tools. Each of these innovations has required careful consideration and has prompted the adoption of new methodological approaches, both within the discipline and in interdisciplinary contexts.

Although it is evident that MIR and (ethno)musicology do not consistently share the same perspectives, the idea of "new" methodologies superseding "traditional" ones predates the development of computationally aided projects and development in, what we perceive at this moment, "newer" tools.

In the recent history of music research, there was first resistance to recording devices and new collection practices (Section 3.2.1), and methodologically-speaking as well as to empirically-centred approaches, which were mostly in the domain of systematic musicology (and later, ethnomusicology) (for a detailed review, refer to [230]).

Expanding on the provocative inquiry "Who stole systematic musicology?", Leman observed that systematic musicology had (already) lost control over its own identity [170]. He highlighted that the early conflicts within the various branches of musicology had evolved. Now, these branches combined all share similar tensions with, once again, newer concepts put forward by MIR and related research, as well as neuroscientists and psychologists. These disciplines have increasingly embraced the empirical study of music, further complicating the landscape of musicological research.

The advent of empirical research methods, growing interest in music as a complex phenomenon, and technological advancements have broadened the scope of music research to include various academic disciplines. This expansion has led to significant changes in the structures of these fields, impacting and encourage the development of new areas such as MIR, evolutionary (ethno)musicology, music psychology, popular music research, and the anthropology of music, among others.

While skepticism persists among some regarding the unconditional embrace of emerging collaboratory changes, for others, technologies and related methodological shifts are seen as a natural extension for quantitative, big-data, and empirical music analyses [57]. Music research began to be conceptualised as an interdisciplinary domain comprising "somewhat equal" sub-disciplines, including musicology and MIR. Discussing these collaborations, [217, 311], many underscored the advantages of multidisciplinary projects in broadening the boundaries of isolated disciplines to achieve more comprehensive outcomes.

In contrast, others warned that in "an era in which interdisciplinarity has become a kind of mantra, verbally subscribed to by nearly everyone, disciplines continue to police their own boundaries [17]." Similarly, [170] and [230] highlighted that although they valued the discussions and involvement in interdisciplinary projects, it remained rare for researchers to step outside the boundaries of their own disciplines. Furthermore, the transfer of knowledge is far from seamless between computational scientists and musicologists (as well as other disciplines), resulting in ideas expanding inadequately, if at all [269]. Skepticism continued regarding the unconditional acceptance of the emerging collaborative changes, and thus, recognising the need for improvements in knowledge transfer was deemed imperative [266].

In 2005, [57] critically addressed the prospects and challenges of collaborations between MIR and musicology, many of which were revisited by [70]. The latter further examined the implications of new technologies and reflected on the interdisciplinary dynamics during the first 10 years of ISMIR, the principal MIR community. He highlighted its shortcomings, such as the inability to communicate the produced tools to the user (performer, musicologists, ...), favoring low-level over high-level features and audio over other symbolic music representations, and so forth. These and other unresolved issues persistently hinder collaboration to this day [57], partially due to the inability to pinpoint the theoretical, methodological, and institutional origins of the persistent miscommunication among all involved parties.

To tackle the most persistent issues encountered in knowledge transfer between disciplines, at least three topics need to be considered: the incorporation of and relationship to *new media and/or technologies*, as well as the foundations, structure, and *theoretical concerns* behind the *notion of disciplines and institutions*.

6.2 The Influence of New Media and Technology

In recent years, a growing number of musicologists, along with humanities research -ers in general, have shown a preference for working with digital materials over physical ones [72]. However, the application of computation to their music research is diverse. There are general-purpose software tools, such as word processors or spreadsheet editors, and music-oriented software, such as Sibelius, Finale, and Audacity. Additionally, there are programming music/MIR platforms and libraries, such as Humdrum [119], music21 [62], Librosa [198], and Essentia [27]. Furthermore, there are methods and algorithms developed by the MIR community, as exemplified by [164, 207, 43, 82], among others (see [266] for a detailed review). None of these are clearly distinct from one another as well as in most cases, including our own project[1], combine more than one of the above. While computer usage is prevalent among most researchers, there are fewer musicologists, as well as humanities and social science scholars, who adopt or contribute to the "making" of computational methodologies, one of those being the project of this thesis, introduced in the previous chapters. With the rise of new media and computational advancements, both music and our relationship to it underwent significant transformations [188]. Given the expanding scope of what is considered significant in the realm of music, it raises the question of whether familiarity with computational languages, or at least the ability to understand the concepts of contemporary computational processes, is becoming essential for research endeavors.

The methodological visions of "fundamentally-renewed" music research, according to [302], had "not [yet] taken over the majority of musicological scholarship." Nonetheless, the methodology of research has already shifted, as there is an ascending trend of new research tools and digitised (music) data representations, a lot of them consciously used by musicologists, and if not them, music researchers from other (emerging) domains.

Computational methods assist researchers in handling larger and more varied datasets, but, would musicologists agree that "working with [these] datasets [have] open[ed] up new areas of musicology [57]?" Or, has this shift made the "common" musicological concerns obsolete and evoked new areas of research, which are now (almost) independent from the musicological domain? Who are the tools made for and what are the origins of resistance from these new inventions?

[1] We used excel sheets to make the first annotations, pdf readers to acquire all information on the dataset, MuseScore to edit the song examples, music21 to automatically extract some music descriptors, and created our own set of algorithms for pattern matching tasks.

6.2.1 The Surface

In the process of epochal technological change, there has never been a sudden replacement of old technology for a brand-new solution, nor was this technology initially considered an extension of human capabilities, but rather as a separate "entity". The adoption of new technologies thus typically occurred through a gradual process of transition rather than an immediate change of medium. This process, in a computational sense, took advantage of "the known," thus familiar images of gadgets or encouraged the usage through the "instinctive" features that resembled physical actions of an old apparatus, a predecessor of the new medium. These types of mediators are employed to facilitate the understanding of the basic functionality of the digital tool, even though the underlying mechanisms of it may be entirely incomprehensible to an average user.

In contrast to the leisurely (or similar) use of digital gadgets, the computational analysis and research (of music) has undergone a much less gradual leap. The background the of most computationally analytical (music) tools are driven by adapted mathematical operations woven into a web of algorithms and ciphers that have little, if anything at all in common with the "older", more familiar symbolic representation of an object or the process of (physical, music) analysis. The abrupt transition from the physical world to the realm of incomprehensible code has made the process of integration more challenging for (music) researchers with limited computational skills[2].

Instead of experiencing a transitional moment, the musicologist is confronted with abstract algorithms and their ready-made products, such as sequences of "mathematical" symbols describing or representing musical phenomena. For the majority of researchers who cannot or choose not to adopt computational processes like machine learning, neural networks, or even simpler algorithmic structures, inadequate familiarity with methodological procedures hinders their ability to utilise the results obtained from such analyses. Simply put, they may not understand how or where these solutions emerged from. Hence, since most new technologies (or languages) for music analysis "skipped" the transitional era, they are often incomprehensible or non-intuitive for the average musicologist. As a result, these computational products fail to communicate effectively with them.

There seems to be a "clear disconnect between how MIR tasks are designed to evaluate systems, and how end users are supposed to use those systems [...] [making them] difficult and costly to implement [266]". Consequently, the results, produced by such processes also become unusable, as the "involvement in the wheel of algorithms is indispensable for musicological research [188]."

I have introduced several projects that deemed to overcome this loop alongside

[2] For example, while the introduced methods for this thesis' project remain in the scope of this "sudden" transition, the platforms, such as Dezrann try to facilitate the severity of changes or mediate between the older and newer medium.

of this thesis. However, it still holds that if such obfuscation fails to deter the average user of an algorithmically guided interface, such as the automatic retouching of images with smartphones or the current trend of casually creating iconographic images with linguistic markers, active involvement in the inner workings of algorithms becomes indispensable for the research work of the musicologist (and others). The urge to engage in the latter primarily stems from functional reasons, whereas participation in everyday digital activities is largely determined by the desire to freely control the outcome of our interactions, rendering a thorough understanding of the underlying mechanisms irrelevant as long as one feels engaged.

The "surface", according to the upper discussion and surveys like [122], which addressed the relationship between "regular researchers" and the emergence of new technologies, suggests, that "[...] efforts should be made into supporting the development of their digital skills and providing usable, useful and reliable software created with a "musicology-centred" design approach." Otherwise, the "data richness will lead to information overload [63]." As [63] expressed in 2012, there are many tools for music collection and analysis, of which many "suffer from various shortcomings, such as specificity to a certain repertoire or approach, lack of robustness and flexibility, flawed user interfaces, or output is difficult to interpret."

A similar concern has been expressed by others, such as [148] and [195], or, for textual analysis [67]. All of them advocate not only for a more *accessible and flexible computational methods*, but also express the need to understand *what these methods do and how*. Alongside epistemological confusion and other (methodological) drawbacks, a similar problem was stressed by Aucouturier and Bigand. Their dialogue-style paper also revealed (among others) similar flaws and introduced some prospects for collaborations between MIR and music research (specifically music cognition) [10][3].

6.2.2 The Core

In the process of transductive ergomimesis, as defined by Magnusson, "new digital media drastically reposition the people" [188] and repeatedly evoke new (motor) skills and techniques, professions, and multidisciplinary actions (see also [135]). "What new instruments translate from earlier technologies are not simply the simulation of an interface, but a whole constellation of embodied contexts, where trained movements, musical actions, human-instrument relationships and other processes are translated to a technology of a different material substratum (from the organic to the digital material) [188]." The challenges of contemporary music research are to understand these changes while simultaneously undergoing similar transformation.

[3]The process from digitising to analysing Slovenian folk songs tried to remedy the "distance" between MIR and ethnomusicology such that it, more or less successfully, pursued the collaboration with various different profiles of researchers and their institutions.

This section has already highlighted that new media technologies continuously influence our development, both as individuals and as a collective community, culture, society, and beyond. Insights from works such as [97, 173, 135, 191, 211, 272, 277, 189], along with discussions of the prominence, evolution, and challenges of "new technology" in historical narratives and both past and contemporary popular culture, emphasise the evolving nature of our attitudes and relationships towards technologies that were once considered novel. It is often overlooked that certain everyday technological inventions of the past were, at one point, groundbreaking innovations. While some segments of society, driven by diverse motivations, readily embrace "the new", others maintain a sense of hesitancy.

While it may appear that computers, and more recently AI, have led to the most significant impacts, and although the degree of automation introduced by AI does set the current changes apart from those of previous eras, feelings of "technophobia," "technonostalgia," or similar resistance towards new technological ideas have been common in nearly every historical period (for example, the resistance against recording devices in earlier decades of 20th century of ethnomusicological fieldwork in Slovenia, as noted in Chapter 3.2.1).

Hence, the discontinuity between fields such as MIR and musicology, potentially caused by new technologies[4], is not solely due to the inability to utilise the tools themselves, but also the incapacity to contextualise one's position in relation to these technological advances. This often entails abandoning "traditional" and well-established methods, a cost that some are not immediately willing to incur.

Contrary to those who work with or directly create new media, the dissonance between the two groups may also lie in understanding what constitutes "new" technologies, as this definition is far from universal across different social groups. This uneven understanding of the subject can immediately (and partially) hinder fluent knowledge transmission.

A previously mentioned survey review [267] describes that MIR community's main concern is to create ("objectively") powerful tools, reach the user (leisurely listener), or have more datasets, but fail to properly address the reproducibility of their research. Thus, the next issue that needs to be (re)considered is no longer the mere technological shift, but the functionality of fundamental, theoretical ideas behind the new-established sphere of all (computational) music matters.

6.3 Theoretical Concerns

Musicological study frequently focuses on the internal structure of music, analysing aspects such as harmony, melody, rhythm, form, and instrumentation. However, it

[4]For instance, a notable and one of the oldest documented instances of apprehension towards technology can be found in a dialogue between Socrates and Phaedrus, where the former expresses genuine concerns about the written word compared to the spoken word, and the potential negative consequences it may have on one's memory abilities [106].

sometimes overlooks external factors like socio-economic and historical context, cultural influences, or performance practices, which allows for a thorough exploration of music's intricacies, meanings, and relationships.

New computational methodologies are gradually encouraging the the scope of musicological studies to move beyond traditional music analysis theories. These methodologies incorporate assumptions from disciplines such as acoustics, mathematics, and iconography, as well as subfields of computer science and others, alongside the foundational elements of music theory.

In response to Downie's call for improvements [70], some authors discussed *refined measurements* that need to be considered regarding data collection and interpretation, for "obtaining or accessing *high-quality datasets* remains a serious hurdle, especially on a large scale" [239]. These hurdles limit the (digital) quality of music research, but not only that. All music cannot be collected and/or represented in the same manner, and it is not feasible to investigate and discuss all research queries within identical methodological frameworks [195, 212]. They believe that this issue should be considered not only by musicologists, but should also be of equal importance for the field of MIR. Schüler and Huron argued that mutual *theoretical awareness* is essential for musicologists and MIR researchers [121, 269]. Due to the imposed importance of theory *and* "practice" (one of the key components of MIR), some suggest, that there must exist a *cyclical collaboration* between the disciplines [148].

In fields such as MIR, research questions are often addressed using methodologies rooted in the natural sciences, treating the data as "objective." However, scholars in the humanities frequently raise concerns about this approach, particularly when it comes to studying "human" or "societal" aspects such as music. There is a prevailing fear that this type of research can lead to detached results and interpretations, or a lack thereof, and tends to prioritise "facts" and their assessments, which are often equated with algorithmic success [67, 195] (see the introductory paragraphs of Chapter 5 discussing evaluation practices in MIR). Similar arguments, regarding theoretical concerns of what "matters" in music research have, however, already been used long before the emergence of MIR.

The theoretical debates concerning the fundamental aspects of music research hence predate the existence of MIR. Within the realm of musicology and related disciplines, ongoing discussions have centered around inquiries into the essence of musical significance, the contextual influences shaping its interpretation, the subjective nature of meaning attribution, and the methodological frameworks employed. These deliberations underscore the interdisciplinary character of music scholarship, which integrates insights from diverse fields such as philosophy, sociology, anthropology, and psychology to deepen our comprehension of music's role within human culture and experience.

Apart from the already mentioned disputes among different subfields of musicology, it is worth noting that the disagreements also occur on the level of music theory versus musicology, as described in, for example, *Does Music Theory Need Musicology?* [4], or *How We Got Into Analysis, and How to Get Out* [132]. Therefore, computational processes must take into account the ongoing debates within music scholarship and integrate these disputes into their research methodologies. Depending solely on prior music scholarship is inadequate, given that knowledge progression and theory development within this domain does not follow a linear ("proof-based") path.

6.3.1 The Transparency of Algorithmic Tools

When addressing possible dispute resolutions, many authors emphasise the necessity for numerous scholars to actively participate in shaping the framework and methodologies. This involvement requires a high degree of algorithmic transparency. "[I]n the long run, the most "useful" computational analyses will be the ones that are interactive, confronting a human user with the results of computational analysis and allowing that user to modify or intervene in the procedure to arrive at an acceptable or interesting result [195]."

Although authors commonly stress the need for a more transparent "background" of the methodology, the computational methods rarely consist of "universally applicable" models and thus remain very limited in certain areas - such as the analysis of vertical musical structures in relation to horizontal ones, the lack of a strategy for exploring non-Western or non-tempered musical notations (and recordings), music theory versus interpretation with societal factors and so on. Additionally, with the uneven distribution of resources, the digitisation of music material is disproportionately more common in Western institutions, while musical traditions with smaller coverage remain both physically and methodologically less represented in the (computer-supported) research landscape. The collection and accessibility of diverse musical material primarily depends on the direction of financial and infrastructural resources, which are often over-determined by its "market" value and/or attractiveness[5].

If it holds true that "the humanities are not a mere afterthought, simply studying and critiquing the effects of computational methods. [Their theory] can provide ways of thinking differently", as implied by Drucker [71] and similarly by Morreale [210], then the question arises: which specific existing theories and ideas, if any, play a role in either uniting or dividing these fields? Given the inherent tensions within these domains, what contributions can they offer to the emerging interdisciplinary field and how? Can these matters be communicated by resolving the transparency and representation of digitised resources and algorithms alone?

[5]For example, all of the materials used in this thesis had to be digitised, modeled and annotated as part of self-initiated first (pre-)step of the thesis and had not (and probably would not) exist in such format had it not been for this particular project and its flexible funding.

6.3.2 Theoretical Foundations

Huron suggested that there should exist the obligation for both parties (MIR and musicology) to familiarise themselves with each other's methodologies [121]. Additionally, [10] highlighted the importance of knowing which parts of whose methodology are to be used for a fruitful collaboration. However, [170] sees the "failure" of collaboration in the notion of the absence of "concrete planned goal at long term, except some vague idea of what all these research activities are up to [170]." He suggests solving the gap by inducing multi-modality, introducing context-based approaches into empiricism [170]. A more reserved argument in [230] implies that the wall is set by the feeling of superiority on both sides [230].

Although I am not aware of any concrete theoretical solutions in MIR, some humanities scholars [5, 67, 209] have begun to explore "computationally-inspired" alternatives to data handling, suggesting that the theoretical frameworks that preceded computational advancements may be insufficient, and that new methodologies should be leveraged to shape new theoretical perspectives[6].

What authors seem to disagree on is whether the objective should involve a gradual collaborative effort to comprehend and integrate (the entirety of) each other's theoretical concepts, and computationally replicate the ideas of musicological history (whichever these may be), or, instead to focus on establishing new theoretical foundations. If choosing the latter, the new foundations would need to facilitate a more comprehensive approach to researching the phenomenon of "music" within a contemporary, technologically enriched environment, rather than relying on vague notions [170] and momentary collaborative inspiration.

In our previous contribution [30], we discussed on the collaboratory struggles through theory of games or play by Johan Huizinga [118], and then Roger Caillois [41], in which we conclude that games (or science) can only be played when all parties are in agreement with the particular rules, and these rules are mainly composed of institutional organisation and theoretical ideas that support it. MIR is indeed a multidisciplinary environment, however, most of the participants (deriving from natural rather than humanities or social sciences), already play by similar rules (or speak the same language, or form their posture towards musical matters within a similar discourse practice).

As a result, the multidisciplinary activity within MIR continues to be relatively restrictive, and despite surveys [266], it has not adequately addressed all of the constraints mentioned by Stephen Downie about 15 years ago [70]. Furthermore, there has been a lack of genuine initiative to establish any theoretical basis for MIR activity, which transcends a mere comprehension of each other's work. Instead, it should be a comprehensive examination of individual disciplines and the establishment of a unified research subject or objective. Subsequently, it involves reaching a consensus that can foster effective collaboration, whilst also contributing to the long-term

[6] Moretti, for example, challenged the concept of canonical literary works with *distant reading*, analysing a large corpus of novels that never entered the closely curated canon.

legitimisation of their multidisciplinary actions. Merely adapting to each other's rules seems akin to attempting to play both football and handball simultaneously, where the distinct nature of each sport's regulations inevitably precludes reaching a consensus based solely on plausibly shared materials. On the other hand, we must consider a possibility that certain disciplines may find their common objective(s) sooner than others, suggesting that there may be more MIRs, some of which more, and other less, compatible.

6.4 The Notion of Discipline and Institution

In the context of the discordance within the intricate web of various music research paradigms, two overarching issues have been dissected: the inexorable advancement of new technologies and the inflexibility in the evolution of underlying theoretical frameworks. The latter alludes to the third obstacle, which is that the efficacy of contemporary academia is heavily dependent on the structuring of institutional space.

6.4.1 The X-Disciplinarity in Contemporary (Music) Academia

Many of the mentioned authors (and possibly others) who have explored computation in music research often tackle the potential issues through the prism of disciplinary (dis)agreements. Some focus on the potential for their expansion, while others address the tensions between technical or scientific music research disciplines and those oriented towards history, humanities, or social sciences (the latter of which usually fall somewhere in-between). More specifically, Becker asked whether "our failure [is] due to our own shortcomings in not becoming thoroughly versed in the protocols and expectations of another discipline? Or, was the failure due to too stringent protocols and expectations for publication in a [...] journal?", concluding that some disciplinary barriers may be unbreachable due to rigid institutional formations [17].

What commenced as experiments of technological capabilities in the field of computer-assisted music research, is now evolving into a whole new institutional discipline or, better still, a group of sub-disciplines of MIR. Most of the researchers in this movement originate from technical and natural sciences which, with a desire to analyse music, have organised themselves into groups, such as ISMIR or SMC (Sound and Music Computing), as well as conferences such as AIMC (AI and Music Creativity), and similar. In the era of an overwhelming flood of data, the multidisciplinary methodology and applied focus of MIR projects enabled these ideas to migrate to natural sciences in general as well as the social sciences and humanities. In addition to research activity, music information methodologies began to appear in many academic curricula, although to this day they are rarely independent from another, more established discipline (such as computer science, acoustics, physics, music studies - musicology, composition, music theory; and so on).

With the expansion of the field, the domain of the MIR is becoming increasingly heterogeneous, although groups such as ISMIR, despite expanding their initiative, maintain some (unofficial) niche preferences or trends, especially when it comes to questions of methodology[7]. Apart from the already mentioned communities, a wide variety of computational methods for studying, teaching, or making music are also appearing elsewhere, for example, in contributions to Empirical Musicology Review (EMR), Systematic Musicology (SysMus), International Association of Music Libraries (IAML), Folk Music Analysis (FMA), International Conference on Artificial Intelligence in Music, Sound, Art and Design (EvoMUSART), and other, more occasional interdisciplinary publications and events.

The pace of emerging multidisciplinary craze is forecasting more drastic changes in the forseeable future for all disciplines, "but we've got to put in place the [institutional] conditions to make it actually happen [57]". Hence, we must ask ourselves - what is a concept of a discipline and what does it take for it to actually change, transform?

6.4.2 The Concept of an Academic Discipline

Mentioned authors commonly debate about the limitations and hurdles of disciplines, but rarely address the pitfalls of "inherent" concept of a "discipline" itself, and by extension, the almost omnipresent contemporary institutional framework of academia.

The divisions of research into fields or disciplines are ever-changing processes. Roughly summarised, if the Ancient Greeks preoccupied themselves with the fields of *logics*, *ethics*, *physics*, and *metaphysika* (metaphysics, philosophy), medieval minds set the terms for *septem artes liberales*, split into *quadrivium* (arithmetics, geometry, astronomy, and music) and *trivium* (grammatics, rhetorics, dialectics) [158]. Later on, the disciplinary matters were divided into different variations of *"natural"* and *"humanities, social"* sciences. Pantin and Kuhn offer further classification of natural sciences into *restrictive* or *mature* (for example, mathematics) versus *non-restrictive* or *non-mature*, of which the latter needs, due to its loose (theoretical) structure, to concern itself with the first, the paradigm of which is firmly set (for example, biology), but not vice versa [227, 151, 158]. Slightly later, the one-dimensionality is replaced by at least three dimensions, such as hard/soft, pure/applied, and life system/non-life system sciences and similar categories, which separate researchers by their object/subject and methodology of research, work organisation and types of output (manuscripts versus reports, papers, individual versus group work, theoretical considerations versus product development, etc.) [23, 139, 105, 16, 158]. All of these categories, as Becher suggests [16], can intertwine, thus the classifications are not to be taken as absolute. A more specific overview of disciplinary construction is taken

[7] Trends can be observed on the surface level, meaning the percentage of thematic selection of yearly accepted proceedings to the ISMIR conference as well as the topics of awarded papers and popularity in citations of the selected papers in the years to follow.

by Wallerstein, who observes the situation within the social sciences [314], and Korsyn, who specifically addresses these matters through the lense of music studies [141].

Nevertheless, the fundamental distinction between hard sciences (natural sciences) and soft sciences (social sciences, humanities), a division still acknowledged today, lies in their respective methodologies. Hard sciences rely on precise measurements and the validation of new discoveries, often leading to the assimilation or replacement of older findings, thus exhibiting a linear progression of events. In contrast, soft sciences engage in detailed synthesis and reinterpretation of ideas spanning historical and contemporary eras, while grappling with phenomena that span across past, present, and future epochs [158]. Both hard and soft sciences can operate in pure or applied contexts where, in the applied realm, both disciplines produce tangible outcomes: hard sciences yield products or techniques such as medical tools, computer applications, or patents, while soft sciences result in guidelines, laws, or educational textbooks.

These and especially further disciplinary divisions "have all struggled continuously on a number of different fronts - intellectual, ideological, and political" [314], to maintain their individual activity, "reputation" and prosper within academic (and social) distribution of power. The latter is measured by universally accepted systems [314] (acquired financial means, e.g., *economic capital* (see [35]), and to some extent, academic hierarchical positions, points and award systems, e.g., *social* and *cultural* capital) [8], or are established on a more symbolic level (*symbolic capital*, (see [35])), which is being constantly re-shaped through a chain of historical and socio-political processes or events.

What occurs when multidisciplinarity arises, particularly involving disciplines from diverse "generalised" spheres characterised by varying means of economic, social, cultural, and symbolic capital? For cooperation to occur, at least one of the disciplines involved must align with the system of the primary institution that (financially) initiates the collaboration. The determination of who possesses the resources to initiate such participation seldom arises solely from individual scientific endeavors. Instead, it typically emerges from internal dynamics within the academic world (research institutes, academia), which adapt to evolving trends in the external environment (the state, economy, politics). These external forces often invest in science with public or private resources [158].

All of the above often complicates the conditions for the free interdisciplinary (or similar) transformation of academic work and curricula, the latter being a key stakeholder in maintaining the status quo of the individual disciplines [158], and is severely restrictive for the flow of knowledge between disciplines and their institutional carriers. The emergence of fields like MIR has instigated numerous changes

[8] It was especially the Bologna Process for higher education that further empowered these divisions, especially by restructuring the productivity scoring in academia, which encourages competition and the need for constant justification of individuals (and their discipline) on the "academic market" by "universal" markers that are by no means universally applicable.

within traditional music disciplines. However, the challenge lies in the fact that the pressure to reconfigure organisational structures based on new intellectual categories is tackled on a country-by-country, university-by-university, project-by-project basis [314].

Wallerstein suggests that in addition to the current disciplinary frameworks, there should be "new avenues for dialogue and exchange beyond (and not merely between) the existing disciplines [314]". He argues for the need to "amplify the organisation of intellectual activity without attention to current disciplinary boundaries [314]". This viewpoint underscores an awareness of the continually evolving nature of disciplinary processes, indicating that wisdom is not monopolized, and knowledge is not reserved solely for individuals with specific university degrees [314]. While this notion is appreciated, those who have the opportunity to explore interdisciplinary realms often face challenging circumstances. For instance, they may encounter difficulties navigating diverse academic scoring systems across multiple disciplines, face hurdles in securing positions that value their multidisciplinary background, and/or find themselves obliged to teach conventional courses unrelated to their research interests, among other obstacles.

The disciplinary structures, primarily focused on administrative and financial concerns, often exert significant pressure on scholars, leading them either to transcend university frameworks to pursue their work [314] or to exit academia altogether. Despite calls for multidisciplinary approaches, many scholars who choose to remain within these structures persist in justifying the importance of their "primary" discipline. They do so by adhering to canonical works, using the vocabulary of a specific disciplinary discourse, and contributing their domain-specific expertise to interdisciplinary projects rather than collaboratively developing new, shared areas with adapted theoretical frameworks.

6.5 Conclusion

In conclusion, this chapter has examined the complex relationship between musicology and Music Information Retrieval (MIR), highlighting both the potential for collaboration and the significant barriers that exist between these (and similar) fields. While this discussion may initially seem peripheral to the main focus of the thesis, it is essential to understanding the broader academic and practical challenges that shape the future of this work. Building on two key contributions—a cross-disciplinary issue of Musicological Annual on computational musicology and a study presented at ISMIR 2023—this chapter has revealed the limited impact of MIR research on musicological scholarship. Despite the large amount of contributions produced by MIR, only a small percentage of it finds its way into musicological discourse, underscoring the need for greater cross-disciplinary engagement.

Furthermore, by addressing the broader theoretical issues of technological impact, academic discipline formation, and internal conflicts, this chapter draws on

Korsyn's concept of disciplinary legitimation to explain why these barriers persist. The tensions within and between fields like MIR and musicology are not just practical but deeply rooted in how (and why) these disciplines define their purpose and identity. While this thesis does not offer definitive solutions, it provides a critical lens through which these disciplinary divisions can be examined. Ultimately, it encourages further introspection and dialogue across academic boundaries to foster more meaningful collaboration in the future.

Chapter 7

Conclusion

7.1 Summary

This thesis explored the potential of pattern matching methods to accommodate the transcriptions of orally transmitted folk songs from Slovenian regions. The methodological focus was on integrating melodic pattern matching with a set of additional descriptors and metadata, while maintaining or enhancing the algorithmic efficiency of such tasks. Furthermore, the research aimed to provide a viable and well-documented dataset that conforms to these methods, serving as a resource for future endeavors by the research or educational community.

As the work undertaken spans a range of diverse topics, I will summarise the contributions by revisiting the initial objectives. As stated, the goal of Objective 1 was to digitise, organise, curate, and thoroughly document materials from selected Slovenian folk song ballads. This was achieved by curating the digitised, though not yet digitally published, scores and annotating them with complementary metadata where possible [32, 33].

Additionally, we examined all factors influencing these sources, ensuring that the dataset is not mistaken for a ”ground truth” corpus on how Slovenian folk songs sound. Rather, it is presented as a reflection of singing trends, available musical knowledge, and the continually evolving practices of collection and digitisation. Both statistical information and in-depth discussions are provided on specific topics, such as few-tone melodies, the challenges of harmonisation, and the relationship between the year of collection and the age of the song.

Second, as a continuation of the previous objective, Objective 2 aimed to annotate the materials and create a publicly accessible digital dataset on the Dezrann platform. This was accomplished by developing a system for retrieving annotations, either automatically or manually, including features such as phrase numbers, melodic contours, phrase labels (melody, verse), time signatures, tone sets, and similar attributes. Upon completing the annotation and curation of the materials, the collection, along with most of the annotations, was transferred to the Dezrann platform. The complete scores, accompanied by metadata, lyrics, and annotations, are now visualised and publicly accessible, as well as released as open data [12][1].

Third, Objective 3 focused on designing algorithms capable of incorporating as much of the available data as possible. We developed algorithms that accommodate the format of our dataset while remaining flexible enough to handle other types of queries and materials with minimal adjustments. For the purposes of this thesis, we designed three different algorithms, each dealing with melodic sequences, descriptors, or a combination of both. A portion of this contribution can also be read in the published paper at ISMIR 2023 [32].

Descriptor pattern matching was divided into two groups: the first handled only one descriptor element per query (e.g., querying for a convex contour but not any others), while the second approach introduced multiple descriptor queries, allowing for the inclusion of multiple elements per descriptor (e.g., convex or descending contour, and similar). This approach afforded us greater flexibility, which is particularly important when working with orally transmitted music, as such music often loses exact information when transferred into symbolic notation, like scores.

Fourth, Objective 4 aimed to expand the evaluation of the methodology by dividing it into a technical performance assessment, two real-life case studies, and a broader discussion on the principles within the wider context of music research.

Fifth, Objective 5 extended the evaluation and [30] discussions by assessing the overall contribution of MIR to fields such as ethnomusicology, and vice versa. This prompted reflection on why certain issues in one discipline do not easily translate to the other, highlighting three underlying causes: (1) miscommunication regarding the relationship to new media and technologies in research, (2) the foundations of distinct disciplinary theoretical frameworks, and (3) the structural organisation of disciplines and research institutions.

7.2 General Contribution

The five objectives ensured five key contributions. First, the project successfully included underrepresented music in the digital space, thereby enriching existing MIR datasets. Both, the open license and the thorough documentation of the dataset ensure that a wide range of studies can benefit from it.

Second, it established the dataset on a digital platform that not only allows for visualisation of the data, but also enables users to interact with it by adding or removing annotations. This makes the dataset accessible and useful not only to the MIR community, but also to those with limited computational knowledge.

Third, in our exploration of the structures and melodic patterns of Slovenian folk songs, along with their descriptors (a process later repeated with children's songs),

[1] Dezrann platform, `https://www.dezrann.net/explore/slovenian-folk-songs`, accessed on 4th October 2024; Slovenian Folk Song Ballads dataset, `https://gitlab.com/algomus.fr/slovenian-folksongs`, accessed on 4th October 2024.

we found that the proposed algorithms, which combine melodic content with descriptors, offer a valuable tool for discovering insights into the characteristics of these songs. Moreover, the introduced methods are adaptable, as they do not rely on predefined assumptions about what constitutes a song, descriptor or a sequence, and do not require precise information on the type of music one is researching. And, while our particular descriptors cannot be directly transmitted onto corpora that are not close to the music structure and research endeavors of this thesis, it is exactly this that confirmed the flexibility as crucial. By thoroughly examining it on two different cases and reflected on such methods being applied in MIR in general, we demonstrated that not all descriptors are universally applicable to describe all content. Nonetheless, most corpora would probably share the finding that individual (melodic) sequences rarely provide enough understanding on the corpus for the findings to be informative, thus some form of descriptors is necessary.

Fourth, an evaluation supported the aforementioned goals and sparked a debate on the usefulness of these principles, suggesting the need to extend such methodologies to a broader range of sources beyond music notation.

Lastly, we conducted a brief analysis of ISMIR's impact on musicology and related fields, exploring the reasons behind the dissonances between these disciplines. This analysis initiated a debate on the need for changes across at least three key areas.

7.3 Future Perspectives

With all contributions in mind, this type of work can still hardly be considered complete. However, what can be concluded is this kind of relationship between pattern matching and ethnomusicology—or more broadly, music, the humanities, and social sciences.

Some of the most important perspectives, in my opinion, are: (1) connecting different formats of the same music, (2) incorporating "non-musical" ideas and discussions as an integral part of a unified analytical framework, (3) better defining the purpose and reasoning behind this work, supported by theoretical justification, and (4) using this understanding to enhance the exchange of ideas between disciplines.

This contribution, while focused on music scores and providing as much context as possible, remained critical of the initial ideas behind such analyses. Future work should certainly explore musical features like polyphonic melodies, rhythm, scales, and harmonies. Moreover, it should move beyond a single music format to offer more comprehensive insights into music as a whole.

One key realisation during this project was that music transcends the confines of notation, audio recordings, and similar formats. Music is a relationship—between people, symbolic forms, and cultures and supporting social structures, rituals, and needs. As computational tools evolve, it is essential to rethink how we can use them to extend music studies beyond simply replicating manual analysis with more data.

The descriptor method introduced here provides a foundation, but expanding the methodology to handle multiple formats and integrate textual or discourse-based contexts would further enhance our understanding of music and its significance.

Lastly, we must address the issues raised in the chapter on the dynamics between MIR and ethnomusicology: what we do, how we do it, and why. This more theoretical line of inquiry is crucial for enabling the advancements discussed earlier. Establishing a strong foundation for future computational approaches will not only foster more fruitful interdisciplinary collaborations but also ensure that data and computation are used more responsibly, effectively, and sustainably.

Appendix A

Ethnomusicological Datasets

The following table gives an overview of some ethnomusicological datasets, including references, URLs, content descriptions, and details on the availability of notation (♪), recordings/audio (🎧), lyrics (📖), metadata (ℹ), and annotations (✎). The table also provides licensing information and assesses the datasets' suitability for ethnomusicological research (**ER**), music education (**ME**), computational music analysis/processing (**MIR**), and general public use (**GP**). Full details on the latter classification can be found in Section 2.2.1. The symbols under the availability and usability categories indicate: **Y** = available/usable, **N** = not available/usable, **O** = partially available/usable, **?** = undetermined with available data.

Dataset	REF	URL	Content	♪	🎧	📖	ℹ	✎	©	ER	ME	MIR	GP
The Essen Folksong collection	[258]	URL	8473 folk songs	Y	O	O	O	O	Partially copyrighted (unspecified).	O	X	Y	X
Dutch Song Database	[144]	URL	140000 Dutch song descriptions, some of which include text, notations and recordings.	O	O	O	Y	X	The copyright of the Royal Dutch Academy of Sciences (?)	Y	O	X	O
Dunya	[281]	URL	Audio recordings, some scores, and descriptive information for Carnatic, Hindustani, Makam, Jingju, and Andalusian music, with tools for analysing and exploring these musical repertoires.	O	O	O	O	O	Prohibited from commercial use, reproduction, distribution, or modification of the website's contents. Restrictions may vary by corpus.	Y	O	Y	O
Compmusic Dataset: Indian Music Tonic Dataset	[100]	URL	597 audio excerpts and manually annotated tonic pitches for Indian art music, along with editorial metadata for the development and evaluation of automatic tonic identification methods. Each recording is linked to an MBID for further details via the Dunya API.	X	O	X	O	Y	CC BY-NC-ND 4.0	X	X	Y	X
Compmusic Datasets: Carnatic Varnam Dataset	[138]	URL	28 solo vocal recordings used for analysing intonation in Carnatic raagas. It includes audio recordings, taala cycle annotations, and machine-readable notations.	Y	Y	O	O	Y	CC BY-NC-ND 4.0	O	X	Y	X
Compmusic Datasets: Carnatic Music Rhythm Dataset	[283]	URL	Rhythm-annotated corpus of 176 excerpts (16.6 hours) in four taalas with audio, designed for automatic rhythm analysis in Carnatic music. It includes manually annotated time-aligned markers for taala cycles and metadata.	X	Y	O	O	X	🎧: non-commercial research usage, with no further distribution. ✎: CC BY-NC-ND 4.0.	O	X	Y	X
Compmusic Datasets: Hindustani Music Rhythm Dataset	[286]	URL	597 audio excerpts and manually annotated tonic pitches for Indian art music, along with editorial metadata for development and evaluation of automatic tonic identification. Each recording is linked to an MBID via the Dunya API.	X	Y	O	O	Y	🎧: non-commercial research usage, with no further distribution. ✎: CC BY-NC-ND 4.0.	O	X	Y	X
Compmusic Datasets: Mridangam Stroke Dataset	[8]	URL	7162 audio examples of individual Mridangam strokes across various tonics of 10 different strokes played on Mridangams with 6 tonic values.	X	Y	O	?	Y	CC BY-NC 4.0	O	X	Y	X
Compmusic Datasets: Tani-avarthanam Dataset	[163]	URL	Two transcribed tani-avarthanams performed by Padmavibhushan Umayalpuram K. Sivaraman. Recorded at IIT Madras and annotated by professional Carnatic percussionists. It includes 24 minutes of audio and 8800 strokes.	X	Y	X	O	O	CC BY-NC-ND 4.0	X	X	Y	X

TABLE A.1: Overview of ethnomusicological datasets (continued)

Dataset	REF	URL	Content	♪	🎧	📖	ℹ	✎	©	ER	ME	MIR	GP
Compmusic Datasets: Tabla Solo dataset	[101]	URL	38 solo audio recordings with time-aligned syllabic scores of tabla compositions. The data, sourced from the instructional DVD *Shades of Tabla by Pt. Arvind Mulgaonkar* with audio and scores.	O	Y	X	Y	O	🎧: for non-commercial research usage with no further distribution. ✎: CC BY-NC-ND 4.0.	X	X	Y	X
Compmusic Datasets: Turkish Makam Symbolic Phrase Dataset	[127]	URL	480 Turkish makam music scores, segmented into phrases tu support research on melodic similarity between phrases and the relationship between melodic phrasing and meter.	Y	X	X	?	Y	CC BY-NC-SA 4.0	O	X	Y	X
Compmusic Datasets: Turkish Makam Melodic Phrase Dataset	[39]	URL	899 SymbTr scores, manually annotated into melodic segments by three experts, with a total of 31362 phrase annotations.	Y	X	X	?	Y	CC BY-NC-ND 4.0	O	X	Y	X
Compmusic Datasets: Turkish şarki vocal dataset	[73]	URL	Recordings of şarki compositions from MusicBrainz. Version 1: 10 performances (5 male, 5 female) with minimal backing vocals and no percussion, in .wav format. Version 2: 12 performances (8 female, 4 male) of 11 compositions, with some recordings added and others omitted from Version 1.	X	Y	Y	O	O	CC BY-NC-ND 4.0	O	X	O	X
Compmusic Datasets: Turkish makam acapella sections dataset	[74]	URL	Recordings of şarki compositions sung a cappella by professional singers (12 a-cappella performances of 11 compositions) (recordings of Version 2 of Turkish Şarki Vocal Dataset). It provides an a cappella counterpart to polyphonic recordings, and annotations for sections, lyrics phrases, and individual words, all aligned with their corresponding audio segments.	O	Y	Y	O	O	CC BY-NC-ND 4.0	O	X	Y	X
Compmusic Datasets: Turkish Makam Audio - Score Alignment Dataset	[20]	URL	6 audio recordings of 4 peşrev compositions from the classical Ottoman-Turkish tradition with 51 sections in the audio recordings in total. The total number of the note annotations in the audio recordings are 3896.	O	Y	X	?	Y	CC BY 4.0	X	X	Y	X
Compmusic Datasets: Turkish Makam Section Dataset	[328]	URL	2095 sections annotated in 257 audio recordings of 58 compositions.	Y	Y	X	O	Y	Other (Open (?))	X	X	Y	X
Compmusic Datasets: Turkish Composition Identification Dataset	[329]	URL	147 instrumental music scores from the SymbTr collection and 743 audio recordings, of which 360 recordings are linked to 87 music scores.	Y	Y	X	O	Y	CC BY-NC-SA 4.0	O	O	Y	X

TABLE A.1: Overview of ethnomusicological datasets (continued)

Dataset	**REF**	**URL**	**Content**	♪	🎧	📖	ℹ	✎	©	**ER**	**ME**	**MIR**	**GP**
Compmusic Datasets: Turkish Makam Tonic Dataset	[329]	URL	Annotated tonic frequencies for over 2000 audio recordings.	O	Y	X	O	Y	CC BY-NC-SA 4.0	X	X	Y	X
Compmusic Datasets: Turkish Makam Recognition Dataset	[126]	URL	50 recordings for each of the 20 most common Turkish makams to test makam recognition methodologies.	O	Y	O	O	Y	CC BY-NC-SA 4.0	X	X	Y	X
Compmusic Datasets: Beijing Opera Percussion Instrument Dataset	[297]	URL	236 isolated stroke examples spanning the four percussion instrument classes used in Beijing Opera.	O	Y	O	O	O	CC BY 4.0	O	X	Y	X
Compmusic Datasets: Beijing Opera Percussion Pattern Dataset	[285]	URL	133 audio percussion patterns across five pattern classes.	Y	Y	X	X	Y	✎: CC BY-ND 4.0. 🎧: commercial releases restricted from public sharing, but available on request for non-commercial purposes.	O	X	Y	X
Compmusic Datasets: Jingju A Cappella Singing Pitch Contour Dataset	[96]	URL	Pitch contour segment ground truth for 39 Jingju a cappella recordings with annotations of melodic transcription and pitch contour segmentations, extracted from audio recordings.	Y	Y	O	O	O	CC BY-NC 4.0	X	X	Y	X
Compmusic Datasets: Jingju A Cappella Singing Audio and Boundary Annotation Dataset	[24]	URL	Hierarchical boundary annotations for a cappella singing performed by both professional and amateur Jingju singers, including line (phrase), syllable, and phoneme units for Jingju a cappella audio recordings.	O	O	O	O	Y	CC BY-NC 4.0	X	X	Y	X
Compmusic Datasets: Jingju A Cappella Singing Audio Extended Dataset	[95]	URL	120 arias, comprising 1265 melodic lines.	X	Y	X	Y	Y	CC BY-NC 4.0	O	X	Y	X
Compmusic Datasets: Jingju Music Scores Collection	[248]	URL	92 Jingju music scores for analysing Jingju singing within its musical system.	Y	X	Y	Y	Y	CC BY-NC-ND 4.0	Y	X	Y	X

TABLE A.1: Overview of ethnomusicological datasets (continued)

Dataset	REF	URL	Content	♪	🎧	📖	ℹ	✎	©	ER	ME	MIR	GP
Compmusic Datasets: Jingju Lyrics Datasets	[247]	URL	Jingju lyrics datasets for analysing the expressive functions of Jingju rhythmic types.	X	X	Y	O	O	CC BY-NC-ND 4.0	Y	X	Y	X
Compmusic Datasets: Annotated Jingju Arias Dataset	[325]	URL	34 Jingju arias manually segmented at various levels using Praat v5.3.53, including samples of the two main Jingju melodic styles, Xipi and Erhuang, and featuring the five primary singing roles: Dan, Jing, Laodan, Laosheng, and Xiaosheng.	O	X	Y	Y	Y	CC BY-NC-ND 4.0	Y	X	Y	X
Million Song Dataset	[22]	URL	Audio features and metadata for one million contemporary popular music tracks, complemented by community-contributed datasets including SecondHandSongs, musiXmatch, Last.fm, Taste Profile subset, thisismyjam-to-MSD mapping, tagtraum genre annotations, and Top MAGD.	X	X	O	Y	Y	Echo Nest & SecondHandSongs: research only (with potentially more licensing for private companies); MusicBrainz: track "year" - public domain, "tags" and "tag count" - CC BY-NC-SA 2.0; musiXmatch: research only, strictly non-commercial; code: GPL.	X	X	Y	X
The American Folk Song Collection	X	URL	Music scores (PDF), analysis, and sound, primarily for educational use with children's songs with lyrics, background information, game directions, field recordings, Kodály-related short films, and teaching resources.	O	O	Y	Y	O	♪: Allowed to print and share in classrooms, rehearsals, and homes. 🎧: Play recordings for students, but no downloading or digital copies. **Performance:** Non-commercial, educational use in concerts and workshops with appropriate attribution.	Y	Y	O	Y
Saraga	[284]	URL	Time-aligned melody, rhythm, and structural annotations for Carnatic and Hindustani music of 168 tracks from the Carnatic collection and annotations for instruments such as Ghatam, Mridangam, Violin, Voice, and Secondary Voice.	O	Y	X	O	Y	🎧 & ✎: CC BY-NC 4.0. Code and scripts: AGPL-3.0.	Y	X	Y	X

TABLE A.1: Overview of ethnomusicological datasets (continued)

Dataset	REF	URL	Content	♪	🎧	📖	ℹ	✎	©	ER	ME	MIR	GP
The Lyra Dataset	[228]	URL	1570 pieces and approximately 80 hours of data of Greek Traditional and Folk music with timestamped YouTube links for audio and video, and detailed metadata on instrumentation, geography, genre, ...	X	O	X	Y	Y	ℹ: CC BY-NC-SA 4.0. Other data: Unknown.	O	X	Y	X
Erkomaishvili Dataset	[253]	URL	MP3 audio, transcriptions, segment annotations (CSV), F0-annotations (CSV), note-onset annotations (CSV), and digital sheet music (MusicXML) of historic tape recordings of three-voice Georgian songs by master chanter Artem Erkomaishvili.	Y	Y	?	Y	Y	Unknown	O	X	Y	X
UCSB Cylinder Audio Archive	[88]	URL	Over 10000 cylinder recordings held by the UCSB Library, featuring a wide range of recordings from the late 1800s to early 1900s, including popular songs, vaudeville acts, classical and operatic music, comedic monologues, ethnic and foreign recordings, speeches, and readings.	X	Y	X	Y	X	Cylinder 🎧 (<1922): public domain (can be used and downloaded for any purpose). 🎧(mp3) (1923<): CC BY-NC 2.5 with acknowledgment to the University of California (Santa Barbara Library). Original WAV cylinders files (1922<:) available upon request for commercial or non-commercial use, charging a use fee (varies depending on the type of content and purpose).	Y	Y	Y	O
Vaughan Williams Memorial Library Archives	[134, 321]	URL	Indexed songs, tunes, and dances, with some downloadable recordings. It contains diverse materials such as books, pamphlets, periodicals, artworks, photographs, and various audio formats.	X	O	O	Y	X	Unknown	Y	Y	O	Y
Digital Collections of Library of Congress	X	URL	The Library of Congress digital collections encompass textual, image, audio, and score materials. Music: 15618 notated music examples and 59605 audio recordings, featuring notable collections such as Alan Lomax's field recordings (444), African-American Band music (363), New Mexico Folklife Project (112), and 19th-Century American Sheet Music (3500).	O	Y	O	Y	X	Users must verify copyright status and comply with restrictions. Written permission is required for uses beyond fair use.	Y	Y	O	Y

TABLE A.1: Overview of ethnomusicological datasets (continued)

Dataset	**REF**	**URL**	**Content**	♪	🎧	📖	ℹ	✎	©	**ER**	**ME**	**MIR**	**GP**
The Latin Music Database	[275]	URL	3227 MP3 recordings of 10 latin song genres.	X	Y	X	O	Y	Feature vectors: Weka-compatible ARFF format (open); Other data: Unknown	X	X	?	X
Basque Song Collections	[49]	URL	Over 5000 tabs, each with a comprehensive description (title, date, place of collection, etc.), along with scores and melodies in MIDI format.	Y	O	O	Y	O	2011 and later: CC BY-NC-SA 3.0	Y	O	Y	O
Etnofon	X	URL	The Etnofon collection, managed by the Institute of Ethnomusicology of ZRC SAZU, offers annotated sound recordings and accompanying texts online. It replaces the previous CD collection, From the Archives of the Institute of Ethnomusicology. Etnofon features thematic editions of Slovenian musical traditions, highlighting typological, textual, dialectal, musical, and dance characteristics to showcase both commonalities and diversity within the tradition.	X	Y	X	Y	O	Educational and research use only, non-commercial access to digital materials. 🎧: the majority is not copyrighted. Users must secure rights for any use beyond education and research.	Y	Y	X	Y
RISM (Répertoire International des Sources Musicales)	X	URL	Over 1.2 million historical musical sources in searchable database, focusing on the period between 1600 and 1850. Music: 59000 manuscripts, prints, libretti, and treatises.	O	O	O	O	X	The catalogue: CC BY 3.0. Other materials: individually-attributed copyrights.	Y	Y	O	O
Youth Magazines *Ciciban* (GMGM)	X	URL	123 Slovenian children's songs from the outh magazines Ciciban with some annotations.	Y	X	O	O	O	CC BY-NC 4.0	O	O	Y	O
Bohdan Medwidsky Ukrainian Folklore Archives (BMUFA)	X	URL	The largest repository of Ukrainian and Canadian-Ukrainian folklore in North America. It documents, preserves, and studies Ukrainian vernacular culture, with extensive collections including music, student fieldwork, ethnographic studies, family correspondence, and cultural artifacts.	O	O	O	O	X	Under the Copyright Act (Canada). Reproductions may require the copyright owner's authorization. Research and private study are usually exempt, but publication, exhibition, or commercial use requires permission.	O	O	X	O

TABLE A.1: Overview of ethnomusicological datasets (continued)

Dataset	**REF**	**URL**	**Content**	♪	🎧	📖	ℹ	✎	©	**ER**	**ME**	**MIR**	**GP**
Lithuanian Folklore Archive's Database (Sound recordings)	X	URL	Digitised audio recordings from the Lithuanian Folklore Archive, including the oldest folklore recordings in Lithuania, originally captured on phonograph cylinders (1908–1949) and discs (1935–1949), as well as over 8500 hours of recordings on magnetic tapes and cassettes (currently being digitised). It houses metadata and sound recordings from these historic media, featuring authentic recordings of sutartinės.	X	Y	X	O	X	Unknown	Y	O	O	O
Estonian Runsongs' Database (ERAB)	X	URL	108,969 traditional Estonian folk song texts with approximately 6000 transitional and rhyming songs. It includes digitised materials from major Estonian folklore collections, and has ongoing additions from the Estonian Folklore Archive and National Literary Museum, as well as manuscript images through Kivike.	X	X	Y	Y	O	Unknown	Y	O	O	O
Digital Archives of Latvian Folklore (Recordings)	REFS	URL	Phonograph recordings from 1926 onwards, specifically 191 digitised wax cylinders by 1947 and a significant collection of tape reels and compact cassettes, many of which digitised and available. It is also possible to engage with the content by transcribing, translating, and adding information.	X	Y	O	Y	X	Open to the public free of charge (manuscripts, images, audio and video recordings, and other).	Y	Y	Y	Y
Project Dúchas, The Folk Music Archive (NFC): Sound Collection and the Folk Music Archive	REFS	URL	Around 3000 pages of music manuscripts and approximately 2,000 hours of original field recordings. It features a vast collection of songs in both English and Irish, alongside instrumental music from across Ireland. As this is a newer project, not all material is available yet.	?	Y	O	Y	X	CC BY-NC 4.0. Most resources (besides images) are unavailable to download.	Y	Y	X	Y
Lomax Digital Archive	[196, 112]	URL	A comprehensive collection of audio and visual materials compiled by folklorists Alan Lomax and John A. Lomax, spanning seven decades, including recordings, photographs, and transcriptions. It is fully searchable with detailed item-level metadata. Collections are categorized into fieldwork, film and video, radio shows, discussions, lectures & interviews, and Alan Lomax as a performer.	X	Y	Y	Y	O	The rights depend on individual audio, photographic, and video materials, thus their use must be cleared by specific rights holders.	Y	Y	O	Y

TABLE A.1: Overview of ethnomusicological datasets (continued)

Dataset	**REF**	**URL**	**Content**	♪	🎧	📖	ℹ	✎	©	**ER**	**ME**	**MIR**	**GP**
The Global Jukebox	[322]	URL	Thousands of examples of music, dance, and expressive behaviour worldwide, with rich cultural, geographic, and descriptive metadata, including collections of Alan Lomax and his colleagues, featuring studies in Cantometrics, Choreometrics, and other methodologies. The archive is designed for educational and research purposes and offers various search options and tools.	X	Y	Y	Y	Y	Strictly for educational and research purposes. 🎧 & audiovisual files: streaming only without downloading or other uses.	Y	Y	O	Y

Bibliography

[1] Sanu Pulimootil Achankunju. “Music Search Engine from Noisy OMR Data”. In: *Proceedings of the 1st International Workshop on Reading Music Systems*. Paris, France, 2018, pp. 23–24.

[2] Charles R. Adams. “Melodic Contour Typology”. In: *Ethnomusicology* 20.2 (May 1976), pp. 179–215. ISSN: 00141836. DOI: 10.2307/851015.

[3] Donald Adjeroh, Timothy Bell, and Amar Mukherjee. *The Burrows-Wheeler Transform: Data Compression, Suffix Arrays, and Pattern Matching*. Springer Science & Business Media, 2008.

[4] Kofi Agawu. “Does Music Theory Need Musicology”. In: *Current Musicology* 53 (1992), pp. 89–98.

[5] Sven Ahlbäck. “Melody Beyond Notes: A Study of Melody Cognition”. PhD thesis. Göteborgs Universitet, 2004.

[6] Stephen F Altschul et al. “Basic Local Alignment Search Tool”. In: *Journal of molecular biology* 215.3 (1990), pp. 403–410.

[7] Christina Anagnostopoulou, Mathieu Giraud, and Nick Poulakis. “Melodic Contour Representations in the Analysis of Children’s Songs”. In: *International Workshop on Folk Music Analysis (FMA 2013)*. 2013, pp. 40–43.

[8] Akshay Anantapadmanabhan, Ashwin Bellur, and Hema A. Murthy. “Modal Analysis and Transcription of Strokes of the Mridangam Using Non-negative Matrix Factorization”. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013)*. 2013, pp. 181–185.

[9] David Atkinson and Steve Roud. *The English Folk Song: Some Conclusions*. Woodbridge, United Kingdom: Boydell Press, 2007.

[10] Jean-Julien Aucouturier and Emmanuel Bigand. “Mel Cepstrum & Ann Ova: The Difficult Dialog Between MIR and Music Cognition.” In: *International Society for Music Information Retrieval Conference (ISMIR 2012)*. 2012, pp. 397–402.

[11] Jay Ayres et al. “Sequential Pattern Mining Using a Bitmap Representation”. In: *Proceedings of the Eighth International Conference on Knowledge Discovery and Data Mining (ACM SIGKDD 2002)*. 2002, pp. 429–435.

[12] Charles Ballester et al. “Interacting with Annotated and Synchronized Music Corpora on the Dezrann Web Platform”. In: *Transactions of the International Society for Music Information Retrieval (TISMIR)* (Under review).

[13] Béla Bartók. *Hungarian Folk Music*. Oxford, United Kingdom: Oxford University Press, 1924.

[14] Samuel P Bayard. “Prolegomena to a Study of the Principal Melodic Families of British-American Folk Song”. In: *The Journal of American Folklore* 63.247 (1950), pp. 1–44.

[15] David Beach. *Schenkerian Analysis: Perspectives on Phrase Rhythm, Motive and Form*. Routledge, 2019.

[16] Tony Becher and P Trowler. *Tribes and Territories-Intellectual Enquiry and the Culture of Disciplines*. Philadelphia: Open University Press, 1989.

[17] Judith Becker. "Crossing Boundaries: An Introductory Essay". In: *Empirical Musicology Review* 4.2 (2009), pp. 45–48.

[18] Daniel Bedoya, Lawrence Fyfe, and Elaine Chew. "A Perceiver-Centered Approach for Representing and Annotating Prosodic Functions in Performed Music". In: *Frontiers in Psychology* 13 (2022), p. 886570.

[19] Regina Bendix. *In Search of Authenticity: The Formation of Folklore Studies*. Wisconsin, USA: UW Press, 1997.

[20] Emmanouil Benetos and Andre Holzapfel. "Automatic Transcription of Turkish Makam Music". In: *Proceedings of 14th International Society for Music Information Retrieval Conference (ISMIR 2013)*. Curitiba, Brazil, 2013.

[21] Bonnie Berger, Michael S Waterman, and Yun William Yu. "Levenshtein distance, sequence comparison and biological database search". In: *IEEE transactions on information theory* 67.6 (2020), pp. 3287–3294.

[22] Thierry Bertin-Mahieux et al. "The Million Song Dataset". In: *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*. University of Miami, Miami, 2011, pp. 591–596.

[23] Anthony Biglan. "Relationships Between Subject Matter Characteristics and the Structure and Output of University Departments". In: *Journal of Applied Psychology* 57.3 (1973), pp. 204–213.

[24] D. A. A. Black, M. Li, and M. Tian. "Automatic Identification of Emotional Cues in Chinese Opera Singing". In: *13th International Conference on Music Perception and Cognition (ICMPC 2014)*. 2014, pp. 250–255.

[25] Kurt Blaukopf. *Musical Life in a Changing Society: Aspects of Music Sociology*. Portland, Oregon: Amadeus Press, 1992.

[26] Georg Boenn. *Computational Models of Rhythm and Meter*. Cham, Switzerland: Springer, 2018.

[27] Dmitry Bogdanov et al. "Essentia: An Audio Analysis Library for Music Information Retrieval". In: *International Society for Music Information Retrieval Conference (ISMIR 2013)*. 2013, pp. 493–498,

[28] Peter Boot, Anja Volk, and W Bas de Haas. "Evaluating the Role of Repeated Patterns in Folk Song Classification and Compression". In: *Journal of New Music Research* 45.3 (2016), pp. 223–238.

[29] Bogdana Borota et al. *Otrok v svetu glasbe, plesa in lutk*. Ljubljana, Slovenia: Pedagoška fakulteta, 2006.

[30] Vanessa Nina Borsan, Mathieu Giraud, and Richard Groult. *The Games We Play: Exploring the Impact of ISMIR on Musicology*. Milan, Italy, Nov. 2023.

[31] Vanessa Nina Borsan and Leon Stefanija, eds. *Introduction*. Vol. 58/2. Ljubljana, Slovenia: University of Ljubljana Press, 2022, pp. 10–14.

[32] Vanessa Nina Borsan et al. "Adding Descriptors to Melodies Improves Pattern Matching: A Study on Slovenian Folk Songs". In: *International Society for Music Information Retrieval Conference (ISMIR 2023)*. Milan, Italy, Nov. 2023, pp. 474–481.

[33] Vanessa Nina Borsan et al. "Introducing The Digitised Dataset of Slovenian Folk Ballads". In: *Ethnomusicology Forum* (Under review).

[34] Dimitrios Bountouridis et al. "Melodic Similarity and Applications Using Biologically-Inspired Techniques". In: *Applied Sciences* 7.12 (Dec. 2017), p. 1242. (Visited on 08/05/2023).

[35] Pierre Bourdieu. *Outline of a Theory of Practice*. Cambridge, UK: Cambridge University Press, 1977.

[36] Oliver Bown. *Beyond the Creative Species: Making Machines That Make Art and Music*. Cambridge, Massachusetts: The MIT Press, 2021.

[37] Barış Bozkurt. "Computational Analysis of Overall Melodic Progression for Turkish Makam Music". In: *Penser l'improvisation*. Sampzon: Delatour France, July 2015, pp. 289–298.

[38] Barış Bozkurt et al. "Usul and Makam Driven Automatic Melodic Segmentation for Turkish Music". In: *Journal of New Music Research* 43.4 (2014), pp. 375–389.

[39] Barış Bozkurt et al. "Usul and Makam Driven Automatic Melodic Segmentation for Turkish Music". In: *Journal of New Music Research* 43.4 (2014), pp. 375–389.

[40] Michael Burrows. "A Block-Sorting Lossless Data Compression Algorithm". In: *SRS Research Report* 124 (1994).

[41] Roger Caillois. *Man, Play, and Games*. Champaign, Illinois: University of Illinois Press, 2001.

[42] Emilios 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.

[43] Emilios Cambouropoulos. "The Local Boundary Detection Model (LBDM) and its Application in the Study of Expressive Timing". In: *International Computer Music Conference (ICMC 2001)*. 2001.

[44] Rafael Caro Repetto et al. "Comparision of the Singing Style of Two Jingju Schools". In: *International Society for Music Information Retrieval Conference (ISMIR 2015)*. Málaga, Spain, Oct. 2015, pp. 507–13.

[45] Nádia Carvalho and Gilberto Bernardes. "Towards Balanced Tunes: A Review of Symbolic Music Representations and Their Hierarchical Modeling". In: *International Conference on Computational Creativity (ICCC)*. 2020, pp. 236–242.

[46] Rudi Cilibrasi, Paul Vitányi, and Ronald De Wolf. "Algorithmic Clustering of Music". In: *Proceedings of the Fourth International Conference on Web Delivering of Music (EDELMUSIC 2004)*. IEEE. 2004, pp. 110–117.

[47] Martin Clayton et al. "Raga Classification From Vocal Performances Using Multimodal Analysis". In: *International Society for Music Information Retrieval Conference (ISMIR 2022)*. Bengaluru, India, Dec. 2022, pp. 283–290.

[48] Darrell Conklin. "Antipattern Discovery in Folk Tunes". In: *Journal of New Music Research* 42.2 (2013), pp. 161–169.

[49] Darrell Conklin. *Basque Songbook - Eusko Ikaskuntza*. Database. 2011. URL: https://www.eusko-ikaskuntza.eus/en/documentary-collection/basque-songbook/ (visited on 07/20/2023).

[50] Darrell Conklin. "Discovery of Distinctive Patterns in Music". In: *Intelligent Data Analysis* 14.5 (Sept. 2010), pp. 547–554.

[51] Darrell Conklin. "Mining Contour Sequences for Significant Closed Patterns". In: *Journal of Mathematics and Music* 15.2 (May 2021), pp. 112–124.

[52] Darrell Conklin. "Pattern in Music". In: *Journal of Mathematics and Music* 15.2 (May 2021), pp. 95–98.

[53] Darrell Conklin and Christina Anagnostopoulou. "Comparative Pattern Analysis of Cretan Folk Songs". In: *Journal of New Music Research* 40.2 (2011), pp. 119–125.

[54] Darrell Conklin and Christina Anagnostopoulou. "Representation and Discovery of Multiple Viewpoint Patterns". In: *International Computer Music Conference (ICMC 2001)*. 2001, pp. 479–485.

[55] Darrell Conklin and Mathieu Bergeron. "Discovery of Contrapuntal Patterns". In: *International Society for Music Information Retrieval Conference (ISMIR 2010)*. Vol. 2010. 2010, pp. 201–206.

[56] Camelia Constantin et al. "The Melodic Signature Index for Fast Content-based Retrieval of Symbolic Scores". In: *The 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*. Miami, Florida, 2011, pp. 363–368.

[57] Nicholas Cook. "Towards the Compleat Musicologist". In: *International Conference on Music Information Retrieval (ISMIR 2005)*. 2005.

[58] Bas Cornelissen, Willem Zuidema, and John Ashley Burgoyne. "Cosine Contours: A Multipurpose Representation for Melodies". en. In: *22nd International Society for Music Information Retrieval Conference (ISMIR 2021)*. Online, Nov. 2021, pp. 135–42.

[59] James R Cowdery. "A Fresh Look at the Concept of Tune Family". In: *Ethnomusicology* 28.3 (1984), pp. 495–504.

[60] Maxime Crochemore, Christophe Hancart, and Thierry Lecroq. *Algorithms on strings*. Cambridge University Press, 2007.

[61] Maxime Crochemore and Thierry Lecroq. "Text Searching and Indexing". In: *Recent Advances in Formal Languages and Applications* 25 (2006), pp. 43–80.

[62] Michael Scott Cuthbert and Christopher Ariza. "music21: A Toolkit for Computer-aided Musicology and Symbolic Music Data". In: *International Society for Music Information Retrieval Conference (ISMIR 2010)* (2010), pp. 637–642.

[63] Ewa Dahlig Turek et al. *Musicology (Re-) Mapped: Discussion Paper*. European Science Foundation, 2012.

[64] Fred J Damerau. "A Technique for Computer Detection and Correction of Spelling Errors". In: *Communications of the ACM* 7.3 (1964), pp. 171–176.

[65] Roger Dannenberg. "Music Representation Issues, Techniques, and Systems". In: *Computer Music Journal* 17.3 (1993), pp. 20–30.

[66] Jürgen Diet. *Innovative MIR Applications at the Bayerische Staatsbibliothek*. 2018.

[67] James E Dobson. *Critical Digital Humanities: The Search for a Methodology*. Champaign, Illinois: University of Illinois Press, 2019.

[68] J Stephen Downie. “The Scientific Evaluation of Music Information Retrieval Systems: Foundations and Future”. In: *Computer Music Journal (Music Information Retrieval)* 28.2 (2004), pp. 12–23.

[69] J Stephen Downie. “Toward the Scientific Evaluation of Music Information Retrieval Systems”. In: *International Conference on Music Information Retrieval (ISMIR 2003)*. Washington, D.C., 2003.

[70] J Stephen Downie, Donald Byrd, and Tim Crawford. “Ten Years of ISMIR: Reflections on Challenges and Opportunities.” In: *International Society for Music Information Retrieval Conference (ISMIR 2009)*. Kobe, Japan, 2009, pp. 13–18.

[71] Johanna Drucker. “Humanistic Theory and Digital Scholarship”. In: *Debates in the Digital Humanities* 150 (2012), pp. 85–95.

[72] Timothy C Duguid et al. “Music Scholarship Online (MuSO): A Research Environment for a More Democratic Digital Musicology”. In: *Digital Humanities Quarterly* 13.1 (2019).

[73] Georgi Dzhambazov and Xavier Serra. “Modeling of Phoneme Durations for Alignment between Polyphonic Audio and Lyrics”. In: *Sound and Music Computing Conference*. 2015.

[74] Georgi Dzhambazov, Sertan Şentürk, and Xavier Serra. “Searching Lyrical Phrases in A-Capella Turkish Makam Recordings”. In: *16th International Society for Music Information Retrieval Conference (ISMIR 2015)*. 2015.

[75] Ken Déguernel and Bob L. T. Sturm. “Bias in Favour or Against Computational Creativity: A Survey and Reflection on the Importance of Socio-cultural Context in its Evaluation”. In: *Proceedings of the International Conference on Computational Creativity*. Waterloo, Canada, 2023.

[76] Tuomas Eerola. *MIDI toolbox: Matlab tools for music research*. Kopijyvä, Jyväskylä, Finland: University of Jyväskylä, Jan. 2004.

[77] Tuomas Eerola and Micah Bregman. “Melodic and Contextual Similarity of Folk Song Phrases”. In: *Musicae Scientiae* 11.1 (2007), pp. 211–233.

[78] Paolo Ferragina and Giovanni Manzini. “Indexing Compressed Text”. In: *Journal of the ACM (JACM)* 52.4 (2005), pp. 552–581.

[79] Paolo Ferragina and Giovanni Manzini. “Opportunistic Data Structures with Aspplications”. In: *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*. IEEE. 2000, pp. 390–398.

[80] Christoph Finkensiep, Markus Neuwirth, and Martin Rohrmeier. “Generalized Skipgrams for Pattern Discovery in Polyphonic Streams”. In: *International Society for Music Information Retrieval Conference (ISMIR 2018)*. Paris, France, 2018, pp. 547–553.

[81] Christoph Finkensiep et al. “Repetition-Structure Inference With Formal Prototypes”. In: *International Conference on Music Information Retrieval (ISMIR 2015)*. Milan, Italy, 2023.

[82] Christoph Finkensiep et al. “Voice-leading Schema Recognition using Rhythm and Pitch Features”. In: *International Society for Music Information Retrieval Conference (ISMIR 2020)*. 2020, pp. 520–526.

[83] Ruth Finnegan. *Oral Traditions and the Verbal Arts*. London, United Kingdom: Routledge, 1992.

[84] Arthur Flexer. "Statistical Evaluation of Music Information Retrieval Experiments". In: *Journal of New Music Research* 35.2 (2006), pp. 113–120.

[85] Luciano Floridi. *Information: A Very Short Introduction*. Oxford, United Kingdom: Oxford University Press, 2010.

[86] Bradley W Frankland and Annabel J Cohen. "Parsing of Melody: Quantification and Testing of the Local Grouping Rules of Lerdahl and Jackendoff's A Generative Theory of Tonal Music". In: *Music Perception* 21.4 (2004), pp. 499–543.

[87] Albert B. Friedman. *Ballad*. *Encyclopedia Britannica*. 2024. URL: https://www.britannica.com/art/ballad (visited on 08/09/2024).

[88] Pedro Félix. "Cylinder Preservation and Digitization Project". In: *Yearbook for Traditional Music* 47 (2015), pp. 235–236.

[89] Louis Garczynski et al. "Modeling and Editing Cross-Modal Synchronization on a Label Web Canvas". In: *Music Encoding Conference (MEC 2022)*. Halifax, Canada, 2022.

[90] Mathieu Giraud, Richard Groult, and Florence Levé. "Computational Analysis of Musical Form". In: *Computational Music Analysis* (2016), pp. 113–136.

[91] Simon Gog et al. "From Theory to Practice: Plug and Play with Succinct Data Structures". In: *International Symposium of Experimental Algorithms (SEA 2014)*. 2014, pp. 326–337.

[92] Izaro Goienetxea Urkizu et al. "Ontologies for representation of folk song metadata". In: (2012).

[93] Marjetka Golež Kaučič et al. *Slovenske ljudske pesmi V.: Pripovedni pesmi*. sl. Google-Books-ID: Q5dVPgAACAAJ. Založba ZRC, 2007. ISBN: 978-961-254-004-3.

[94] Marjetka Golež Kaučič et al. *Slovenske ljudske pesmi V: Pripovedni pesmi*. Ljubljana, Slovenia: Založba ZRC, 2007.

[95] Rong Gong, Rafael Caro Repetto, and Xavier Serra. "Creating an A Cappella Singing Audio Dataset for Automatic Jingju Singing Evaluation Research". In: *4th International Digital Libraries for Musicology workshop (DLfM 2017)*. Shanghai, China, 2017.

[96] Rong Gong, Yile Yang, and Xavier Serra. "Pitch Contour Segmentation for Computer-aided Jingju Singing Training". In: *Sound and Music Computing (SMC 2016)*. Hamburg, Germany, 2016.

[97] Jack Goody. *The Interface Between the Written and the Oral*. Cambridge, United Kingdom: Cambridge University Press, 1987.

[98] Mark Gotham. "Chromatic Chords in Theory and Practice". In: *24th International Society for Music Information Retrieval Conference (ISMIR 2023)*. Milan, Italy, 2023.

[99] Mark Gotham et al. "The "Measure Map": An Inter-operable Standard for Aligning Symbolic Music". In: *Proceedings of the 10th International Conference on Digital Libraries for Musicology (DLfM 2023)*. 2023, pp. 91–99.

[100] Sankalp Gulati et al. "Automatic Tonic Identification in Indian Art Music: Approaches and Evaluation". In: *Journal of New Music Research* 43.01 (2014), pp. 55–73.

[101] Swapnil Gupta et al. "Discovery of Syllabic Percussion Patterns in Tabla Solo Recordings". In: *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR 2015)*. 2015.

[102] Dan Gusfield. *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*. Cambridge University Press, 1997.

[103] Juan Sebastián Gómez-Cañón et al. "TROMPA-MER: An Open Dataset for Personalized Music Emotion Recognition". In: *Journal of Intelligent Information Systems* 60.2 (Apr. 2023), pp. 549–570.

[104] W Bas de Haas, Anja Volk, and Frans Wiering. "Structural Segmentation of Music Based on Repeated Harmonies". In: *2013 IEEE International Symposium on Multimedia*. IEEE. 2013, pp. 255–258.

[105] Jürgen Habermas. *Knowledge and Human Interests*. Wiley, 2015.

[106] Reginald Hackforth. *Plato: Phaedrus*. Cambridge, United Kingdom: Cambridge University Press, 1972.

[107] Saqib Iqbal Hakak et al. "Exact String Matching Algorithms: Survey, Issues, and Future Research Directions". In: *IEEE access* 7 (2019), pp. 69614–69637.

[108] Masatoshi Hamanaka, Keiji Hirata, and Satoshi Tojo. "Implementing methods for analysing music based on lerdahl and jackendoff's generative theory of tonal music". In: *Computational music analysis* (2016), pp. 221–249.

[109] Dorien Herremans and Ching-Hua Chuan. "Modeling Musical Context Using Word2vec". In: *Proceedings of the First International Workshop on Deep Learning and Music joint with IJCNN*. Anchorage, May 2017, pp. 11–18.

[110] W.B. Hewlett. "A base-40 number-line representation of musical pitch notation". In: *Musikometrika* 4 (1992), pp. 1–14.

[111] Tatsunori Hirai and Shun Sawada. "Melody2Vec: Distributed Representations of Melodic Phrases Based on Melody Segmentation". In: *Journal of Information Processing* 27 (2019), pp. 278–286.

[112] Ana Hofman. "Alan Lomax Archive". In: *Yearbook for Traditional Music* 44 (2012), pp. 228–229, 236.

[113] Henkjan Honing. "Issues on the Representation of Time and Structure in Music". In: *Contemporary music review* 9.1-2 (1993), pp. 221–238.

[114] Mantle Hood. *The Ethnomusicologist*. Kent, Ohio: Kent State University Press, 1982.

[115] Holger H. Hoos et al. "Representing Score-Level Music Using the GUIDO Music-Notation Format". In: *Computing in Musicology* 12 (2001).

[116] Erich M. von Hornbostel and Curt Sachs. "Systematik der Musikinstrumente. Ein Versuch". In: *Zeitschrift für Ethnologie* 46.4-5 (1914), pp. 553–590.

[117] Tongbo Huang et al. "MidiFind: Fast and Effective Similarity Searching in Large MIDI Databases". In: *Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research*. 2013, pp. 209–224.

[118] Johan Huizinga. *Homo Ludens: A Study of the Play-Element in Cultur*. London, United Kingdom: Routledge, 1949.

[119] David Huron. "Music Information Processing Using the Humdrum Toolkit: Concepts, Examples, and Lessons". In: *Computer Music Journal* 26.2 (2002), pp. 11–26.

[120] David Huron. "The Melodic Arch in Western Folksongs". In: *Computing in Musicology* 10 (1996), pp. 3–23.

[121] David Huron. “The New Empiricism: Systematic Musicology in a Postmodern Age”. In: *The 1999 Ernest Bloch Lectures*. Berkeley, University of California, 1999, pp. 2–32.

[122] Charles Inskip and Frans Wiering. “In Their Own Words: Using Text Analysis to Identify Musicologists’ Attitudes Towards Technology”. In: *International Society for Music Information Retrieval Conference (ISMIR 2015)*. 2015, pp. 455–461.

[123] Berit Janssen, Peter van Kranenburg, and Anja Volk. “Finding Occurrences of Melodic Segments in Folk Songs Employing Symbolic Similarity Measures”. In: *Journal of New Music Research* 46.2 (2017), pp. 118–134.

[124] Knud Jeppesen. *The Style of Palestrina and the Dissonance*. engdan. New York: Dover Publications, 1970. ISBN: 978-0-486-22386-5.

[125] Anna Jordanous. “Four PPPPerspectives on Computational Creativity in Theory and in Practice”. In: *Connection Science* (2016).

[126] Altug Karakurt, Sertan Şentürk, and Xavier Serra. “MORTY: A Toolbox for Mode Recognition and Tonic Identification”. In: *3rd International Digital Libraries for Musicology Workshop*. New York, NY, 2016.

[127] MK Karaosmanoglu et al. “A Symbolic Dataset of Turkish Makam Music Phrases”. In: *Folk Music Analysis Workshop (FMA 2014)*. Istanbul, Turkey, 2014.

[128] Adele T Katz. “Heinrich Schenker’s method of analysis”. In: *The Musical Quarterly* 21.3 (1935), pp. 311–329.

[129] Marjetka Golež Kaučič. “Slovene Folk Song at the Crossroads of Influences, Contacts, and Oppositions of East, West, North, and South”. In: *Slovene Studies* 29.1-2 (Jan. 2007), pp. 3–18.

[130] Marjetka Golež Kaučič. *Slovenska ljudska balada*. Ljubljana, Slovenia: Založba ZRC, 2018.

[131] W. James Kent. “BLAT–the BLAST-like Alignment Tool”. In: *Genome Research* 12.4 (Apr. 2002), pp. 656–664.

[132] Joseph Kerman. “How We Got into Analysis, and How to Get Out”. In: *Critical Inquiry* 7.2 (1980), pp. 311–331.

[133] Andrew Killick. “Global Notation as a Tool for Cross-Cultural and Comparative Music Analysis”. In: *Analytical Approaches to World Music* 8.2 (2020), pp. 235–279.

[134] Andrew King. “Resources in the Vaughan Williams Memorial Library: The Ella Mary Leather manuscript collection”. In: *Folk Music Journal* (2010), pp. 749–812.

[135] Friedrich A Kittler. *Gramophone, Film, Typewriter*. Redwood City, Californias, USA: Stanford University Press, 1999.

[136] Marija Klobčar. “"Kako in kdaj so prišle žene do tega, da pojejo pesmi tudi o takih junakih, kakor sta Pegam in Lambergar?": Matija Murko in vprašanje nosilcev pripovednih pesmi”. In: *Glasnik Slovenskega etnološkega društva* 54.3 (2014), pp. 21–29.

[137] Donald E Knuth, James H Morris Jr, and Vaughan Pratt. “Fast Pattern Matching in Strings”. In: *SIAM journal on computing* 6.2 (1977), pp. 323–350.

[138] Gopala Krishna Koduri et al. “Intonation Analysis of Rāgas in Carnatic Music”. In: *Journal of New Music Research* 43.01 (2014), pp. 73–94.

[139] David A Kolb. “Learning Styles and Disciplinary Differences”. In: *The modern American college* 1 (1981), pp. 232–235.

[140] Tina Koren. “Pevski repertoar v drugem vzgojno-izobraževalnem obdobju osnovne šole”. Slovenian. Published in Repository of the University of Ljubljana (RUL). Master’s thesis. Ljubljana, Slovenia: University of Ljubljana, Faculty of Education, 2019. URL: http://pefprints.pef.uni-lj.si/6081/.

[141] Kevin Korsyn. *Decentering Music: A Critique of Contemporary Musical Research*. Oxford, United Kingdom: Oxford University Press, 2003.

[142] Mojca Kovačič. *Glasbena podoba ljudske pesmi v rokopisnih, tiskanih in zvočnih virih v prvih desetletjih 20. stoletja*. Založba Univerze v Ljubljani, 2015.

[143] Peter van Kranenburg. “Computational Approach to Content-based Retrieval of Folk Song Melodies”. PhD thesis. Utrecht University, 2010.

[144] Peter van Kranenburg, Martine de Bruin, and Anja Volk. “Documenting a Song Culture: The Dutch Song Database as a Resource for Musicological Research”. In: *International Journal on Digital Libraries* 20.1 (Mar. 2019), pp. 13–23.

[145] Peter van Kranenburg and Darrell Conklin. “A Pattern Mining Approach to Study a Collection of Dutch Folk-Songs”. In: *Proceedings of the Sixth Workshop on Folk Music Analysis (FMA 2016)*. 2016, pp. 71–73.

[146] Peter van Kranenburg, Martine De Bruin, and Anja Volk. “Documenting a song culture: the Dutch Song Database as a resource for musicological research”. In: *International Journal on Digital Libraries* 20.1 (2019), pp. 13–23.

[147] Peter van Kranenburg and Folgert 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.

[148] Peter van Kranenburg et al. “Collaboration perspectives for folk song research and music information retrieval: The indispensable role of computational musicology”. In: *Journées d’Informatique Musicale (JIM 2010)* 2009 (2010), p. 030.

[149] Peter Van Kranenburg and Remco C Veltkamp. “Musical Models for Melody Alignment”. In: *International Society for Music Information Retrieval Conference (ISMIR 2009)*. Kobe, Japan, 2009, pp. 507–512.

[150] Franjo Ksaver Kuhač. *Južno-slovjenske narodne popievke*. Zagreb, Croatia: Tiskara i litografija C. Albrechta, 1878.

[151] Thomas S. Kuhn. *The Structure of Scientific Revolutions*. Second edition. Chicago, Illinois, USA: The University of Chicago, 1970.

[152] Vijay Kumar, Harit Pandya, and C.V. Jawahar. “Identifying Ragas in Indian Music”. In: *22nd International Conference on Pattern Recognition*. Stockholm, Sweden, 2014, pp. 767–772.

[153] Zmaga Kumer. “Pogled na dosedanje delo baladne komisije”. In: *Ljudske balade med izročilom in sodobnostjo*. Ed. by Marjetka Golež. Ljubljana, Slovenia: Založba ZRC, 1998, pp. 31–32.

[154] Zmaga Kumer. “Slovenske ljudske pesmi z napevi: poročilo o glasbenem gradivu, nabranem 1906-1914 pod Štrekljevim vodstvom, zdaj v Glasbeno narodopisnem inštitutu v Ljubljani”. In: *Slovenski etnograf* 12 (1959).

[155] Zmaga Kumer. *Vloga, zgradba, slog slovenske ljudske pesmi*. Ljubljana: ZRC SAZU, Založba ZRC, Feb. 1996. ISBN: 978-961-6182-11-9. DOI: 10.3986/9616182110. URL: https://omp.zrc-sazu.si/zalozba/catalog/book/1748 (visited on 11/13/2023).

[156] Zmaga Kumer et al. *Slovenske Ljudske Pesmi I: Pripovedne Pesmi*. Ljubljana: ČGP Delo, 1970.

[157] Zmaga Kumer et al. *Slovenske ljudske pesmi I: Pripovedni pesmi*. Vol. 1. Ljubljana: CGP Delo, 1970.

[158] Sonja Kump. *Akademska kultura*. Ljubljana, Slovenia: Znanstveno in publicisti po središče, 1994.

[159] Drago Kunej. "Prva Magnetofonska Snemanja Zvočnega Gradiva Glasbenonarodopisnega Inštituta". In: *Traditiones* 28.2 (1999), pp. 217–232.

[160] Drago Kunej. "Sound Recordings and Karel Štrekelj: The Initiator of a New Approach to Folk Song Research in Slovenia". In: *Muzikologija-Musicology* 33 (Dec. 2022), pp. 39–52.

[161] Drago Kunej. ""We Have Plenty of Words Written Down; We Need Melodies!" the Purchase of the First Recording Device for Ethnomusicological Research in Slovenia". In: *Traditiones* 34.1 (2005), pp. 125–140.

[162] Drago Kunej and Rebeka Kunej. *Protokol za lastnike zvočnih zbirk*. 2020. URL: `ttps://fmh.zrc-sazu.si/wp-content/uploads/2020/11/Protokol_SLO_final.pdf` (visited on 10/07/2023).

[163] Jom Kuriakose et al. "Akshara Transcription of Mrudangam Strokes in Carnatic Music". In: *Proceedings of the 21st National Conference on Communication*. Mumbai, India, 2015.

[164] Olivier Lartillot. "Automated Motivic Analysis: An Exhaustive Approach Based on Closed and Cyclic Pattern Mining in Multidimensional Parametric Spaces". In: *Computational Music Analysis*. Springer International Publishing, 2015, pp. 273–302.

[165] Olivier Lartillot. "In-Depth Motivic Analysis Based on Multiparametric Closed Pattern and Cyclic Sequence Mining". In: *International Society for Music Information Retrieval Conference (ISMIR 2014)*. Taipei, Taiwan, 2014.

[166] Olivier Lartillot. "Motivic Pattern Mining". In: *Knowledge representation for intelligent music processing. Dagstuhl Seminar Proceeding*. Vol. 9051. Schloss Dagstuhl - Leibniz-Zentrum für Informatik. 2009, pp. 1–4.

[167] Olivier Lartillot. "Multi-Dimensional Motivic Pattern Extraction Founded on Adaptive Redundancy Filtering". In: *Journal of New Music Research* 34.4 (2005), pp. 375–393.

[168] Felicity Laurence. "Children's singing". In: *The Cambridge Companion to Singing*. Ed. by John Potter. Cambridge Companions to Music. Cambridge University Press, 2000, pp. 221–230.

[169] Deborah Lee. "Hornbostel-Sachs Classification of Musical Instruments". In: *Knowledge Organization* 47.1 (2019), pp. 72–91.

[170] Marc Leman. "Systematic Musicology at the Crossroads of Modern Music Research". In: *Systematic and comparative musicology: Concepts, methods, findings*. Peter Lang, 2008, pp. 89–115.

[171] Kjell Lemström, Niko Mikkilä, and Veli Mäkinen. "Filtering Methods for Content-Based Retrieval on Indexed Symbolic Music Databases". In: *Information retrieval* 13 (2010), pp. 1–21.

[172] Fred Lerdahl and Ray S Jackendoff. *A Generative Theory of Tonal Music*. Cambridge, Massachusetts: MIT press, 1996.

[173] Andre Leroi-Gourhan. *Gib in beseda II*. Ljubljana: Studia Humanitatis, 1990.

[174] Vladimir I Levenshtein. "Binary Codes Capable of Correcting Deletions, Insertions, and Reversals". In: *Soviet physics doklady*. Vol. 10. 8. Soviet Union, 1966, pp. 707–710.

[175] Ivan Lešnik. "Children's Singing in Slovenia: Vocal Range Analyses". In: *Contemporary Approaches to Music Teaching and Learning*. 2015, pp. 127–141.

[176] Thomas Lidy et al. "On the Suitability of State-of-the-Art Music Information Retrieval Methods for Analyzing, Categorizing, and Accessing Non-Western and Ethnic Music Collections". In: *Signal Processing*. Special Section: Ethnic Music Audio Documents: From the Preservation to the Fruition 90.4 (Apr. 2010), pp. 1032–1048.

[177] David J. Lipman and William R. Pearson. "Rapid and Sensitive Protein Similarity Searches". In: *Science* 227.4693 (1985), pp. 1435–1441.

[178] George List. "The Reliability of Transcription". In: *Ethnomusicology* 18.3 (Sept. 1974), pp. 353–377.

[179] Alan Lomax. *Cantometrics: An Approach to the Anthropology of Music*. University of California Extension Media Center, 1978.

[180] Alan Lomax. *Folk Song Style and Culture*. Piscataway, New Jersey, USA: Transaction Publishers, 1968.

[181] Alan Lomax and Norman Berkowitz. "The Evolutionary Taxonomy of Culture: A Few Behavioral Factors Account for the Regional Variation and Evolutionary Development of Culture". In: *Science* 177.4045 (1972), pp. 228–239.

[182] Marcelo E Rodrıguez López, Anja Volk, and W Bas de Haas. "Comparing Repetition-Based Melody Segmentation Models". In: *Proceedings of the 9th Conference on Interdisciplinary Musicology (CIM14), SIMPK and ICCMR*. 2014, pp. 143–148.

[183] Anna Lubiw and Luke Tanur. "Pattern Matching in Polyphonic Music as a Weighted Geometric Translation Problem." In: *International Conference on Music Information Retrieval (ISMIR 2004)*. Barcelona, Spain, 2004.

[184] Lumen Learning. *Children Singing and Children's Songs*. Accessed: 2024-08-12. n.d. URL: https://courses.lumenlearning.com/suny-music-and-the-child/chapter/chapter-5-children-singing-and-childrens-songs/.

[185] David M. Weigl et al. "Interweaving and Enriching Digital Music Collections for Scholarship, Performance, and Enjoyment". In: *6th International Conference on Digital Libraries for Musicology*. The Hague Netherlands: ACM, Nov. 2019, pp. 84–88.

[186] Bin Ma, John Tromp, and Ming Li. "PatternHunter: Faster and More Sensitive Homology Search". In: *Bioinformatics* 18.3 (2002), pp. 440–445.

[187] Tom Madden. "The BLAST sequence analysis tool". In: *The NCBI handbook* (2003).

[188] Thor Magnusson. *Sonic writing: technologies of material, symbolic, and signal inscriptions*. Bloomsbury Publishing, 2019.

[189] Thor Magnusson. *Sonic Writing: Technologies of Material, Symbolic, and Signal Inscriptions*. Jan. 2019.

[190] Dimos Makris, Ioannis Karydis, and Spyros Sioutas. "The Greek Music Dataset". In: *Proceedings of the 16th International Conference on Engineering Applications of Neural Networks (INNS)*. EANN '15. New York, NY: Association for Computing Machinery, Sept. 2015, pp. 1–7.

[191] Lev Manovich. *The Language of New Media*. Cambridge, Massachusetts: MIT press, 2002.

[192] Giovanni Manzini. "The Burrows-Wheeler Transform: Theory and Practice: Invited Lecture". In: *International Symposium on Mathematical Foundations of Computer Science*. Springer. 1999, pp. 34–47.

[193] Matija Marolt et al. "Automatic Segmentation of Ethnomusicological Field Recordings". In: *Applied Sciences* 9.3 (2019), p. 439.

[194] Alan Marsden. "Automatic Derivation of Musical Structure: A Tool for Research on Schenkerian Analysis." In: *International Conference on Music Information Retrieval (ISMIR 2007)*. Vienna, Austria, 2007, pp. 55–58.

[195] Alan Marsden. "Music Analysis by Computer: Ontology and Epistemology". In: *Computational Music Analysis*. Springer, 2016, pp. 3–28.

[196] Scott L. Matthews. "The Alan Lomax Archive". In: *American Quarterly* 68.2 (2016), pp. 429–437.

[197] Edward M McCreight. "A Space-Economical Suffix Tree Construction Algorithm". In: *Journal of the ACM (JACM)* 23.2 (1976), pp. 262–272.

[198] Brian McFee et al. "librosa: Audio and Music Signal Analysis in Python". In: *Python in Science Conference (SCIPY 2015)*. Vol. 8. 2015, pp. 18–25.

[199] Scott McGinnis and Thomas L. Madden. "BLAST: At the Core of a Powerful and Diverse Set of Sequence Analysis Tools". In: *Nucleic Acids Research* 32.suppl$_2$ (2004), W20–W25.

[200] Deborah L McGuinness, Frank Van Harmelen, et al. "OWL Web Ontology Language Overview". In: *W3C recommendation* 10.10 (2004).

[201] Cory McKay, Julie Cumming, and Ichiro Fujinaga. "jSymbolic 2.2: Extracting Features from Symbolic Music for uUe in Musicological and MIR Research". In: *International Society for Music Information Retrieval Conference (ISMIR 2018)*. Paris, France, 2018, pp. 348–354.

[202] D Meredith, K Lemström, and GA Wiggins. "Algorithms for Discovery Repeated Patterns in Multidimensional Representations of Polyphonic Music". In: *Journal of New Music Research* 31.4 (2003), pp. 321–354.

[203] Dave Meredith, Geraint A Wiggins, and Kjell Lemström. "Pattern induction and matching in polyphonic music and other multidimensional datasets". In: *Proceedings of the 5th World Multiconference on Systemics, Cybernetics and Informatics (SCI2001)*. 2001, pp. 22–25.

[204] David Meredith. "Point-Set Algorithms for Pattern Discovery and Pattern Matching in Music". In: *Dagstuhl Seminar on Content-based Retrieval*. Vol. 06171. 2006, pp. 23–28.

[205] Alan P. Merriam. *The Anthropology of Music*. Northwestern University Press, 1964.

[206] Isabelle Mills. "The Heart of the Folk Song". In: *Canadian Folk Music Journal* 2 (1974), pp. 29–34.

[207] Marcel Mongeau and David Sankoff. "Comparison of Musical Sequences". In: *Computers and the Humanities* 24.3 (1990), pp. 161–175.

[208] Bhavya Mor, Sunita Garhwal, and Ajay Kumar. "A Systematic Literature Review on Computational Musicology". In: *Archives of Computational Methods in Engineering* 27 (2020), pp. 923–937.

[209] Franco Moretti. *Distant Reading*. London, United Kingdom: Verso Books, 2013.

[210] Fabio Morreale. "Where Does the Buck Stop? Ethical and Political Issues with AI in Music Creation". In: 4 (1 2021), pp. 105–113.

[211] Tessa Morris-Suzuki. *Beyond Computopia: Information, Automation and Democracy in Japan*. First issued in paperback. Japanese Studies. London New York: Routledge, 2014.

[212] Stefan Münnich. "FAIR for Whom? Commentary on Hofmann et al. (2021)". In: *Empirical Musicology Review* 16.1 (2021), pp. 151–153.

[213] Matija Murko. "Velika zbirka slovenskih narodnih pesmi z melodijami". In: *Etnolog (Ljubljana)* 3 (1929), pp. 5–54.

[214] Shreyas Nadkarni et al. "Exploring the Correspondence of Melodic Contour With Gesture in Raga Alap Singing". In: *International Society for Music Information Retrieval Conference (ISMIR 2023)*. Milan, Italy, 2023.

[215] Saul B Needleman and Christian D Wunsch. "A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins". In: *Journal of Molecular Biology* 48.3 (1970), pp. 443–453.

[216] Bruno Nettl. *Folk and Traditional Music of the Western Continents*. Prentice-Hall, 1960.

[217] Kerstin Neubarth, Mathieu Bergeron, and Darrell Conklin. "Associations Between Musicology and Music Information Retrieval". In: *International Society for Music Information Retrieval Conference (ISMIR 2011)*. Miami, Florida, USA, 2011, pp. 429–434.

[218] Kerstin Neubarth and Darrell Conklin. "Contrast Pattern Mining in Folk Music Analysis". In: *Computational Music Analysis*. Springer, 2016, pp. 393–424.

[219] Kerstin Neubarth et al. "Association Mining of Folk Music Genres and Toponyms". In: *International Society for Music Information Retrieval Conference (ISMIR 2012)* (2012), pp. 7–12.

[220] Han-Wen Nienhuys and J. Nieuwenhuizen. "Lilypond, a System for Automated Music Engraving". In: *XIV Colloquium on Musical Informatics*. Firenze, Italy, May 2003, pp. 1–6.

[221] Danica Š. Novosel. "Ob 50-letnici Cicibana". In: *Otrok in knjiga* 22.39/40 (1995), pp. 120–121.

[222] Thomas Nuttall et al. "Contributing to New Musicological Theories with Computational Methods: The Case of Centonization in Arab-Andalusian Music". In: *International Society for Music Information Retrieval Conference (ISMIR 2019)*. Delft, Netherlands, Nov. 2019, pp. 223–8.

[223] Sergio de la Ossa. *A Basic Guide to Folksong Analysis*. Budapest, Hungary: Liszt Academy of Music, 2019.

[224] Yuto Ozaki et al. "Similarities and Differences in a Global Sample of Song and Speech Recordings [Stage 2 Registered Report]". In: *PsyArXiv (preprint)* (2022).

[225] Maria Panteli. "Computational Analysis of World Music Corpora". PhD Thesis. Queen Mary University of London, School of Electronic Engineering and Computer Science, 2018.

[226] Maria Panteli, Emmanouil Benetos, and Simon Dixon. "A Review of Manual and Computational Approaches for the Study of World Music Corpora". In: *Journal of New Music Research* 47.2 (2018), pp. 176–189.

[227] Carl Frederick Abel Pantin. *Relations between sciences*. Cambridge University Press, 1968.

[228] Charilaos Papaioannou et al. "A Dataset for Greek Traditional and Folk Music: Lyra". In: *International Society for Music Information Retrieval Conference (ISMIR 2022)*. Bengaluru, India, 2022.

[229] Nikos Papasarantopoulos, Nick Poulakis, and Christina Anagnostopoulou. "A Computational Analysis Study of Children's Songs from Different Countries". In: *Proceedings of the Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music*. Ed. by J. Ginsborg et al. Manchester, United Kingdom, 2015, pp. 626–630.

[230] Richard Parncutt. "Systematic Musicology and the History and Future of Western Musical Scholarship". In: *Journal of Interdisciplinary Music Studies* 1.1 (2007), pp. 1–32.

[231] Marcus Pearce and Geraint Wiggins. "Improved Methods for Statistical Modelling of Monophonic Music". In: *Journal of New Music Research* 33.4 (2004), pp. 367–385.

[232] Ewald Peiszer, Thomas Lidy, and Andreas Rauber. "Automatic Audio Segmentation: Segment Boundary and Structure Detection in Popular Music". In: *Proceedings of the 2nd International Workshop on Learning the Semantics of Audio Signals (LSAS)*. Vol. 106. 2008, pp. 45–59.

[233] Celia Pendlebury. "Tune Families and Tune Histories: Melodic Resemblances in British and Irish folk tunes". In: *Folk Music Journal* 11.5 (2020), pp. 67–158.

[234] Matevž Pesek, Aleš Leonardis, and Matija Marolt. "SymCHM—An Unsupervised Approach for Pattern Discovery in Symbolic Music with a Compositional Hierarchical Model". In: *Applied Sciences* 7.11 (Nov. 2017), p. 1135.

[235] Jeremy Pickens and Tim Crawford. "Harmonic Models for Polyphonic Music Retrieval". In: *Proceedings of the Eleventh International Conference on Information and Knowledge Management*. 2002, pp. 430–437.

[236] Alastair Porter and Xavier Serra. "An Analysis and Storage System for Music Research Datasets". In: *Proceedings of the 1st International Workshop on Digital Libraries for Musicology*. London, United Kingdom: ACM, Sept. 2014, pp. 1–3.

[237] Alastair Porter, Mohamed Sordo, and Xavier Serra. "Dunya: A System for Browsing Audio Music Collections Exploiting Cultural Context". In: *International Society for Music Information Retrieval Conference (ISMIR 2013)*. Curitiba, PR, Brazil, Nov. 2013, pp. 101–6.

[238] Polina Proutskova et al. "From Music Ontology Towards Ethno-Music-Onthology". In: *International Society for Music Information Retrieval Conference (ISMIR 2020)*. Online, Nov. 2020, pp. 923–931.

[239] Laurent Pugin. "The Challenge of Data in Digital Musicology". In: *Frontiers in Digital Humanities* 2 (2015), p. 4.

[240] Michael O Rabin and Richard M Karp. "Efficient Randomized Algorithms for String Matching". In: *Journal of the ACM (JACM)* 31.1 (1984), pp. 16–36.

[241] Yves Raimond and Mark Sandler. "Evaluation of the Music Ontology Framework". In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 7295. 2012, pp. 255–269.

[242] Yves Raimond et al. "The Music Ontology". In: *International Conference on Music Information Retrieval (ISMIR 2007)*. Vienna, Austria, Sept. 2007.

[243] Iris Yuping Ren. "Closed Patterns in Folk Music and Other Genres". In: *Proceedings of the 6th International Workshop on Folk Music Analysis (FMA 2016)*. 2016, pp. 56–58.

[244] Iris Yuping Ren et al. "Analysis by Classification: A Comparative Study of Annotated and Algorithmically Extracted Patterns in Symbolic Music Data". In: *19th International Conference on Music Information Retrieval*. Paris, France, Sept. 2018, pp. 539–546.

[245] Iris Yuping Ren et al. "In Search of the Consensus Among Musical Pattern Discovery Algorithms". In: *International Society for Music Information Retrieval Conference (ISMIR 2017)*. Suzhou, China, 2017, pp. 671–678.

[246] Kai Renz. "Algorithms and Data Structures for a Music Notation System based on GUIDO Music Notation". phd. Darmstadt: Technische Universität, Oct. 2002.

[247] Rafael Caro Repetto. *Jingju Lyrics Datasets (1.0)*. 2017. URL: https://doi.org/10.5281/zenodo.1285632 (visited on 08/08/2024).

[248] Rafael Caro Repetto and Xavier Serra. "A Collection of Music Scores for Corpus Based Jingju Singing Research". In: *International Society for Music Information Retrieval Conference (ISMIR 2017)*. Suzhou, China, 2017, pp. 46–52.

[249] Timothy Rice. *Modeling Ethnomusicology*. Oxford, New York: Oxford University Press, 2017.

[250] David Rizo and Alan Marsden. "An MEI-Based Standard Encoding for Hierarchical Music Analyses". In: *International Journal on Digital Libraries* 20.1 (Mar. 2019), pp. 93–105.

[251] Tatiana Rocher, Mathieu Giraud, and Mikaël Salson. "Indexing Labeled Sequences". In: *PeerJ Computer Science* 4 (2018), pp. 1–14.

[252] Bruce A. Rosenberg. *Narrative Folksong, New Directions: Essays in Appreciation of W. Edson Richmond*. First Edition. Boulder, Colorado, USA: Westview Press, 1985.

[253] Sebastian Rosenzweig et al. "Erkomaishvili Dataset: A Curated Corpus of Traditional Georgian Vocal Music for Computational Musicology". In: *Transactions of the International Society for Music Information Retrieval* 3 (Apr. 2020), pp. 31–41.

[254] Joseph Rothstein. *MIDI: A Comprehensive Introduction*. Middleton, Wisconsin, USA: A-R Editions, Inc., May 1992.

[255] Mikael Salson et al. "A Four-Stage Algorithm for Updating a Burrows–Wheeler Transform". In: *Theoretical Computer Science* 410.43 (2009), pp. 4350–4359.

[256] Carolina L. dos Santos and Carlos Nascimento Silla. "The Latin Music Mood Database". In: *EURASIP Journal on Audio, Speech, and Music Processing* 2015.1 (Aug. 2015), p. 23.

[257] Craig Stuart Sapp. *Computational Methods for the Analysis of Musical Structure*. Stanford, California, USA: Stanford University, 2011.

[258] Craig Stuart Sapp. "Online Database of Scores in the Humdrum File Format". In: *International Conference on Music Information Retrieval (ISMIR 2005)*. London, United Kingdom, Sept. 2005, pp. 664–665.

[259] Natalie Sarrazin. "Music and the Child". In: Milne Publishing, 2016. Chap. 5. URL: https://milnepublishing.geneseo.edu/music-and-the-child/chapter/chapter-5/.

[260] Patrick E Savage. "Alan Lomax's Cantometrics Project: A Comprehensive Review". In: *Music & Science* 1 (2018).

[261] Patrick E Savage. *Measuring the Cultural Evolution of Music: Cross-Cultural and Cross-Genre Case Studies*. Oct. 2020.

[262] Patrick E Savage and Quentin D Atkinson. "Automatic Tune Family Identification by Musical Sequence Alignment". In: *International Society for Music Information Retrieval Conference (ISMIR 2015)*. Vol. 163. Málaga, Spain, 2015.

[263] Patrick E. Savage et al. "CantoCore: A New Cross-cultural Song Classification Scheme". In: *Analytical Approaches to World Music* 2.1 (2012), pp. 87–137.

[264] Patrick E Savage et al. "Sequence Alignment of Folk Song Melodies Reveals Cross-Cultural Regularities of Musical Evolution". In: *Current Biology* 32.6 (2022), pp. 1395–1402.

[265] Helmut Schaffrath. "Essen Corpus of German Folksong Melodies". In: *Oxford Text Archive Core Collection* (1987).

[266] Markus Schedl, Emilia Gómez, Julián Urbano, et al. "Music Information Retrieval: Recent Developments and Applications". In: *Foundations and Trends in Information Retrieval* 8.2-3 (2014), pp. 127–261.

[267] Markus Schedl, Emilia Gómez, and Julián Urbano. "Music Information Retrieval: Recent Developments and Applications". In: *Foundations and Trends in Information Retrieval* 8.2-3 (2014), pp. 127–261.

[268] Frank Scherbaum, Nana Mzhavanadze, and Sebastian Rosenzweig. "Mulit-media Recordings of Traditional Georgian Vocal Music for Computational Analysis". In: *The 9th International Workshop on Folk Music Analysis (FMA 2019)*. 2019.

[269] Nico Schüler. "Reflections on the History of Computer-Assisted Music Analysis I: Predecessors and the Beginnings". In: *Musicological Annual* 41.1 (2005), pp. 31–43.

[270] Anthony Seeger. "Ethnomusicology and Music Law". In: *Ethnomusicology* 36.3 (1992), pp. 345–359.

[271] Xavier Serra. "The Computational Study of a Musical Culture Through Its Digital Traces". In: *Acta Musicologica* 89.1 (2017), pp. 24–44.

[272] Michel Serres. *Thumbelina: The Culture and Technology of Millennials*. Rowman & Littlefield International, 2015.

[273] William R. Shadish, Thomas D. Cook, and Donald T. Campbell. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. 2nd. Boston: Cengage Learning, 2001.

[274] Cecil J Sharp and Olive Dame Campbell. *English Folk Songs from the Southern Appalachians*. Oxford University Press, 1917.

[275] Carlos Nascimento Silla, Alessandro Lameiras Koerich, and Celso AA Kaestner. "The Latin Music Database". In: *International Conference on Music Information Retrieval (ISMIR 2008)*. Philadelphia, Pennsylvania, 2008, pp. 451–456.

[276] Carlos Nascimento Silla Jr, Alessandro L. Koerich, and Celso AA Kaestner. "The Latin Music Database." In: *9th International Society for Music Information Retrieval Conference*. Drexel University, Philadelphia, Sept. 2008, pp. 451–456.

[277] Gilbert Simondon. *On the Mode of Existence of Technical Objects*. First University of Minnesota Press. Minneapolis University of Minnesota Press, 2017.

[278] Jared T Simpson and Richard Durbin. "Efficient Construction of an Assembly String Graph Using the FM-Index". In: *Bioinformatics* 26.12 (2010), pp. i367–i373.

[279] Yeshwant Singh et al. *Indian Folk Music Dataset*. May 2022. URL: https://zenodo.org/record/6584021 (visited on 07/11/2023).

[280] Temple F Smith and Michael S Waterman. "Identification of Common Molecular Subsequences". In: *Journal of Molecular Biology* 147.1 (1981), pp. 195–197.

[281] Mohamed Sordo et al. "A Musically Aware System for Browsing and Interacting with Audio Music Collections". In: *Proceedings of 2nd CompMusic Workshop*. Istanbul, Turkey, 2012, pp. 20–24.

[282] Ajay Srinivasamurthy, Andre Holzapfel, and Xavier Serra. "Informed Automatic Meter Analysis of Music Recordings". In: *International Society for Music Information Retrieval Conference (ISMIR 2017)*. Suzhou, China, 2017, pp. 679–685.

[283] Ajay Srinivasamurthy and Xavier Serra. "A Supervised Approach to Hierarchical Metrical Cycle Tracking from Audio Music Recordings". In: *Proceedings of the 39th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)*. Florence, Italy, 2014, pp. 5237–5241.

[284] Ajay Srinivasamurthy et al. "Saraga: Open Datasets for Research on Indian Art Music". In: *Empirical Musicology Review* 16.1 (2021), pp. 85–98.

[285] Ajay Srinivasamurthy et al. "Transcription and Recognition of Syllable based Percussion Patterns: The Case of Beijing Opera". In: *International Society for Music Information Retrieval Conference (ISMIR 2014)*. Taipei, Taiwan, 2014.

[286] Ajay Srinivasasmurthy et al. "A Generalized Bayesian Model for Tracking Long Metrical Cycles in Acoustic Music Signals". In: *Proceedings of the 41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016)*. Shanghai, China, 2016.

[287] Star Musicals Press. *Kids' Vocal Range - What Is the Best Range for Kids to Sing?* n.d. URL: https://www.starmusicalspress.com/kids-vocal-range/.

[288] Leon Stefanija et al. *Slovenska mladinska in otroška glasba 1945-1991*. https://korpusi.musiclab.si/. Accessed: 2024-09-24. 2024.

[289] Leon Stefanija et al. "Zgodovina in izzivi digitalne etno/muzikologije v Sloveniji". In: *Musicological Annual* 58.2 (2022), pp. 15–49.

[290] Estera Stojko. "Pevski repertoar v prvem vzgojno-izobraževalnem obdobju osnovne šole: magistrsko delo". Master's thesis. University of Ljubljana, Faculty of Education, 2016.

[291] Gregor Strle and Matija Marolt. “Computational Folkloristics: A Semantic Analysis and Visualization of Topic Distribution of Song Types”. In: *Glasnik SED* 54 (Jan. 2015), pp. 36–43.

[292] Gregor Strle and Matija Marolt. “The EthnoMuse Digital Library: Conceptual Representation and Annotation of Ethnomusicological Materials”. en. In: *International Journal on Digital Libraries* 12.2-3 (Aug. 2012), pp. 105–119.

[293] Bob Sturm and Arthur Flexer. “A Review of Validity and its Relationship to Music Information Research”. In: *International Society for Music Information Retrieval Conference (ISMIR 2023)*. Milan, Italy, 2023.

[294] Wai Man Szeto and Man Hon Wong. “A Graph-Theoretical Approach for Pattern Matching in Post-Tonal Music Analysis”. In: *Journal of New Music Research* 35.4 (2006), pp. 307–321.

[295] David Temperley. “Meter and Grouping in African Music: A View from Music Theory”. In: *Ethnomusicology* 44.1 (2000), pp. 65–96. ISSN: 0014-1836. DOI: 10.2307/852655. URL: https://www.jstor.org/stable/852655 (visited on 07/15/2023).

[296] Marko Terseglav. *Ljudsko pesništvo*. Literarni leksikon: Študije 32. Ljubljana: Državna založba Slovenije, 1987.

[297] Mi Tian et al. “A Study of Instrument-wise Onset Detection in Beijing Opera Percussion Ensembles”. In: *Proceedings of ICASSP 2014*. Florence, Italy, 2014.

[298] Mi Tian et al. “Towards the Representation of Chinese Traditional Music: A State of the Art Review of Music Metadata Standards”. In: *International Conference on Dublin Core and Metadata Applications*. Lisbon, Portugal, Sept. 2013, pp. 71–81.

[299] Darian Tomašević et al. “Exploring Annotations for Musical Pattern Discovery Gathered with Digital Annotation Tools”. In: *Journal of Mathematics and Music* 15.2 (2021), pp. 194–207.

[300] Yuen-Hsien Tseng. “Content-based Retrieval for Music Collections”. In: *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. 1999, pp. 176–182.

[301] Julián Urbano, Markus Schedl, and Xavier Serra. “Evaluation in Music Information Retrieval”. In: *Journal of Intelligent Information Systems* 41.3 (2013), pp. 345–369.

[302] Michelle Urberg. “Pasts and Futures of Digital Humanities in Musicology: Moving Towards a ”Bigger Tent””. In: *Music Reference Services Quarterly* 20.3-4 (2017), pp. 134–150.

[303] Peter Van Kranenburg and Eoin Kearns. “Cross-Corpus Melodic Similarity For Enriching Archival Collections”. In: *Proceedings of the 10th International Conference on Digital Libraries for Musicology*. New York, NY, USA: Digital Libraries for Musicology (DLfM 2023), Nov. 2023, p. 1.

[304] Peter Van Kranenburg, Anja Volk, and Frans Wiering. “AA Comparison Between Global and Local Features for Computational Classification of Folk Song Melodies”. In: *Journal of New Music Research* 42.1 (2013), pp. 1–18.

[305] Peter Van Kranenburg et al. “Towards Integration of MIR and Folk Song Research.” en. In: *International Conference on Music Information Retrieval (ISMIR 2007)*. Vienna, Austria, Sept. 2007, pp. 505–508.

[306] Hugues Vinet. “The Representation Levels of Music Information”. In: *International Symposium on Computer Music Modeling and Retrieval*. Springer. 2003, pp. 193–209.

[307] Jasmina Vodeb Baša. “Ciciban razveseljuje otroke že 70 let”. In: *Regionalna obala* (2014). URL: `https://regionalobala.si/novica/ciciban-razveseljuje-otroke-ze-70-let` (visited on 08/08/2024).

[308] Valens Vodušek. *Etnomuzikološki članki in razprave*. Ed. by Marko Terseglav and Robert Vrčon. Folkloristika 1. Založba ZRC, Jan. 2003.

[309] Aanja Volk, W. Bas de Haas, and Peter van Kranenburg. “Towards Modelling Variation in Music as Foundation for Similarity”. In: *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music*. Thessaloniki, Greece, Sept. 2012, pp. 1085–1094.

[310] Anja Volk and Peter van Kranenburg. “Melodic Similarity Among Folk Songs: An Annotation Study on Similarity-Based Categorization in Music”. In: *Musicae Scientiae* 16.3 (2012), pp. 317–339.

[311] Anja Volk, Frans Wiering, and Peter van sKranenburg. “Unfolding the Potential of Computational Musicology”. In: *Proceedings of the 13th International Conference on Informatics and Semiotics in Organisations*. 2011.

[312] Anja Volk et al. “A Manual Annotation Method for Melodic Similarity and the Study of Melody Feature Sets.” In: *International Conference on Music Information Retrieval (ISMIR 2008)*. Philadelphia, Pennsylvania, 2008, pp. 101–106.

[313] Anja Volk et al. “Applying Rhythmic Similarity Based on Innner Metric Analysis to Folksong Research”. In: *International Conference on Music Information Retrieval (ISMIR 2007)*. Vienna, Austria, 2007, pp. 293–296.

[314] Immanuel Wallerstein. *Open the Social Sciences: Report of the Gulbenkian Commission on the Restructuring of the Social Sciences*. Stanford, California, USA: Stanford University Press, 1996. ISBN: 978-0-8047-2727-3.

[315] David Weigl and Kevin Page. “A Framework for Distributed Semantic Annotation of Musical Score: ”Take It to the Bridge!”” In: *International Society for Music Information Retrieval Conference (ISMIR 2017)*. Suzhou, China, 2017.

[316] David Weigl et al. “FAIR Interconnection and Enrichment of Public-Domain Music Resources on the Web”. en. In: *Empirical Musicology Review* 16.1 (Dec. 2021), pp. 16–33. ISSN: 1559-5749. (Visited on 11/20/2023).

[317] David Weigl et al. “Notes on the Music: A social data infrastructure for music annotation”. In: *Proceedings of the 8th International Conference on Digital Libraries for Musicology (DLfM 2021)*. 2021, pp. 23–31.

[318] Frans Wiering and Emmanouil Benetos. “Digital Musicology and MIR: Papers, Projects and Challenges”. In: *International Society for Music Information Retrieval Conference (ISMIR 2013)*. Curitiba, PR, Brazil, 2013.

[319] Donald Knight Wilgus. “A Type-Index of Anglo-American Traditional Narrative Songs”. In: *Journal of the Folklore Institute* 7.2/3 (1970), pp. 161–176.

[320] Mark D. Wilkinson et al. “The FAIR Guiding Principles for scientific data management and stewardship”. In: *Scientific Data* 3.1 (Mar. 2016), p. 160018.

[321] Lyn A. Wolz. "Resources in the Vaughan Williams Memorial Library: The Anne Geddes Gilchrist Manuscript Collection". In: *Folk Music Journal* (2005), pp. 619–639.

[322] Anna Wood et al. "The Global Jukebox: A public database of performing arts and culture". In: *PLOS One* 17.11 (2021).

[323] Borut Žalik et al. "Can Burrows-Wheeler Transform be Replaced in Chain Code Compression?" In: *Information Sciences* 525 (2020), pp. 109–118.

[324] Tasos Zembylas, ed. *Kurt Blaukopf on Music Sociology: An Anthology*. Frankfurt am Main and elsewhere: Peter Lang, Apr. 2012.

[325] Shuo Zhang, Rafael Caro Repetto, and Xavier Serra. "Study of the Similarity between Linguistic Tones and Melodic Pitch Contours in Beijing Opera Singing". In: *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014)*. Taipei, Taiwan, 2014, pp. 343–348.

[326] Eve Zheng, Melody Moh, and Teng-Sheng Moh. "Music Genre Classification: A n-Gram Based Musicological Approach". In: *2017 IEEE 7th International Advance Computing Conference (IACC)*. IEEE. 2017, pp. 671–677.

[327] Tiange Zhu and Raphaël Fournier-S'niehotta. "FACETS: A Tool for Management and Navigation of Symbolic Music Collections". In: *International Society for Music Information Retrieval Conference (ISMIR 2011)*. Miami, Florida, USA, 2011.

[328] Sertan Şentürk, Andre Holzapfel, and Xavier Serra. "Linking Scores and Audio Recordings in Makam Music of Turkey". In: *Journal of New Music Research* 43 (2014), pp. 34–52.

[329] Sertan Şentürk and Xavier Serra. "Composition Identification in Ottoman-Turkish Makam Music Using Transposition-Invariant Partial Audio-Score Alignment". In: *Proceedings of 13th Sound and Music Computing Conference (SMC 2016)*. Hamburg, Germany, 2016, pp. 434–441.

[330] Karel Štrekelj. *Navodila in vprašanja za zbiranje in zapisovanje narodnih pesmi, narodne godbe, narodnih plesov in šeg, ki se nanašajo na to*. Ljubljana: Slovenski delovni odbor za publikacijo "Avstrijske narodne pesmi", 1906.

[331] Dragica Žvar. *Grlica za otroke: izbor otroških zborovskih pesmi revije Grlica od 1953 do 1988*. 1. izd., 1. natis. [Glasbeni tisk]. Ljubljana: Zavod Republike Slovenije za šolstvo, 2012.

[332] Dragica Žvar. "Izhodišča za izbiranje otroške zborovske pesmi pri obravnavi v šolskem pevskem zboru". In: *Glasba v šoli in vrtcu* XII.1 (2007), pp. 48–55.