MIRchiving: Challenges and opportunities of connecting MIR research and digital music archives

Reinier de Valk
DANS
reinier.de.valk@dans.knaw.nl

Anja Volk
Utrecht University
a.volk@uu.nl

Andre Holzapfel
KTH Royal Institute of Technology
holzap@kth.se

Aggelos Pikrakis
University of Piraeus
pikrakis@unipi.gr

Nadine Kroher
University of Seville - IMUS
nkroher@us.es

Joren Six
Ghent University - IPEM
joren.six@ugent.be

ABSTRACT

This study is a call for action for the music information retrieval (MIR) community to pay more attention to collaboration with digital music archives. The study, which resulted from an interdisciplinary workshop and subsequent discussion, matches the demand for MIR technologies from various archives with what is already supplied by the MIR community. We conclude that the expressed demands can only be served sustainably through closer collaborations. Whereas MIR systems are described in scientific publications, usable implementations are often absent. If there is a runnable system, user documentation is often sparse—posing a huge hurdle for archivists to employ it. This study sheds light on the current limitations and opportunities of MIR research in the context of music archives by means of examples, and highlights available tools. As a basic guideline for collaboration, we propose to interpret MIR research as part of a value chain. We identify the following benefits of collaboration between MIR researchers and music archives: new perspectives for content access in archives, more diverse evaluation data and methods, and a more application-oriented MIR research workflow.

KEYWORDS

Applying MIR, documentation, collaboration, digital music archives

1 INTRODUCTION

Research in MIR has resulted in the development of a variety of tools for the extraction of information from musical data. This data may take the shape of audio signals, machine-readable notation and graphic scores (*symbolic* data), or audiovisual recordings. Common is the processing of music as a time series, with the aim to extract information from the signal. Time series analysis tools for music can have a large value for applications in music archives. For instance, they can help to segment field recordings into music and non-music sections [11], to identify recordings of the same tune in large collections [10], or to detect duplicates of recordings [18]. Increasingly, projects aim at digitization and preservation of archives,[1] or at improving exchange and accessibility of music data resources.[2] Various recent projects, some described in more detail below, have shown the potential of a closer collaboration between music archives and MIR researchers. As illustrated in this paper, an important obstacle to a more extended application of MIR tools by archivists is the lack of such tools' availability in a useful form.

To increase the availability of MIR tools, we need an understanding of how cross-fertilization between music archives and MIR research can drive progress in both fields. To this end, dialogue and exchange must be stimulated, and guidelines and best practices must be determined. In this paper, we present the outcomes of discussions between archivists, musicologists, and MIR researchers held at the Lorentz workshop *Computational Ethnomusicology: Methodologies for a new field,*[3] and subsequently with archivists at DANS.[4] The discussions identified those MIR tasks with highest value for music archives. Starting with these tasks, collaborations between MIR research and music archives can improve availability and applicability of analysis tools, and will lead to a series of impacts for both fields:

- Opening of new roads for evaluating MIR tools. Data different than commonly used in MIR research will become available, including data encoding a wider variety of musical styles from various cultures, studio and field recordings, and mixtures of notation, audio, and video in various formats.
- The creation of a complete value chain, connecting research code to user-friendly software products. Within this value chain, demands for MIR algorithms can be more explicit, and development can become more user-oriented.
- Improved sustainability of tools due to a more user-oriented approach, improved accessibility of archival material, and improved functionality for archivists to prepare, process, and archive data.

---

[1]See, e.g., http://www.tape-online.net/ and http://www.bl.uk/manuscripts/
[2]See, e.g., http://www.europeanasounds.eu/ and http://www.music-encoding.org/
[3]http://www.lorentzcenter.nl/lc/web/2017/866/info.php3?wsid=866
[4]http://www.dans.knaw.nl/

2 RELATED WORK

MIR research aims at developing computational methods for extracting, processing and organizing musical information from data collections. Digital music archives can be partners in this endeavor, providing the data collections and research questions regarding their specific collections.

A collaboration between Utrecht University and the Meertens Institute in Amsterdam stimulated the development of computational methods for searching the Dutch Song Database based on musical content [23], as well as a wide range of research into oral transmission [24]. Starting from the musicological archivists' concept of *tune families* as groups of similar songs in the Database, the collaboration has fostered research on perceived melodic similarity [26], symbolic similarity measures [8, 25], and automatic pattern finding and compression [1]. The computational methods helped to provide the general public access to the collection, and to unlock the collection's potential for musicological research questions.

In a similar manner, the Digital Music Lab project [27] aimed to apply MIR tools to large audio and other music data collections held by the British Library, with the goal to provide research findings relevant for musicology. This required a close collaboration between all partners in order to ensure the complete research chain between audio feature extraction, a semantic web framework, and visualization interfaces.

In the DIADEMS project,[5] several ethnological and anthropological institutes collaborated in research into singer turns and instrumental timbre [5, 21], using the audio content of field recordings from the CNRS archive. Within this project, a collaborative multimedia asset management system, Telemeta, was integrated into the archive's website.[6] This system incorporates automatic analysis functionalities, such as speech detection and segmentation. The content of this database is now available for listening all over the world, enabling communities from which historical field recordings originate to interact with their cultural heritage.

In collaboration with the archive of the Royal Museum for Central Africa in Tervuren, Belgium, research on tempo estimation and pitch profiles in Central African musics [2, 17] resulted in tools for automated processing of music from a particular geographical region. The collaboration was initiated by the digitization project DEKKMMA,[7] whose goals were preservation, database organization, and content analysis tool development, and which made the content accessible to a wider audience. Furthermore, it posed novel challenges to the development of tempo estimation algorithms [2].

In the Ethnomuse project [12], a collaboration between the University of Ljubljana and the Slovenian Institute of Ethnomusicology, segmentation methods were developed to partition field recordings into music and non-music parts [11]. The prototypes were further developed into user-friendly software.[8] The devised tools are used by archivists and researchers to organize metadata and access the music collection, but public access so far is limited to melody and query-by-humming functionalities on a part of the collection.[9]

The research conducted within the Single Interface for Music Score Searching and Analysis project [6] at McGill University is arranged along two strands: a *content* strand, in which optical music recognition techniques to transform digital images of scores into machine-readable representations are improved, and an *analysis* strand, in which tools for large-scale analysis of symbolic music notation are developed. The project's goal is to increase the accessibility of digital collections from libraries and museums around the globe—among the collaborators are the Hathi Trust Research Center and the Bavarian State Library—, thus opening the way to large-scale data-driven analysis, study, and performance.[10]

While the described projects are promising examples of collaborations between MIR researchers and musical archives, the full potential of using MIR methods both for organizing collections and for investigating research questions that are relevant in the context of these collections has not yet been exploited.

3 LIMITATIONS AND PERSPECTIVES

Despite the progress in MIR tasks during the past decades, most developed technologies have so far not found application in digital music archives. In discussions with musicologists, archivists, and MIR researchers, we identified three main reasons for this:

- Archivists are not aware of the existence of common MIR tasks, or have a misleading understanding of the type of challenges MIR research addresses.
- Conversely, MIR research questions are not end user-driven but rather formulated as technological challenges.
- State-of-the-art MIR systems are often unavailable as user-friendly, well-documented implementations.

Typically, MIR research follows a common workflow. Researchers address a specific task, for example beat tracking or automatic transcription, and propose a method which improves the accuracy. This is usually measured by its performance on standardized test collections, or in evaluation initiatives such as the Music Information Retrieval Evaluation eXchange (MIREX).[11] Results are then published in peer-reviewed conference proceedings or journals.

In recent years, the issue of reproducibility of research has gained importance in the MIR community, and several conferences and journals require or encourage provision of source code and evaluation data. As a consequence, many MIR systems have become publicly available in the form of source code, distributed through open repositories. However, while the provided materials ensure reproducibility of the reported experimental results, they are neither designed to serve as ready-to-use tools for end users, nor as interoperable modules for software developers. To reach this level, further development with respect to scalability, documentation, maintenance and interoperability is necessary—*but this usually exceeds the interest area, skill set, and available funding of MIR researchers*. So far, only few efforts have attempted to define a unified algorithm format allowing a flexible incorporation into existing software frameworks—an example is the VAMP plugin interface.[12]

In addition, MIR algorithms are usually developed and tested under laboratory conditions, and the reported accuracies may only

---

[5]http://www.irit.fr/recherches/SAMOVA/DIADEMS/
[6]http://archives.crem-cnrs.fr/
[7]http://music.africamuseum.be/english/index.html
[8]http://lgm.fri.uni-lj.si/portfolio-view/sefire/
[9]http://www.etnofletno.si/

[10]See also the project website at http://www.simssa.ca/.
[11]http://www.music-ir.org/mirex/wiki/MIREX_HOME
[12]http://www.vamp-plugins.org/

generalize under certain assumptions. Furthermore, commonly used evaluation datasets contain mainly Western music, and MIR technology overall is too much focused on Western musical concepts [13]. Lastly, MIR algorithms often rely on latent assumptions (e.g., about tonality or rhythmic organization), which do not apply universally. Often such assumptions are not made explicit, and only become apparent when the internal mechanisms of the algorithm are analyzed. That being said, a recent trend in MIR focuses on genre-specific systems that explicitly target non-Western music.

Despite these limitations, MIR tools have great potential for archivists as well as archive users. In particular, they can help to describe, visualize and query high-level information encoded in music, and facilitate navigation through digital archives, potentially attracting a larger group of users. Furthermore, they can aid archivists in editing and maintaining content and metadata, in fully or partially automated procedures. Also, automatic content description technologies set the basis for big data analysis and large-scale musicological studies. Conversely, close collaboration with archivists can help MIR researchers to identify new challenges and evaluate their algorithms in real-world environments.

4 EXISTING MIR TOOLS: A CASE STUDY

In this section, two potential MIR applications for digital archives, as identified in the abovementioned discussions, are described.

The task of **segmenting audio recordings** refers to the automatic splitting of audio recordings into non-overlapping parts, such that each part contains only one *audio class* from a taxonomy of classes (e.g., speech, *a capella* singing voice, instrumental and background sounds). The size and organization of taxonomies can vary greatly. For example, Google's Audio Set [7] is a hierarchical ontology of 632 diverse audio classes, whereas the Urban Sounds Datasets [15] is a smaller collection of sounds from urban environments. A related subtask is the segmentation of long, uninterrupted field recordings, whose accompanying metadata cannot be directly linked to individual parts of the audio file. A related 2015 MIREX task used data from the British Library's music collections for a comparative study on the problem of speech/music segmentation.[13] This particular binary task has formed a research problem for over a decade [14]. Leaving the two-class scenario, SeFiRe, a downloadable tool for the segmentation and visualization of field recordings, segments and labels an audio file into five classes: speech, solo and choir singing, instrumental, and bell chiming.

Audio-to-score alignment refers to the task of synchronizing a score with an audio file during playback. From a technical perspective, two cases are distinguished: offline (computations are performed before playback) and online (computations are performed at runtime) alignment, or score following. Both are likely to be useful for digital archives. In recent years, many systems have been developed for both cases, and score following has been a MIREX task since 2008, where methods are evaluated based on their performance on Western classical scores. Consequently, it is not clear without in-depth study of the algorithms and their underlying assumptions how well the reported performance generalizes to different music traditions. In addition, most presented algorithms

are only documented in the form of a technical report, and do not provide source code or user-level tools. A recent genre-specific approach to audio-to-score alignment in Turkish makam music [16] has been made publicly available as a Python toolbox[14] and a component of the Dunya framework,[15] an online music visualization tool for non-Western musics.

Consequently, the suitability of audio-to-score alignment methods would have to be tested on the music genres contained in a particular archive, and further software development would be needed to implement and incorporate state-of-the-art technologies into the archive's existing infrastructure.

5 THE MIR VALUE CHAIN

As shown above, successful examples of MIR research exist that bridge algorithmic and interface development with the goal to access the contents of music archives. In order to arrive at a systematic framework for such processes, it is helpful to consider MIR research within the following five-layer *value chain* [9]:

(1) **MIR research and development layer** Research ideas drive the development of algorithms, which aim at solving specific tasks and are evaluated on a set of music corpora.

(2) **Software development layer** Promising research ideas and concepts are taken up and developed into systems that follow specifications regarding their interfaces, outputs, and efficiency. In collaborations with archives, this part of the value chain is accomplished by MIR researchers.

(3) **Product design layer** Interfaces are designed that are aimed to provide access to the functionalities of developed systems. In some of the described projects, these design aspects were handled by web technicians working at archives.

(4) **Publishing layer** The designed products are made accessible to users; in case of software products usually by means of online publishing. In our examples, the archives incorporate the designed access interfaces into their websites, and stream audio content or provide it for download.

(5) **End user layer** To the best of our knowledge, so far there is no quantitative data on archive access through the novel interfaces.

5.1 Suggested action points

The discussed archive projects connect the first four layers of the value chain. However, as described in Section 3, the lack of awareness among archivists about MIR, as well as the often weak connection between MIR research and user studies, are problems that indicate a lack of connectivity of the value chain. As a consequence, some MIR research results promising in the context of digital music archives do not proceed through the value chain.

An integrated development process following the described value chain will provide MIR research access to music data largely absent from evaluation corpora until now. Thus, the existing bias towards Western musics can be tempered, and the ability of algorithms to generalize can be examined on a wider basis. As a benefit for music listeners, digital music archives will increase the diversity of available streaming content by making accessible non-commercial

---

[13]http://www.music-ir.org/mirex/wiki/2015:Music/Speech_Classification_and_Detection

[14]http://www.github.com/sertansenturk/tomato/
[15]http://dunya.compmusic.upf.edu/makam/

archive recordings. In turn, the end users of the archives' content interfaces provide an immensely valuable extension to the mainly music corpus-based evaluations in MIR research.

Application-oriented MIR research by means of collaboration between archivists, interface designers, and user studies experts is a promising approach to improve the connectivity of the value chain. Software development stands separate from, but interacts with, such research. In the short-term perspective, the following action points for MIR reseachers and research groups are suggested:

- An active contribution to delivering reproducible results. Here, providing access to data used in research and well-documented prototypes are essential. The importance of making data available has been acknowledged, but software sustainability is a relatively new topic [3, 19, 20]. Furthermore, when developing prototypes, where possible standardized interfaces should be adhered to so that software developers can engage with the research product [22].
- The use of standardized, open data formats. This encourages reusability and sustainability of research data, and facilitates linking across seemingly unrelated datasets—enabling researchers to discover new relationships between them. A prime example of an archive suggesting so-called *preferred formats* is the Library of Congress; see [4] for a discussion.
- The organization of workshops for researchers and archivists, to which archivists can bring samples of their own data so that the functionality of newly developed tools can be demonstrated in mimicked real-world situations.
- Offering students following an MIR curriculum the possibility to take up an internship at a digital music archive.

6   CONCLUSIONS AND FUTURE WORK

The mutual benefits of collaboration between MIR researchers, who develop computational methods to extract information from digital music data, and digital music archives, which provide well-maintained collections of such data as well as relevant research questions, require little further explanation. However, despite several recent digitization and archiving projects, collaboration between the two parties is currently suboptimal. We identify three reasons for this: archivists do not always have a clear understanding of the common MIR tasks and problems, MIR researchers tend to formulate their questions as technological challenges rather than end user-driven solutions, and state-of-the-art MIR systems usually are prototypes instead of end-user friendly implementations.

However, as the case study described underlines, MIR tools can be of great value for the end user, in our case the archivists—provided their close involvement in the research from the earliest possible stage on. We therefore propose to incorporate the MIR research process into a five-layer value chain, connecting research, software development, product design, publishing, and the end user. To realize this, we draft a number of action points. These include an active contribution to delivering reproducible results by providing well-documented tool prototypes whose implementation comforms to standardized software interfaces; using standardized, open data formats; organizing hands-on workshops; and incorporating archive internships in academic curricula.

We foresee three main benefits of the suggested collaboration between MIR researchers and archivists: new perspectives for content access in archives, more diverse evaluation data and methods, and an MIR research workflow that is more application-oriented.

ACKNOWLEDGMENTS

REFERENCES

[1] P. Boot, A. Volk, and W. B. de Haas. 2016. Evaluating the role of repeated patterns in folk song classification and compression. *J. New Music Res.* 45, 3 (2016), 1–18.
[2] O. Cornelis et al. 2013. Evaluation and recommendation of pulse and tempo annotation in ethnic music. *J. New Music Res.* 42, 2 (2013), 131–149.
[3] S. Crouch et al. 2013. The Software Sustainability Institute: Changing research software attitudes and practices. *Comput. Sci. Eng.* 15, 6 (2013), 74–80.
[4] R. de Valk, M. van Berchum, and P. van Kranenburg. 2017. Music encoding, formats, and data sustainability. In *Proc. MEC*. Forthcoming.
[5] D. Fourer et al. 2014. Automatic timbre classification of ethnomusicological audio recordings. In *Proc. 15th ISMIR*. 295–300.
[6] I. Fujinaga, A. Hankinson, and J. E. Cumming. 2014. Introduction to SIMSSA (Single Interface for Music Score Searching and Analysis). In *Proc. 1st DLfM*. 1–3.
[7] J. F. Gemmeke et al. 2017. Audio Set: An ontology and human-labeled dataset for audio events. In *Proc. 42nd IEEE ICASSP*. 776–780.
[8] B. Janssen, P. van Kranenburg, and A. Volk. 2017. Finding occurrences of melodic segments in folk songs employing symbolic similarity measures. *J. New Music Res.* 46, 2 (2017), 118–134.
[9] R. Kaplinsky and M. Morris. 2001. *A handbook for value chain research.* International Development Research Centre, Ottawa.
[10] C. C. S. Liem and A. Hanjalic. 2009. Cover song retrieval: A comparative study of system component choices. In *Proc. 10th ISMIR*. 573–578.
[11] M. Marolt. 2009. Probabilistic segmentation and labeling of ethnomusicological field recordings. In *Proc. 10th ISMIR*. 75–80.
[12] M. Marolt, J. F. Vratanar, and G. Strle. 2009. Ethnomuse: Archiving folk music and dance culture. In *Proc. IEEE EUROCON*. 322–326.
[13] D. Moelants et al. 2007. The problems and opportunities of content-based analysis and description of ethnic music. *IJIH* 2 (2007), 57–68.
[14] A. Pikrakis, T. Giannakopoulos, and S. Theodoridis. 2008. A speech music discriminator of radio recordings based on dynamic programming and Bayesian networks. *IEEE Trans. Multimedia* 10, 5 (2008), 846–857.
[15] J. Salamon, C. Jacoby, and J. P. Bello. 2014. A Dataset and taxonomy for urban sound research. In *Proc. 22nd ACM ICM*. 1041–1044.
[16] S. Sentürk et al. 2015. A tool for the analysis and discovery of Ottoman-Turkish makam music. In *Proc. 16th ISMIR*.
[17] J. Six, O. Cornelis, and M. Leman. 2013. Tarsos, a modular platform for precise pitch analysis of Western and non-Western music. *J. New Music Res.* 42, 2 (2013), 113–129.
[18] J. Six and M. Leman. 2014. Panako - A scalable acoustic fingerprinting system handling time-scale and pitch modification. In *Proc. 15th ISMIR*. 259–264.
[19] A. M. Smith, D. S. Katz, and K. E. Niemeyer. 2016. Software citation principles. *PeerJ Comput. Sci.* 2, e86 (2016). https://doi.org/10.7717/peerj-cs.86
[20] S. Sufi et al. 2014. Software in reproducible research: Advice and best practice collected from experiences at the collaborations workshop. In *Proc. 1st ACM SIGPLAN TRUST*. 2.
[21] M. Thlithi et al. 2014. Segmentation in singer turns with the Bayesian Information Criterion. In *Proc. 15th INTERSPEECH*. 1988–1992.
[22] M. van Gompel et al. 2016. *Guidelines for software quality.* Technical Report. CLARIAH.
[23] P. van Kranenburg et al. 2009. Musical models for folk-song melody alignment. In *Proc. 10th ISMIR*. 507–512.
[24] P. van Kranenburg, M. de Bruin, and A. Volk. 2017. Documenting a song culture: The Dutch Song Database as a resource for musicological research. *Int. J. Digit. Libr.* (2017). https://doi.org/10.1007/s00799-017-0228-4
[25] P. van Kranenburg, A. Volk, and F. Wiering. 2013. A comparison between global and local features for computational classification of folk song melodies. *J. New Music Res.* 42, 1 (2013), 1–18.
[26] A. Volk and P. van Kranenburg. 2012. Melodic similarity among folk songs: An annotation study on similarity-based categorization in music. *Music. Sci.* 16, 3 (2012), 317–339.
[27] T. Weyde et al. 2014. Big data for musicology. In *Proc. 1st DLfM*. 1–3.