Multilayer Music Representation and Processing: Key Advances and Emerging Trends

Federico Avanzini, Luca A. Ludovico
LIM – Music Informatics Laboratory
Department of Computer Science
University of Milan
Via G. Celoria 18, 20133 Milano, Italy

{federico.avanzini,luca.ludovico}@unimi.it

***Abstract*—This work represents the introduction to the proceedings of the *1st International Workshop on Multilayer Music Representation and Processing (MMRP19)* authored by the Program Co-Chairs. The idea is to explain the rationale behind such a scientific initiative, describe the methodological approach used in paper selection, and provide a short overview of the workshop's accepted works, trying to highlight the thread that runs through different contributions and approaches.**

***Index Terms*—Multilayer Music Representations, Multilayer Music Applications, Multilayer Music Methods**

I. Introduction

An effective digital description of music information is a non-trivial task that involves a number of knowledge areas, technological approaches, and media. Catching and profitably handling such a richness are needs increasingly felt by industry and academia, thus crossing the narrow boundaries of sound and music computing community.

This idea is not a novel one: for instance, early formats for the multilayer representation of music dates back to decades ago. In this sense, a milestone initiative was the *1st IEEE International Conference on Musical Application using XML (MAX 2002)* held at the Department of Computer Science, University of Milan in 2002. In that occasion, world renowned experts gathered together to discuss the state of the art and the future of multilayer digital representations of music. In the meanwhile, digital technology, computer-based approaches, and formats have been evolving in directions that could hardly be predicted more than 15 years ago: the advent of new devices such as smartphones and wearable devices, the growing interest in augmented and virtual reality, the evolution of artificial intelligence towards deep learning, the drastic changes in user's entertainment and communication habits are only a few examples of what happened in the last years. The way music and multimedia information is preserved, manipulated and presented can not ignore the most recent technological advances and emerging trends.

This is the background that led to the organization of the *1st International Workshop on Multilayer Music Representation and Processing (MMRP19)*, a scientific initiative dealing with the already-explored subject of multilayer representation, but facing it under a new light. Following on in the tradition of MAX 2002, this event has been organized once again at the Department of Computer Science of the University of Milan.

From the point of view of MMRP19 organizers, a key result to achieve was inclusiveness. This aspect can be appreciated, for example, in the composition of the Scientific Committee, which gathered world-renowned experts from both academia and industry, covering different geographical areas (see Figure 1) and bringing their multi-faceted experience in sound and music computing. Inclusiveness was also highlighted in the call for papers, focusing on novel approaches to bridge the gap between different layers of music representation and generate multilayer music contents with no reference to specific formats or technologies. For instance, suggested application domains embraced heterogeneous fields: computational musicology, intangible cultural heritage, music education and training, libraries and archives, multimedia entertainment, etc. Similarly, solicited subjects included not only multilayer representation, but also signal processing, machine learning and understanding of music, optical music recognition, etc. In Section II we will explain the way accepted works have been clustered around big themes, in order to confer an explicit and easily understandable structure to the workshop.

We were glad to receive submissions from several countries, with authors of submitted papers spanned over three continents (see again Figure 1). We believe that this response from the scientific community confirmed that the workshop addressed a timely topic. Thanks to the hard work of the Scientific Committee, and of additional reviewers, each submitted paper received three independent reviews from experts in the field.

The workshop was held in conjunction with the kick-off event of the IEEE Working Group (WG) for XML Musical Application. The IEEE 1599-2008™ Standard for music representation was first released in 2008, with the aim of providing a comprehensive description of the multi-layered information related to a given music piece, in a single XML document. Ten years after the release of the standard, the new WG will work at updating and extending it, in the light of new techniques and new application domains emerged during the last decade. The contributions to the MMRP workshop and the fruitful discussion among participants has provided the WG with novel suggestions and insights.

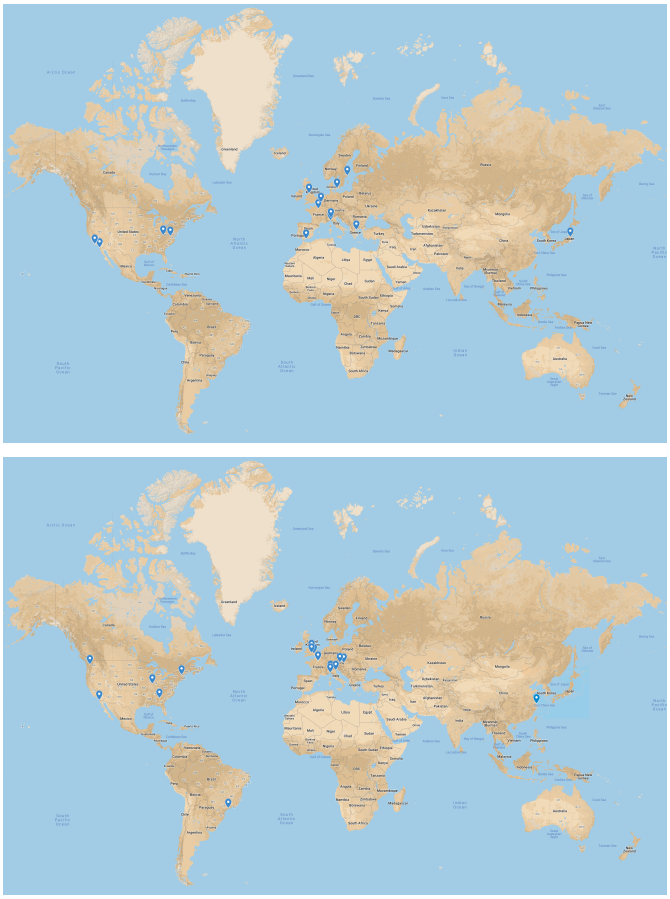

Fig. 1. Geographical distribution of the affiliations of MMRP19 Scientific Committee (top panel) and of the MMRP19 paper authors (bottom panel).

II. Structure of the Workshop

The MMRP19 workshop accepted 13 papers that have been presented during 3 sessions. The main themes identified in order to cluster contributions were multi-layer music representations, applications, and methods. The goal of next subsections is to provide an overview of each session subject, remarking heterogeneous approaches and finding common trends.

A. Representations

The first paper session, entitled *Representations*, dealt with complementary ways to face the problem of music description, taking into account recent technological advancements and new needs.

Reference [1] addressed smart musical instruments, an emerging category of musical instruments characterized by sensors, actuators, wireless connectivity, and embedded intelligence. This contribution aimed to provide the key requirements of an interoperable file format for the exchange of content produced by this new class of instruments in the context of Internet of (Musical) Things.

Reference [2] represented a critical survey on multimodal collaborative processing and retrieval of music information. The goal was to highlight how multimodal algorithms, working simultaneously on audio and video recordings, symbolic music scores, mid-level representations, motion and gestural data, etc., can help Music Computing applications. In next years, approaches based on information fusion will be a challenging subject in the fields of Music Information Retrieval and Sound and Music Computing.

Reference [3] addressed the problem of music browsing and music information retrieval on the base of semantic descriptions. Currently, the increased availability of musical content and the constitution of large music libraries requires new paradigms for music recommendation, browsing and retrieval able to overcome the traditional metadata-based search methods. In response to such a need, the authors proposed a browsing framework based on the navigation into a three-dimensional space where high-level semantic descriptors of musical items can be represented.

Finally, Reference [4] critically analyzed and compared the most promising extensible formats for multilayer music representation. Specifically, the contribution focused on IEEE 1599, MEI and MusicXML. Each format was described in its key features, strengths and weaknesses. Since the communities of each format are very active, this work aimed to shed a light on their future perspectives.

In conclusion, session *Representations* provided multi-faceted views on the complex issue of digital music representation, describing the features of current formats [4] and proposing new ones [1], highlighting how a suitable representation can turn into an effective way to improve music information computing and retrieval [2], and proposing new forms of semantic representation of music items [3].

B. Applications

The second MMRP19 session, entitled *Applications*, focused on some practical issues solved by multilayer approaches.

Reference [5] presented an innovative system for real-time machine listening in the context of human-machine free improvisation. Anthony Braxton's Language Music system was adopted as a grammatical model for contextualizing real-time audio feature data within free improvisation. In this way, musical material unknown to the system can be organized into a fluid, coherent, and expressive musical language, thus yielding idiosyncratic interactions, full machine autonomy, and novel musical output.

Also Reference [6] focused on automated music generation, but starting from the representation of hierarchical musical structures in terms of multilayered maps that refer to the composer's mental representation as well as the listeners' perception of the piece. The paper described a computational method to perform this kind of analysis and proposed an implementation to generate short musical phrases and perform resulting melodies.

Heading in the direction of a higher degree of abstraction, Reference [7] dealt with semantic Web technologies to build new musical experiences. In more detail, this contribution described the evolution and the outcomes of the *FAST project (Fusing Audio and Semantic Technology for Intelligent Music Production and Consumption)*, an initiative aiming to realize

a new multilayer musical ecosystem in response to current users' requirements: richness in linked contents, flexibility, interactivity, adaptivity, etc.

Finally, Reference [8] applied the concepts of multilayer description and multimodal interactive systems to the field of music training and education, specifically to the automated analysis of postural and movement qualities of violin players. This contribution demonstrated how collecting and analyzing multimodal performances in order to provide the learner with feedback can improve instrumental practice, maximize efficiency and minimize injuries.

In conclusion, session *Applications* introduced some novel approaches based on multilayer descriptions for music generation [5], [6], instrumental practice [8], and music experience [7].

C. Methods

The third and last MMRP19 session, entitled *Methods*, collected papers dealing with techniques for extracting information from music representations, and in some cases for generating musical information. Perhaps not surprisingly, almost all of the contributions in this session made use of neural networks and deep learning approaches.

Reference [9] applied Deep Neural Networks (DNNs) to the problem of sound source separation, with an emphasis on the musical problem of separating the leading singing voice from a polyphonic musical mixture. The separation network is informed with the frame-level vocal activity, thus learning to differentiate between vocal and non-vocal regions and ultimately reducing artifacts in the separation results. Outcomes on a known dataset show that the proposed approach provides better separation with respect to state-of-the-art methods.

Reference [10] addressed a different problem, namely the estimation of the dominant melody in polyphonic audio. Similarly to other recent works, the authors experiment the use of a neural network originally designed in the context of image processing. To this end, the U-net is here trained with a novel sequential method and with careful pre-conditioning of the training data, and is shown to outperform plain convolutional networks.

Unlike the previous contributions, Reference [11] focused on the development of meaningful audio features, that can be used for machine learning tasks. Specifically, the authors introduce a novel feature called Chroma Interval Content, which is well suited to describe key-independent harmonic progressions and can be computed efficiently. They subsequently explore its power in representing chord progressions and its use in specific MIR tasks that are related to harmony.

With Reference [12] we are back to neural networks, although in this case the context is computational creativity and generative models rather than analysis and retrieval. The authors explored a technique for symbolic melody generation, constrained by a given chord progression. In particular, they compare two generative models based on Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs). Results are assessed in terms of both subjective judgments and objective analysis through Variable Markov Oracle.

Finally, Reference [13] dealt with the problem of automatic transcription (specifically, polyphonic piano transcription) within the framework of multitask learning, with multiple targets, predictions, and loss functions. The performance of the proposed approach was investigated using various CNN architectures on a large dataset of piano performances accompanied by ground-truth information.

In summary, session *Methods* dealt with music audio features [11] and deep learning approaches [9], [10], [12], [13], in the context of relevant music information retrieval [9]–[11], [13] and automatic music generation [12] tasks.

III. Emerging trends

The diversity of the topics covered by the workshop sessions provide useful indications of relevant emerging research trends.

Session *Applications* showed that the potential applications of multi-layer music representation formats span several domains, including (but not limited to) creative uses (automatic accompaniment systems, automatic music generation), education (music instrument learning), and music production. With respect to ten years ago, the industry is now ready to exploit these application domains.

This richness of potential applications and industrial exploitation, however, calls for the development of music representation formats and standards that are able to represent different information layers in a single and coherent data structure. Papers in the session *Representations* have addressed these issues from different points of view.

Well structured formats, in turn, need efficient and accurate techniques that allow to automate content production for such formats, by analyzing and synchronizing information across various representation layers. In a complementary fashion, the availability of music information structured in this way may allow to extract higher-level meaning using appropriate features and machine learning approaches.

References

[1] L. Turchet and P. Kudumakis, "Requirements for a file format for smart musical instruments," in *Proc. of the 1st Int. Workshop on Multilayer Music Representation and Processing*. IEEE CPS, 2019.

[2] F. Simonetta, S. Ntalampiras, and F. Avanzini, "Multimodal collaborative music information processing and retrieval: Survey and future challenges," in *Proc. of the 1st Int. Workshop on Multilayer Music Representation and Processing*. IEEE CPS, 2019.

[3] S. Cherubin, C. Borrelli, M. Buccoli, M. Zanoni, A. Sarti, and S. Tubaro, "Three-dimensional mapping of high-level music features for music browsing," in *Proc. of the 1st Int. Workshop on Multilayer Music Representation and Processing*. IEEE CPS, 2019.

[4] A. Baratè, G. Haus, and L. A. Ludovico, "State of the art and perspectives in multi-layer formats for music representation," in *Proc. of the 1st Int. Workshop on Multilayer Music Representation and Processing*. IEEE CPS, 2019.

[5] H. Brown and M. Casey, "Heretic: Modeling anthony braxton's language music," in *Proc. of the 1st Int. Workshop on Multilayer Music Representation and Processing*. IEEE CPS, 2019.

[6] F. Carnovalini and A. Rodà, "A multilayered approach to automatic music generation and expressive performance," in *Proc. of the 1st Int. Workshop on Multilayer Music Representation and Processing*. IEEE CPS, 2019.

[7] M. Sandler, S. Benford, and D. De Roure, "Semantic web technology for new experiences throughout the music production-consumption chain," in *Proc. of the 1st Int. Workshop on Multilayer Music Representation and Processing*. IEEE CPS, 2019.

[8] E. Volta and G. Volpe, "Automated analysis of postural and movement qualities of violin players," in *Proc. of the 1st Int. Workshop on Multilayer Music Representation and Processing*. IEEE CPS, 2019.

[9] R. V. Swaminathan and A. Lerch, "Improving singing voice separation using attribute-aware deep network," in *Proc. of the 1st Int. Workshop on Multilayer Music Representation and Processing*. IEEE CPS, 2019.

[10] G. Doras, P. Esling, and G. Peeters, "On the use of u-net for dominant melody estimation in polyphonic music," in *Proc. of the 1st Int. Workshop on Multilayer Music Representation and Processing*. IEEE CPS, 2019.

[11] M. Queiroz and R. Borges, "Chroma interval content as a key-independent harmonic progression feature," in *Proc. of the 1st Int. Workshop on Multilayer Music Representation and Processing*. IEEE CPS, 2019.

[12] K. Chen, W. Zhang, S. Dubnov, G. Xia, and W. Li, "The effect of explicit structure encoding of deep neural networks for symbolic music generation," in *Proc. of the 1st Int. Workshop on Multilayer Music Representation and Processing*. IEEE CPS, 2019.

[13] R. Kelz, S. Böck, and G. Widmer, "Multitask learning for polyphonic piano transcription, a case study," in *Proc. of the 1st Int. Workshop on Multilayer Music Representation and Processing*. IEEE CPS, 2019.