AI4media

ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

ROADMAP ON AI TECHNOLOGIES & APPLICATIONS FOR THE MEDIA INDUSTRY

SECTION: "AI FOR MUSIC"

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911

info@ai4media.eu www.ai4media.eu

Authors	**Rémi Mignot** (Institut de Recherche et de Coordination Acoustique/Musique - IRCAM) **Axel Roebel** (Institut de Recherche et de Coordination Acoustique/Musique - IRCAM)

This report is part of the deliverable D2.3 - "*AI technologies and applications in media: State of Play, Foresight, and Research Directions*" of the AI4Media project.

You can site this report as follows:

F. Tsalakanidou et al., Deliverable 2.3 - AI technologies and applications in media: State of play, foresight, and research directions, AI4Media Project (Grant Agreement No 951911), 4 March 2022

This report was supported by European Union´s Horizon 2020 research and innovation programme under grant number 951911 - AI4Media (A European Excellence Centre for Media, Society and Democracy).

Copyright

AI for Music

Current status

In the early stages of ***automatic music processing***, such as in musical representation or sound signal modelling, ad hoc methods and expert algorithms were explicitly designed by the researchers or engineers to achieve specific tasks. The targeted applications were for example: tempo estimation, harmonic key recognition, automatic transcription, audio indexing, similarity estimation, automatic mixing, beat and downbeat detection; for which the experts of the field are able to formalise explicitly the models for the properties to analyze. Nevertheless, these approaches were not sufficient to explain more complex concepts of the music, such as the ***musical genre or the emotion*** provided by a song. Therefore, since almost two decades, machine learning has been used to build models able to automatically learn a musical concept, without having to formalise it mathematically. Frequently inspired by automatic speech recognition, these methods have been able to solve more complex tasks based on a dataset of examples, such as the musical genre recognition[1].

For some years now, deep learning has made possible to go even further. Many ad hoc methods, which worked fine, have been overtaken by deep neural network models. For example, whereas previous tempo estimation models were based on a dedicated signal transformation, pre-established rhythmic patterns and pattern recognition methods[2], more recent approaches are based on deep learning. These models are automatically learned using a dataset of annotated song recordings, for which the tempo is known[3].

These technology advances have a positive impact on the analysis and the processing of the music at the ***signal level***, but also at the ***symbolic level***, that is the score of a piece of music. In the context of the creation of contemporary and experimental music, some computer tools have been developed to help musicians when composing and orchestrating innovative music pieces[4]. Inspired by serialism[5] and spectralism music[6], some approaches are able to generate musical excerpts based on chosen constraints or a given musical sample[7] (Figure 1).

[1] G. Tzanetakis, and P. Cook. “Musical genre classification of audio signals.” *IEEE Transactions on speech and audio processing,* 2002.
[2] G. Peeters “Template-based estimation of time-varying tempo”. EURASIP JASP, 2007.
[3] H. Foroughmand, and G. Peeters, “Deep-rhythm for tempo estimation and rhythm pattern recognition”, ISMIR, 2019.
[4] J. Bresson, C. Agon, and G. Assayag. "OpenMusic: visual programming environment for music composition, analysis and research." *19th ACM International Conference on Multimedia*. 2011.
[5] Wikipedia article on Serialism: https://en.wikipedia.org/wiki/Serialism
[6] Wikipedia article on Spectral music: https://en.wikipedia.org/wiki/Spectral_music
[7] C. Cella, et al. “OrchideaSOL: a dataset of extended instrumental techniques for computer-aided orchestration.” 2020, arXiv preprint arXiv:2007.00763.

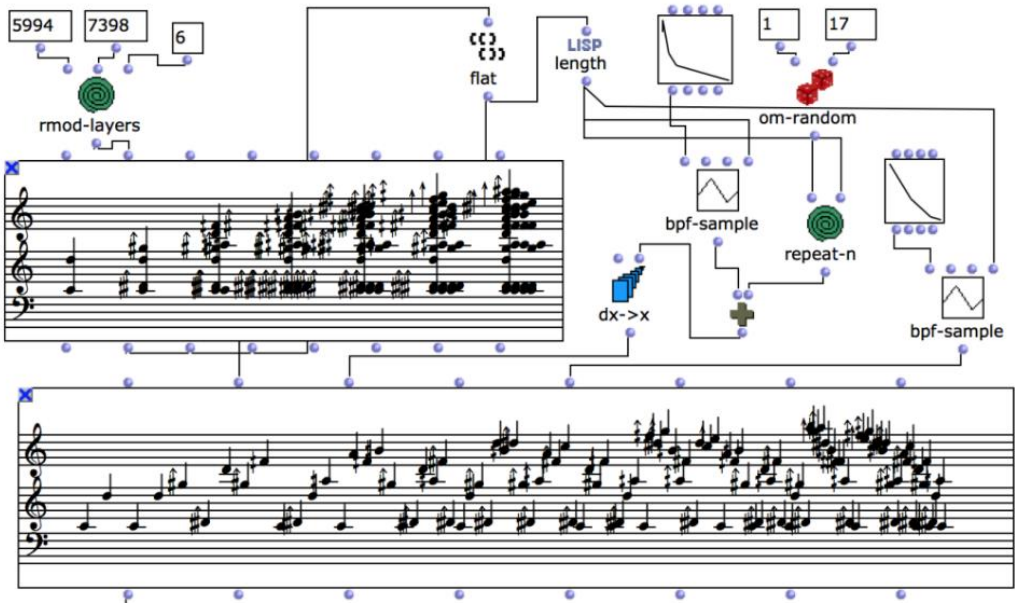

Figure 1: Automatic music generation tool based on rules and constraints[8].

Previously limited to experimental music, for which it is often possible to formalise the musical concept by mathematical constraints, nowadays this research field tends more towards ***popular music***, thanks to deep learning, among other reasons. Indeed, compared to classical and experimental music for example, popular music is more related to the musician's feeling, quite difficult to describe, rather than intellectual concepts. So, based on relevant and numerous examples, the flexibility of the deep neural network models makes possible to infer more sensitive concepts in music.

Research challenges

One standard problem in machine learning, and especially in deep learning, is the ***lack of data***. For example, traditional tempo estimation models are learned using a supervised training which needs some songs labelled to their tempo. First, the size of the training dataset must be as large as possible; second, the used examples must be of a good quality with a number of wrong annotations as low as possible; and third, the used dataset must be representative of the world of interest. For example, a tempo model trained with classical music may fail with electronic songs.

Unfortunately, for most of the applications in music processing, the creation of such datasets is difficult and can be expensive. To deal with this problem, many research results have been proposed for machine learning, and more specifically deep learning: data augmentation, non-supervised generative models (e.g. Variational Auto-Encoder), student teacher learning, data synthesis, domain adaptation, generative adversarial networks, differentiable processing, and transfer learning. This data issue is not restricted to music and sound processing, it is also a strong limitation for most of the other applications of deep learning. Even if it is possible to mix

[8] Open Music software, image source: http://christophertrapani.com/wordpresssite/computer-assisted-composition/

these mentioned methods, a challenge to researchers is to continue to develop ***new learning methods able to be trained with scarce data***.

A second challenge for AI research in music and sound processing is the "***artistic innovation***". In music history, the evolution has been usually preceded by technological advances. For example the romanticism of the 19th century is partly linked to the creation of the piano, which replaced the harpsichord of the 18th century. Nowadays, the use of AI can provide new ways to create music, and one could think that it makes possible the emergence of new styles. Nevertheless, there is the risk of the opposite effect. Indeed, because deep learning models are trained on existing music examples, it seems impossible for a machine to properly imagine the music of tomorrow; ***humans must stay in the loop***.

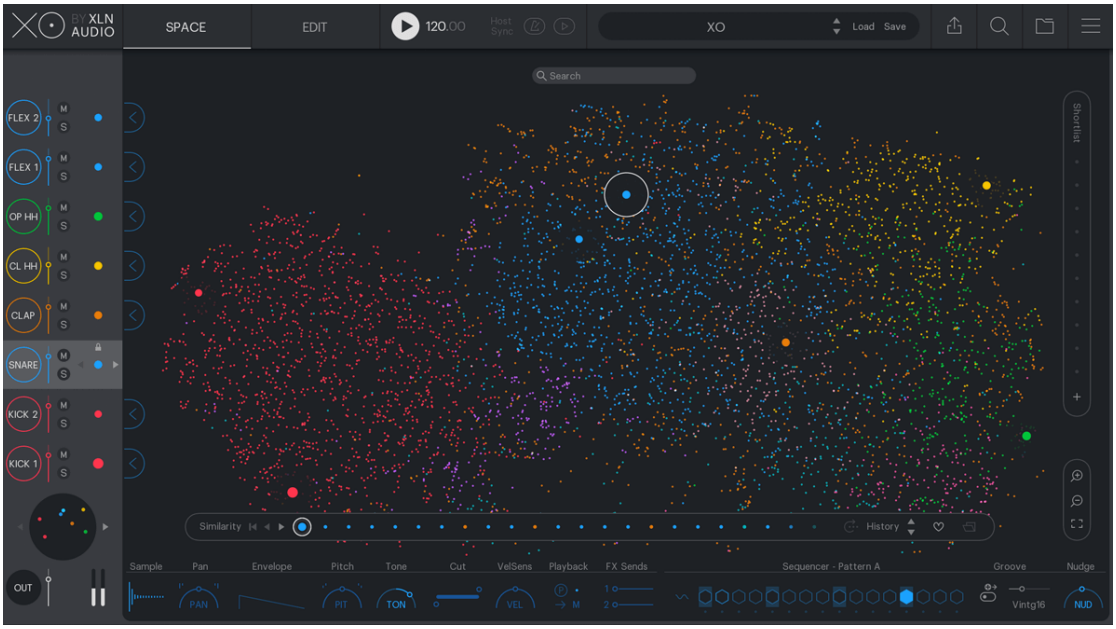

Figure 2: Graphical Interface for a smart exploration of a catalogue of drum sounds[9].

Consequently, most of the adopted strategies consist in providing an ***artificial assistant*** to musicians, but the question of creativity remains. For example, among the current AI assistants, the most advanced tools propose some generated musical samples according to the inputs given by the musician (Figure 2). Sometimes he/she has the possibility to transform the solution by moving in a low-dimensional latent space. For now, these algorithms provide limited degrees of liberty to the musicians, and a challenge for next research in music and sound processing is to ***give more control to the musicians*** to enhance the musical innovation. In other words, the musicians should be able to explore sounds and music out of the current distributions, but in an artistic and explainable way, and not in a random way.

[9] Software Xo of xln-Audio, image source: https://www.xlnaudio.com/products/xo

Societal and media industry drivers

Vignette 1: Composing music for a movie soundtrack with the help of an AI assistant

Joannah is a composer for a movie soundtrack. Her friend Stevia is a movie director, and she commissioned the composition and the orchestration of her next movie. During the shoot, Joannah composed the musical themes. Now the video editing is finished and she needs to align the orchestration on the video before the rehearsal and the recording with the musicians. Unfortunately the time is short, so she uses her AI assistant for the orchestration and the synchronisation of the music and the video. This consists of modifying the prepared music score to align it in time with the fixed video according to the ambiances, e.g. romantic, suspense, action, and also to punctual events. Joannah does not want the AI assistant to compose for her, she wants an intelligent assistant able to help her during the orchestration but without Joannah losing the control on the generated score. Thanks to it, Joannah prepared the music on time, and using the integrated sound synthesiser, she could produce a first preview, which Stevia validated before the recording with the real musicians.

Vignette 2: AI assistant for automatic music arrangement and sound mixing in live performances

Tom and Guillem form an electronic music band which is popular in the nightclubs. They like to improvise in live and to experiment with new things every night. Rather than to mix different sound recordings or drum machine samples, they use an AI assistant, which is able to generate different music tracks (e.g. bass lines, drums, synth drones) automatically adapted to the sound features of the played music, e.g. rhythm, tempo, harmony, ambiance. Contrary to standard drum machines, this assistant knows their particular musical style that differentiates them from other DJs. It is a great help to manage several instruments during the improvisations; for example, Tom and Guillem can decide to suddenly change the ambiance, and the assistant automatically changes the music tracks accordingly to the parameters controlled by the duo. They like it because, they keep full control on the assistant, it frees them from having to do boring tasks, it fits their musical style, and it improves their creativity during live performances.

Future trends for the media sector

The analysis and the generation of music already has an important role in the media sector. Nevertheless, there is a room for AI to improve its impact on media. Some future research trends are summarised below:

- Improve ***musical recommendation systems***, e.g. for video games, radios, documentaries, a) to better fit the user taste and the targeted features, and b) to take into account the context of the listening.

- Allow ***training on small datasets*** created by the end users for a specific task: a) improve the learning ability with scarce data, and b) design generalised models able to learn different tasks without the need to be used by expert users.
- Improve the ***learning of emotions*** provided by the music and felt by humans: a) for the analysis of music recordings, and b) for the generation or treatment of musical sounds and scores.
- Improve the ***learning of the musical language***. This could also improve the analysis and the synthesis of musical audio signal.
- Give ***more controls to the musicians*** during the process. For example, propose some understandable and controllable constraints which makes sense from an artistic point of view.

Goals for next 10 or 20 years

Musical AI applications will provide an improved understanding about what makes sense in the music: style, emotion, etc. But the machine should not be a competitor for musicians; it should stay an assistant which will be able to help them when doing tedious, repetitive, or long tasks. It is illusory to think that an AI model can be innovative in music and can propose new styles in a proper and artistic way, without a random process. For music creation, humans must stay in the loop, in consequence the research will be active to find some ways to easily control the AI process.

Nowadays, most AI applications for automatic composition and arrangement are based on digital scores in the MIDI format. First, we can have access to thousands of MIDI files but with a poor quality, without realistic nuances and variations; secondly, quality digital scores are few; and third, pairs of sound mixes with quality scores are even rarer. Nevertheless, we can access millions of recorded songs with high audio quality, many of them having partial annotations, but without the associated scores.

The main goal for the next 5-10 years will be to make the link between scores and sound mixes. To achieve this goal, some advances in deep learning are particularly interesting: for example because we do not have synchronised audio-score data, unsupervised methods and generative adversarial networks are very promising approaches, and some interesting research projects have already started. For a longer-term perspective, in the next 10-20 next years, using this future knowledge on the relation between symbolic music and sound recording, we will be able to learn efficient embedded musical representations of compositions and performances. These models will implicitly learn all the standard musical rules but also how to break them in an appropriate way like many talented musicians do.

Finally, a strong limitation still remains in capturing the emotions and moods provided by a song. A future solution can be given by neuroscientists. Indeed, the study of emotions is an active topic in neuroscience and we can expect that future results can be reused in musical AI applications. For example, we can imagine that in two decades, the emotion of a human subject will be efficiently detected by lightweight and portable encephalographic devices, and that volunteers will help in making emotion datasets just by listening to the music when travelling by train or by plane.

AI4media

ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911

info@ai4media.eu www.ai4media.eu