Neural Computing and Applications
https://doi.org/10.1007/s00521-019-04166-0

EDITORIAL

The emergence of deep learning: new opportunities for music and audio technologies

Dorien Herremans[1] · Ching-Hua Chuan[2]

Received: 18 March 2019 / Accepted: 20 March 2019

There has been tremendous interest in deep learning across many fields of study. Recently, these techniques have gained popularity in the field of music. Projects such as Magenta (Google's Brain Team's music generation project), Jukedeck, and IBM Watson Beat testify to their potential. Due to this rising interest in using deep neural networks to tackle tasks in the domain of audio and music, the guest editors organized the first International Workshop on Music and Audio as part of the International Joint Conference on Neural Networks (IJCNN) in Anchorage, Alaska in 2017. The current NCAA issue on "Deep learning for music and audio" was born out of the workshop.

While humans can rely on their intuitive understanding of musical patterns and the relationships between them, it remains a challenging task for computers to capture and quantify musical structures. Recently, researchers have attempted to use deep learning models to learn features and relationships that allow us to accomplish tasks such as music transcription, audio feature extraction, emotion recognition, music recommendation, and automated music generation. With this special issue, we aim to present a collection of research that advances the state of the art in machine intelligence for music and audio. This enables us to critically review and discuss cutting-edge research so as to identify grand challenges, effective methodologies, and potential new applications. The current issue therefore contains a wide variety of manuscripts that touch upon an important number of topics which remain of particular interest to the field of music and audio technology, including:

- deep learning for computational music research;
- modeling hierarchical and long-term music structures using deep learning;
- modeling ambiguity and preference in music;
- applications of deep networks for music and audio such as audio transcription, voice separation, music generation, music recommendation, etc.;
- novel architectures designed to represent music and audio.

We present a selection of papers on state-of-the-art approaches, current challenges, and future directions in deep learning for music and audio. Novel approaches are explored in various applications, including chord labeling, voice separation, and music generation. For instance, *Koops et al.* discuss how to model ambiguity and individual preferences when performing automatic chord labeling from audio, by using a merged representation of a dense deep neural network. Singing voice separation in audio recordings was tackled by *Lin et al*., who use an ideal binary mask to train a deep convolutional neural network. With regards to music generation, *Hadjeres and Nielsen* propose a new network architecture for generating (harmonized) soprano parts of Chorales that incorporates user-constraints in a recurrent neural network. In addition, *Dean and Forth* examine the use of neural networks to generate music in a rather unexplored style (post-tonal improvisation) and manage to obtain promising initial results. *Oore et al.* show that recurrent neural networks are able to generate expressive music. Their system received positive feedback from musicians. For readers who are new to music generation and deep learning, *Briot and Pachet*'s paper provides an introductory overview of the problem, approaches, and remaining challenges. Finally, the question of using CNNs for audio style transfer is examined by *Shahrin and Wyse*. While this problem remains hard, the authors showed that the network learns meaningful features, as audio texture is revealed in the gram matrices.

In addition to applications, a number of papers in this special issue also examine meaningful concepts that deep networks can learn from music and audio, as well as

✉ Dorien Herremans
dorien.herremans@gmail.com

[1] Singapore University of Technology and Design, Singapore, Singapore

[2] University of Miami, Coral Gables, USA

Published online: 01 April 2019

compare the performance of different architectures on feature learning, and investigate the impact of challenging scenarios in acoustic signals. *Chuan et al.* show that musical concepts such as key and chords can be captured by statistical learning methods such as word2vec, a commonly used technique in the field of natural language processing. Convolutional neural networks for audio emotion recognition are explored by *Wieser et al.* who found that these networks can learn meaningful features related to certain emotions. *Deng et al.* propose a novel deep time–frequency LSTM for audio restoration, whereby temporal and spectral dynamics are explicitly captured, thus allowing for more effective low bitrate audio restoration. *Dörfler et al.* show that the design of the audio filter and the time–frequency resolution can affect the accuracy of convolutional neural networks when used as a classifier. *Kiskin et al.* focus on the detection of low signal-to-noise ratio acoustic events (e.g., detecting the presence of mosquitoes in audio recordings) through convolutional neural networks and other machine learning techniques, using acoustic features extracted by different transforms. Finally, the effect of different deep architectures and multiple learning sources on a model's ability to learn efficient musical representations is examined by *Kim et al.*

We hope the readers will enjoy the manuscripts in this special issue. Our thanks goes out to all of the authors, reviewers, editor-in-chief, and the editorial office of NCAA for their support. Exciting times are ahead for the field of audio and music technologies.