Tutorial 1

Fundamentals of Music Processing: An Introduction using Python and Jupyter Notebooks

Meinard Müller and Frank Zalkow

Abstract

This tutorial will give an easy-to-understand introduction to music processing with a particular focus on audio-related analysis and retrieval tasks. In particular, the tutorial is aimed at non-experts and researchers who are new to the field. Based on well-established topics in Music Information Retrieval (MIR) as motivating application scenarios, we present fundamental techniques and algorithms that apply to a wide range of analysis and retrieval problems. We intend to explain the main ideas and techniques in an intuitive fashion using various figures and sound examples. Besides the theory, we also show how these techniques can be implemented going through specific Python code examples. All material, including the introduction of MIR scenarios, illustrations, sound examples, technical concepts, mathematical details, and code examples, are integrated into a comprehensive framework based on Jupyter notebooks. The notebooks are organized along with the eight chapters of the textbook on Fundamentals of Music Processing (FMP) (Springer 2015, www.music-processing.de). Another important goal of this tutorial is to show how the notebooks can be used to generate educational material for lectures and presentations. The notebooks (as well as HTML exports and multimedia examples) can be accessed via https://www.audiolabs-erlangen.de/FMP.

Reference

Meinard Müller
Fundamentals of Music Processing — Audio, Analysis, Algorithms, Applications
Springer Verlag, ISBN: 978-3-319-21944-8, 2015.

Meinard Müller studied mathematics (Diplom) and computer science (Ph.D.) at the University of Bonn, Germany. In 2002/2003, he conducted postdoctoral research in combinatorics at the Mathematical Department of Keio University, Japan. In 2007, he finished his Habilitation at Bonn University in the field of multimedia retrieval. From 2007 to 2012, he was a member of the Saarland University and the Max-Planck Institut für Informatik. Since September 2012, Meinard Müller holds a professorship for Semantic Audio Processing at the International Audio Laboratories Erlangen, which is a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and the Fraunhofer-Institut für Integrierte Schaltungen IIS. His recent research interests include music processing, music information retrieval, audio signal processing, and motion processing. Meinard Müller has been a member of the IEEE Audio and Acoustic Signal Processing Technical Committee from 2010 to 2015 and is a member of the Board of Directors of the International Society for Music Information Retrieval (ISMIR) since 2009. He wrote a monograph titled "Information Retrieval for Music and Motion" (Springer, 2007) as well as a textbook titled "Fundamentals of Music Processing" (Springer, 2015, www.music-processing.de).

Frank Zalkow studied Music Informatics and Musicology (Bachelor) and Music Informatics (Master) at the University of Music Karlsruhe, Germany. Since 2016, he has been working towards his Ph.D. degree in the Semantic Audio Processing Group headed by Meinard Müller at the International Audio Laboratories Erlangen. Previously, he worked for the Max-Reger-Institute Karlsruhe (2008–15) as well as Institute for Musicology at Saarland University (2015–16). His research interests include music retrieval, machine learning, as well as cross-connections between musicology and music information retrieval.

Tutorial 2

Generating Music with GANs: An Overview and Case Studies

Hao-Wen Dong and Yi-Hsuan Yang

Abstract

This tutorial aims to provide an overview of generative adversarial networks (GANs) and their use in generating music. The format of the tutorial will include lectures, demonstration of sample systems and technical results with illustrative musical examples.

- We will start by discussing the scope of music generation and introduce various tasks that can broadly be regarded as music generation. For each task, we will then discuss its challenges, commonly used approaches and some notable systems proposed in the literature.
- In the second part, we will explain the machine learning fundamentals for GANs. We will also present some interesting applications of GANs in other fields to showcase their potentials.
- The following section will contain the case studies of four different tasks—symbolic melody generation, symbolic arrangement generation, symbolic musical style transfer and musical audio generation. In each part, we will first provide an overview of the task and then introduce several models proposed in the literature as examples.
- We will conclude the tutorial by discussing the current limitations of GAN-based models and suggesting some possible future research directions.

The tutorial is targeted to students and newcomers who are interested in or working on music generation research, and also machine learning specialists who want to see how GANs can be applied to music generation.

Hao-Wen Dong is currently a research internship in the Research and Development Division at Yamaha Corporation. He will be starting a Ph.D. this fall in Electrical and Computer Engineering at University of California, San Diego. Previously, he was a research assistant under the supervision of Dr. Yi-Hsuan Yang in the Music and AI Lab at Academia Sinica. He received his bachelor's degree in Electrical Engineering at National Taiwan University. His research interests lie at the intersection of machine learning and music.

Yi-Hsuan Yang is an Associate Research Fellow with Academia Sinica, where he leads a research lab called the Music and AI Lab. He received his Ph.D. degree in communication engineering from National Taiwan University in 2010. He is also a Joint-Appointment Associate Professor with the National Cheng Kung University. His research interests include music information retrieval, affective computing, and machine learning. Dr. Yang was a recipient of the 2011 IEEE Signal Processing Society Young Author Best Paper Award, the 2012 ACM Multimedia Grand Challenge First Prize, and the 2015 Best Conference Paper Award of the IEEE Multimedia Communications Technical Committee. In 2014, he served as a Technical Program Chair of the International Society for Music Information Retrieval Conference (ISMIR). He gave a tutorial on "Music Affect Recognition: The State-of-the-art and Lessons Learned" in ISMIR 2012. He was an Associate Editor for the IEEE Transactions on Affective Computing and the IEEE Transactions on Multimedia in 2016-2019. He is currently on a sabbatical leave to work with a privately funded research organization in Taipei called the Taiwan AI Labs.

Tutorial 3

Audiovisual Music Processing

Zhiyao Duan, Slim Essid, Bochen Li, and Sanjeel Parekh

Abstract
Music is a multimodal art form. While sound plays a key role, other modalities, especially visual, are also critical to enhancing the musical experience. Recently, the MIR field has witnessed a rapid growth of interest in audiovisual processing of music.

This tutorial is intended to introduce this emerging research direction to the broader MIR community. It extends a recently published overview article on audiovisual analysis of music performances [1] into general audiovisual music processing. Specifically, it provides a comprehensive overview of state-of-the-art research in different aspects of audiovisual music processing, including music performance analysis, content-based retrieval, and music creation. It summarizes datasets, tools and other resources in this field, and articulates challenges and opportunities for future research. An interesting aspect of this tutorial is that it contains two hands-on case studies (30 min each) for the audience to personally experience audiovisual research. Instructions of software environments and starter code will be provided prior to the tutorial for preparation.

To our best knowledge, this is the very first tutorial on audiovisual processing at ISMIR. This tutorial is designed for students and researchers who have general knowledge of music information retrieval and who are interested in learning the state of the art and gaining hands-on experience of audiovisual music processing research. The comprehensive overview and categorization of different aspects of this field will help the audience gain a global view of the research problems, methods, tools, challenges, and opportunities. The hands-on case studies will provide the audience a first-hand experience of the research, helping them quickly arrive at the research frontier. We especially look forward to ideas and inspirations that the MIR community has to offer through this interactive and hands-on tutorial.

[1] Zhiyao Duan*, Slim Essid*, Cynthia C. S. Liem*, Gaël Richard*, and Gaurav Sharma*, "Audiovisual analysis of music performances: overview of an emerging field," IEEE Signal Processing Magazine, vol. 36, no. 1, pp. 63-73, 2019. (* authors in alphabetical order)

Zhiyao Duan is an assistant professor in Electrical and Computer Engineering, Computer Science and Data Science at the University of Rochester. He received his B.S. in Automation and M.S. in Control Science and Engineering from Tsinghua University, China, in 2004 and 2008, respectively, and received his Ph.D. in Computer Science from Northwestern University in 2013. His research interest is in the broad area of computer audition, i.e., designing computational systems that are capable of understanding sounds, including music, speech, and environmental sounds. He is also interested in the connections between computer audition and computer vision, natural language processing, and augmented and virtual reality. He co-presented a tutorial on Automatic Music Transcription at ISMIR 2015. He received a best paper award at the 2017 Sound and Music Computing (SMC) conference, a best paper nomination at the 2017 International Society for Music Information Retrieval (ISMIR) conference, and a CAREER award from the National Science Foundation.

Slim Essid received his state engineering degree from the École Nationale d'Ingénieurs de Tunis, Tunisia, in 2001, his M.Sc. (D.E.A.) degree in digital communication systems from the École Nationale Supérieure des Télécommunications, Paris, France, in 2002, his Ph.D. degree from the Université Pierre et Marie Curie (UPMC), Paris, France, in 2005, and his Habilitation à Diriger des Recherches degree from UPMC in 2015. He is a professor in Telecom ParisTech's Department of Images, Data, and Signals and the head of the Audio Data Analysis and Signal Processing team. His research interests are machine learning for audio and multimodal data analysis. He has been involved in various collaborative French and European research projects, among them Quaero, Networks of Excellence FP6-Kspace, FP7-3DLife, FP7-REVERIE, and FP-7 LASIE. He has published over 100 peer-reviewed conference and journal papers, with more than 100 distinct coauthors. On a regular basis, he serves as a reviewer for various machine-learning, signal processing, audio, and multimedia conferences and journals, e.g., a number of IEEE transactions, and as an expert for research funding agencies.

Bochen Li received his B.S. from University of Science and Technology of China in 2014. He is currently pursuing a Ph.D. degree in the Department of Electrical and Computer Engineering at the University of Rochester in the USA, under the supervision of Professor Zhiyao Duan. His research interests lie primarily in the inter-disciplinary area of audio signal processing, machine learning, and computer vision towards multimodal analysis of music performances, such as video-informed multipitch estimation and streaming, source separation and association, and expressive performance modeling and generation.

Sanjeel Parekh received B. Tech (hons.) degree in Electronics and Communication engineering from LNM Institute of Information Technology, India in 2014 and M.S. in Sound and Music Computing from Universitat Pompeu Fabra (UPF), Spain in 2015. His Ph.D. thesis titled 'Learning representations for robust audio-visual scene analysis' was completed in collaboration with Technicolor R&D and Telecom ParisTech, France between 2016-19. His research focusses on developing and applying machine learning techniques to problems in audio and visual domains. Currently, he is with LTCI lab at Telecom ParisTech, France.

Tutorial 4

Computational Modeling of Musical Expression: Perspectives, Datasets, Analysis and Generation

Carlos Cancino-Chacón, Katerina Kosta, and Maarten Grachten

Abstract
The aim of this tutorial is to introduce the theory and practice of music performance research to a broad MIR audience. A music performance—an acoustic or audio-visual rendering of it—provides a much richer musical experience than the symbolic or notated representation of the performed music. This richness is arguably an important aspect of our engagement with music and is shaped by the musician's interpretation of the intentions of the music, as conveyed through their performance. The means of expressing these intentions vary from one instrument to another, and can include tempo, timing, dynamics, articulation, timbre, and intonation, among others.

In this tutorial we will give a brief overview of the music performance literature and highlight how expressive dimensions affect the perception and the creation of music. Furthermore we will showcase some state-of-the-art computational methods for both analysis and synthesis of expressive piano performances. We include a hands-on part in which we share easily operated and adaptable code written in Python using Jupyter iPython notebook for demonstrating how to get started with computational analysis and synthesis of musical expression.

This work is partially supported by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 670035 (project "Con Espressione").

Carlos Cancino-Chacón is a postdoctoral researcher at the Austrian Research Institute for Artificial Intelligence (OFAI), Vienna, Austria. His research focuses on studying expressive music performance, music cognition and music theory with machine learning methods. He pursued a doctoral degree on computational models of expressive performance at the Institute of Computational Perception of the Johannes Kepler University Linz, Austria. He received an M.Sc. degree in Electrical Engineering and Audio Engineering from the Graz University of Technology, a degree in Physics from the National Autonomous University of Mexico and a degree in Piano Performance from the National Conservatory of Music of Mexico.

Katerina Kosta is a senior machine learning researcher at ByteDance AI lab. She pursued her Ph.D. from the Centre for Digital Music at the Computer Science and Electronic Engineering department of Queen Mary University of London, conducting research on computational modelling and quantitative analysis of expressive changes of dynamics in performed music. Research interests during her studies included time series analysis, custom data structures, pattern recognition, audio processing, and machine learning for music synthesis and analysis of perceived emotion in music audio. She received degrees from University of Athens (Mathematics) and Filippos Nakas Conservatory, Athens (Piano), and a Sound and Music Computing Masters from the Music Technology Group at Universitat Pompeu Fabra, Barcelona.

Maarten Grachten is a senior researcher in machine learning for music and sound technology, currently active as an independent machine learning consultant. He holds an M.Sc. in Artificial Intelligence from University of Groningen (The Netherlands, 2001), and a Ph.D. in Computer Science and Digital Communication from Pompeu Fabra University (Spain, 2007). He has worked on computational modeling of musical expression in jazz and classical music since 2001. His work has been funded by European and national research grants at research institutions including the Austrian Research Institute for Artificial Intelligence (OFAI) and Johannes Kepler University (Austria), and has been published in international peer-reviewed conferences and journals.

Tutorial 5

Waveform-based music processing with deep learning

Jongpil Lee, Jordi Pons, and Sander Dieleman

Abstract

A common practice when processing music signals with deep learning is to transform the raw waveform input into a time-frequency representation. This pre-processing step allows having less variable and more interpretable input signals. However, along that process, one can limit the model's learning capabilities since potentially useful information (like the phase or high frequencies) is discarded. In order to overcome the potential limitations associated with such pre-processing, researchers have been exploring waveform-level music processing techniques, and many advances have been made with the recent advent of deep learning.

In this tutorial, we introduce three main research areas where waveform-based music processing can have a substantial impact:

1) Classification: waveform-based music classifiers have the potential to simplify production and research pipelines.
2) Source separation: making possible waveform-based music source separation would allow overcoming some historical challenges associated with discarding the phase.
3) Generation: waveform-level music generation would enable, e.g., to directly synthesize expressive music.

Jongpil Lee received the B.S. degree in electrical engineering from Hanyang University, Seoul, South Korea, in 2015, the M.S. degree, in 2017, from the Graduate School of Culture Technology, Korea Advanced Institute of Science and Technology, Daejeon, South Korea, where he is currently working toward the Ph.D. degree. He interned at Naver Clova Artificial Intelligence Research in the summer of 2017 and at Adobe Audio Research Group in the summer of 2019. His current research interests include machine learning and signal processing applied to audio and music applications.

Jordi Pons is a researcher at Dolby Laboratories. He is finishing a PhD in music technology, large-scale audio collections, and deep learning at the Music Technology Group (Universitat Pompeu Fabra, Barcelona). Previously, he received a MSc in sound and music computing (Universitat Pompeu Fabra, Barcelona), and his BSc was in telecommunications engineering (Universitat Politècnica de Catalunya, Barcelona). He also interned at IRCAM (Paris), at the German Hearing Center (Hannover), at Pandora Radio (USA, Bay Area), and at Telefónica Research (Barcelona).

Sander Dieleman is a Research Scientist at DeepMind in London, UK, where he has worked on the development of AlphaGo and WaveNet. His current research interest is large-scale generative modeling of perceptual signals (images, audio, video). He was previously a PhD student at Ghent University, where he conducted research on feature learning and deep learning techniques for learning hierarchical representations of musical audio signals. In the summer of 2014, he interned at Spotify in New York, where he worked on implementing audio-based music recommendation using deep learning on an industrial scale.

Tutorial 6

Fairness, Accountability and Transparency in Music Information Research (FAT-MIR)

Andre Holzapfel, Marius Miron, Bob L. Sturm, Emilia Gómez

Abstract

This tutorial focuses on the timely issues of ethics, fairness, accountability and transparency, with particular attention paid to research in applications in music information research. These topics arise from a broader consideration of ethics in the field – related work of which was recently published in TISMIR (https://transactions.ismir.net/articles/10.5334/tismir.13). These topics are also receiving attention in the broader domain of machine learning and data science, e.g., the FAT-Machine Learning (ML) conference 2014-2018, Explainable AI workshops 2017-2018, Interpretable Machine Learning workshops, and in the context of the HUMAINT project and winter school on ethical, legal, social and economic impact of Artificial Intelligence (https://ec.europa.eu/jrc/communities/en/community/humaint). This tutorial is suitable for researchers and students in MIR working in any domain, as these issues are relevant for all MIR tasks, from low- to high-level, from system to user-centered research. There are no prerequisites for taking this tutorial.

Andre Holzapfel received M.Sc. and Ph.D. degrees in computer science from the University of Crete, Greece, and a second Ph.D. degree in music from the Centre of Advanced Music Studies (MIAM) in Istanbul, Turkey. He worked at several leading institutes in computer engineering as postdoctoral researcher, with a focus on rhythm analysis in music information retrieval. His field work in ethnomusicology was mainly conducted in Greece, with Cretan dance being the subject of his second dissertation. In 2016, he became Assistant Professor in Media Technology at the KTH Royal Institute of Technology in Stockholm, Sweden. Since then, his research subjects incorporate the computational analysis of human rhythmic behavior by means of sensor technology, and the investigation of ethical aspects of computational approaches to music.

Marius Miron is a Postdoctoral researcher for European Commission's Joint Research Centre within the project HUMAINT, working on fairness and interpretable machine learning and on assessing the influence of artificial intelligence on humans. He has a PhD (2018) in Computer Science (Audio Signal Processing and Machine Learning) from Pompeu Fabra University, Barcelona. His PhD thesis concerned separating the audio corresponding to the instruments in an orchestral music mixture. He has completed internships at Computational Perception Group, Johannes Kepler University, Linz where he worked on deep learning for source separation, and at Telefonica Research, Barcelona, where he worked on catastrophic forgetting in machine learning. During 2011-2013 he was a research engineer for the research institute INESC in Porto for a project aiming at modelling groove in music.

Bob L. Sturm received the B.A. degree in physics from University of Colorado, Boulder in 1998, the M.A. degree in Music, Science, and Technology, at Stanford University, in 1999, the M.S. degree in multimedia engineering in the Media Arts and Technology program at University of California, Santa Barbara (UCSB), in 2004, and the M.S. and Ph.D. degrees in Electrical and Computer Engineering at UCSB, in 2007 and 2009. In Dec. 2014, he became a

Lecturer at the Centre for Digital Music at Queen Mary University of London. In July 2018 he became an associate professor of computer science at the Royal Institute of Technology KTH in Stockholm Sweden.

Emilia Gómez leads the HUMAINT (HUman and MAchine INTelligence) team at the Centre for Advanced Studies, Joint Research Centre (European Commission) and the MIR (Music Information Research) lab of the Music Technology Group, Universitat Pompeu Fabra in Barcelona. Her research background is in music information retrieval, where she has particularly focused on pitch, melody and tonal processing of music audio signals. She also researches more widely on the impact of artificial intelligence technologies on human behaviour. She is a Telecommunication Engineer (Universidad de Sevilla, Spain), DEA in Acoustics, Signal Processing and Computer Science applied to Music (IRCAM, Paris) and Ph.D. in Computer Science (Universitat Pompeu Fabra). Emilia Gómez has co-authored more than 130 scientific publications in peer-reviewed scientific journals and conferences, and contributed to several open datasets and software libraries. She is currently president of the International Society for Music Information Retrieval (ISMIR), and particularly interested in promoting diversity in the MIR field.