

Tutorial 5

Deep learning for automatic mixing

Christian J. Steinmetz, Soumya Sai Vanka, Gary Bromham, Marco A. Martínez Ramírez

Abstract

Mixing is a central task within audio post-production where expert knowledge is required to deliver professional quality content, encompassing both technical and creative considerations. Recently, deep learning approaches have been introduced that aim to address this challenge by generating a cohesive mixture of a set of recordings as would an audio engineer. These approaches leverage large-scale datasets and therefore have the potential to outperform traditional approaches based on expert systems, but bring their own unique set of challenges. In this tutorial, we will begin by providing an introduction to the mixing process from the perspective of an audio engineer, along with a discussion of the tools used in the process from a signal processing perspective. We will then discuss a series of recent deep learning approaches and relevant datasets, providing code to build, train, and evaluate these systems. Future directions and challenges will be discussed, including new deep learning systems, evaluation methods, and approaches to address dataset availability. Our goal is to provide a starting point for researchers working in MIR who have little to no experience in audio engineering so they can easily begin addressing problems in this domain. In addition, our tutorial may be of interest to researchers outside of MIR, but with a background in audio engineering or signal processing, who are interested in gaining exposure to current approaches in deep learning.

Biographies of Presenters

[Christian J. Steinmetz](#) is PhD researcher working with Prof. Joshua D. Reiss within the Centre for Digital Music at Queen Mary University of London. He researches applications of machine learning in audio with a focus on differentiable signal processing. Currently, his research revolves around high fidelity audio and music production, which involves enhancing audio recordings, intelligent systems for audio engineering, as well as applications that augment and extend creativity. He has worked as a Research Scientist Intern at Adobe, Facebook Reality Labs, and Dolby Labs. Christian holds a BS in Electrical Engineering and BA in Audio Technology from Clemson University, as well as an MSc in Sound and Music Computing from the Music Technology Group at Universitat Pompeu Fabra.

[Soumya Sai Vanka](#) is a first year PhD researcher at the Centre for Digital Music, Queen Mary University of London. She is part of the AI and Music, Centre for Doctoral Training. Her research focus is mainly on exploring the idea of Music Mix similarity, Music Mix Style transfer, and Intelligent Multitrack Mixing using Self-Supervised, Semi-Supervised, and Unsupervised Learning architectures. She also writes music, produces and plays saxophone. Her educational background is a mixture of Masters in Physics and Courses in Music Production.

Gary Bromham is a part-time PhD researcher at Queen Mary University of London, researching the role that traditional studio paradigms and retro aesthetics play in intelligent music production systems (2016 -). He has several publications in this field and has contributed a chapter to the recent Routledge publication, ‘Perspectives on Music Production: Mixing Music’ (2017). He was also a research assistant on the EPSRC funded project called FAST (Fusing Audio and Semantic Technologies) where he is employed as an industry advisor (2017 - 2020). In addition to his research interests, Gary is a practising music producer, songwriter and audio engineer, with over 30 years’ experience (1989 - 2020). He has worked with artists as diverse as Bjork, Wham, Blur and U2, during a period that has witnessed several technological changes. Gary is well versed in most popular music making software and has extensive knowledge of using analog hardware, acting as a product designer and specialist for the renowned mixing desk company, Solid State Logic. He is also a frequent guest lecturer and external advisor at several universities in the UK, Norway and Sweden; speaking on songwriting, music production aesthetics and audio engineering and bringing some of his extensive knowledge and experience to both Undergraduate and Master’s degree level programs.

Marco A. Martínez Ramírez is music technology researcher at Sony in the Tokyo R&D center, where he is part of the Creative AI Lab. His research interests lie at the intersection of machine learning, digital signal processing, and intelligent music production, with a primary focus on deep learning architectures for music processing tasks. Previously, he was an audio research intern at Adobe and received his PhD from the Centre for Digital Music at Queen Mary University of London. He has a MSc in digital signal processing from the University of Manchester, UK, and a BSc in electronic engineering from La Universidad de Los Andes, Colombia. Marco also has a background in music production and mixing engineering.