MDX
Music Demixing Workshop

Deep learning for automatic mixing: challenges and next steps

Christian J. Steinmetz[1]

1 Centre for Digital Music, Queen Mary University of London

In partnership with

Abstract

The process of transforming a set of audio recordings into a cohesive mixture for film, radio, or music production encompasses a number of artistic and technical considerations. Due to the inherent complexity in this task, a great deal of training and expertise on the part of the audio engineer is generally required. This has motivated the design of automated or assistive systems for the audio production process that aid both amateur and expert users (Moffat & Sandler, 2019). While classical machine learning approaches and expert systems have been investigated, they often fail to generalize to the diversity and scale of real-world projects, with the inability to adapt to a varying number of sources, capture stylistic elements across genres, or apply sophisticated processing (De Man et al., 2017). Recently, deep learning approaches have shown promise in addressing these limitations and bring the potential to model the complex mixing process directly from data (Martinez Ramirez et al., 2021; Christian J. Steinmetz, 2020; Christian J. Steinmetz et al., 2021). However, automatic mixing poses a number of unique challenges for deep learning approaches including very low tolerance of artifacts, high sample rates, need for interpretability and controllability, along with limited multitrack data. This talk will provide an overview of the field of automatic mixing and outline recent developments in deep learning approaches, with a focus on potential directions for advancement. This overview will include ideas on how source separation may play a role in advancing automatic mixing, such as data generation (Ward et al., 2017) and audio manipulation (Choi et al., 2021), along with tasks like audio effect removal (Gorlow et al., 2014).

Choi, W., Kim, M., Ramírez, M. A. M., Chung, J., & Jung, S. (2021). AMSS-net: Audio manipulation on user-specified sources with textual queries. *arXiv Preprint arXiv:2104.13553*.

De Man, B., Reiss, J. D., & Stables, R. (2017). *Ten years of automatic mixing*.

Gorlow, S., Reiss, J. D., & Duru, E. (2014). Restoring the dynamics of clipped audio material by inversion of dynamic range compression. *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*.

Martinez Ramirez, M., Stoller, D., & Moffat, D. (2021). A deep learning approach to intelligent drum mixing with the wave-u-net. *Journal of the Audio Engineering Society*.

Moffat, D., & Sandler, M. B. (2019). Approaches in intelligent music production. *Arts*.

Steinmetz, Christian J. (2020). *Learning to mix with neural audio effects in the waveform domain* [Master's thesis, Universitat Pompeu Fabra]. https://doi.org/10.5281/zenodo.4091203

Steinmetz, Christian J., Pons, J., Pascual, S., & Serrà, J. (2021). Automatic multitrack mixing with a differentiable mixing console of neural audio effects. *ICASSP*.

Ward, D., Wierstorf, H., Mason, R., Plumbley, M., & Hummersone, C. (2017). Estimating the loudness balance of musical mixtures using audio source separation. *3nd AES Workshop on Intelligent Music Production*.