

Master 2 research internship - 2024

Deep-learning vs model-based approach for non-linear audio effects inversion

Feb. 2024 - Aug. 2024 (6 months)

Supervisor(s) : Dominique Fourer and Hichem Maaref
Team / Laboratory : SIAM / IBISC (EA 4526) - Univ. Évry/Paris-Sacay
Collaborators : IRCAM and TelecomParis
Contact : dominique.fourer@univ-evry.fr

Abstract : Recent audio methods which involve deep learning approaches [2, 7] are not always sufficient for machine learning models, especially when the invariance to unwanted signal transformations (i.e. when working on different datasets) is needed. Indeed, neglecting the audio quality in machine listening algorithms can lead to unstable results due to a “horse”. For example, the “album” and the “artist” effects are among well-known problems reported by music signal processing researchers [5]. Such problems can be explained by a lack of generalization of the trained audition model which focuses on unwanted and irrelevant audio features resulting from the recording conditions and/or the production effects related to “audio quality”. Hence, the goal of the present internship research is to develop methods for analyzing one or several audio effects related to audio quality and to propose the corresponding inversion/cancellation method allowing the restoration of the original signal.

keywords : audio processing, music information retrieval, dynamic range compression, deep learning

Goals

- Bibliographical study for identifying existing inversion methods for non-linear audio effects such as dynamic range compression [8, 6]
- Implementation/proposal of new techniques for analyzing and inverting the investigated audio effects
- Proof of concept by using the proposed method to improve an existing machine-learning-based MIR system

Methodology

As instrumental sound “timbre” is defined as all sound characteristics which are not related to pitch, loudness and duration, “audio quality” refers here as everything related to the sound characteristics which is not related to the content sources. This therefore includes the choice of microphone, recording media (tapes, vinyls, digital and related potential artifacts), audio production chain (equalization, compression, reverberation) and diffusion (such as mp3 data-reduction). Some of them are considered as artifacts (or degradation), some of them are related to artistic choices. Of course, when content sources are artificial (such as synthesized) separating both content and quality is ill-defined. Hence, audio quality is defined as the set of audio effects independent from the signal content creating a unique listening experience.

As a first step, the recruited internship researcher will propose analysis-synthesis audio processing chains for a selection of audio effects to pursue our previous work related to audio quality prediction [1, 6]. Then, we will focus on the inversion of one or several effects and investigate the best approach by comparing a model-based approach (eg. [3] for dynamic range compression) with a deep learning approach [4]. During this study, we plan to identify all the relevant parameters of each audio effect and to compare the prediction and inversion/cancellation results provided by the competing approaches. More precisely, we will compare the proposed methods in terms of prediction accuracy and in terms of objective and perceptual signal quality for the audio restoration problem.

Finally, we will investigate the relevance of the audio quality simulation (data augmentation) and/or restoration in the context of a music information retrieval (MIR) task such as music genre detection using a machine-learning-based state-of-the-art method.

Required profile

- good machine learning and signal processing knowledge
- mathematical understanding of the formal background
- excellent programming skills (Python, Matlab, C/C++, keras, pytorch, tensorflow, etc.)
- good motivation, high productivity and methodical works

Salary an perspectives

According to background and experience (a minimum of 577.50 euros/month). Possibility to pursue with a 3-year-funded PhD contract in the field of audio/music signal processing (ANR Project).

Références

- [1] D. Fourer and G. Peeters. Objective characterization of audio signal quality : Application to music collection description. In *Proc. IEEE ICASSP*, pages 711–715, New Orleans, LA, USA, March 2017.
- [2] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [3] S. Gorlow and S. Marchand. Reverse engineering stereo music recordings pursuing an informed two-stage approach. In *Proc. Digital Audio Effects Conf. (DAFx'13)*, 2013.
- [4] Johannes Imort, Giorgio Fabbro, Marco A Martínez Ramírez, Stefan Uhlich, Yuichiro Koyama, and Yuki Mitsufuji. Removing distortion effects in music using deep neural networks. *arXiv preprint arXiv :2202.01664*, 2022.
- [5] Y. E Kim, D. S Williamson, and S. Pilli. Towards quantifying the album effect in artist identification. In *International Society on Music Information Retrieval Conference (ISMIR)*, 2006.
- [6] Côme Peladeau and Geoffroy Peeters. Blind estimation of audio effects using an auto-encoder approach and differentiable signal processing. *arXiv preprint arXiv :2310.11781*, 2023.
- [7] Wenwu Wang. *Machine Audition : Principles, Algorithms and Systems : Principles, Algorithms and Systems*. IGI Global, 2010.
- [8] Udo Zölzer, Xavier Amatriain, Daniel Arfib, Jordi Bonada, Giovanni De Poli, Pierre Dutilleul, Gianpaolo Evangelista, Florian Keiler, Alex Loscos, Davide Rocchesso, et al. *DAFX-Digital audio effects*. John Wiley & Sons, 2002.