

Classical GAS - Classical methods for Generative Audio Synthesis.

Keywords: Audio synthesis, C/C++ programming, generative models, frugal AI, deep learning,
Gratification: In accordance with current legislation
Laboratory: LISTIC, Université Savoie Mont Blanc
Location: Annecy, France
Duration: 6 months starting from February/Mars 2025

Context:

Signal-based audio synthesis methods like the one described in [1, 2] have seen a regain of interest recently after the introduction of the Differential Digital Signal Processing (DDSP) package [3, 4]. It resort in using a neural architecture to produce control parameters for a signal based synthesis method. The original DDSP implementation perform succesfully in a variety of audio synthesis tasks (synthesis, transfer, ...) using a Variational Autoencoder and a Sine plus noise generative model. It offers an expressive generative model without compromising explainability and control as well as an opportunity for real time sound synthesis.

In contrast, diffusion models lacks interpretability and intuitive control, need an intensive training and suffer from a long inference time. Yet, the training of a generative model remains challenging due the unstable gradients produced by signal-based audio synthesis methods. Notably, the success of such a training of a DDSP inspired generative model is impacted by (i) the signal based synthesis method, (ii) the control neural architecture (iii) the training methodology.

In [5], a GAN-based synthesizer (StyleWaveGAN, SWG) have shown promising results on percussion synthesis with improved expressivity in control, opening way to instruments synthesis using DDSP inspired models.

Project summary:

During this internship, a study of DDSP based audio generative model will be performed. After a comprehensive study of DDSP inspired architectures on piano and drums datasets [gillet2006enst, 6], the students will extend the work proposed in [5] to more complex signal-based synthesizers. In particular the scalability of controls to achieve a perceptive control method similar to [7] will be studied. The intership will result in the implementation of a DDSP as virtual instrument on a Digital Audio Workstation (DAW) or on an embedded system.

Candidate Profile: She/he should be enrolled in a M2 or engineer diploma in one or more of the following fields: signal and image processing, computer science, embedded systems. The candidate should have strong progamming abilities as well as good writing and oral communication skills. A strong interest and/or experience with audio signal processing will be appreciated.

Contacts and Application procedure:

Send a detailed CV and motivation letter to antoine.lavault@univ-smb.fr and yassine.mhiri@univ-smb.fr

References:

- [1] Xavier Serra. "Musical Sound Modeling with Sinusoids plus Noise". In: *1997* (1997).
- [2] John Chowning. "The Synthesis of Complex Audio Spectra by Means of Frequency Modulation". In: *Journal of The Audio Engineering Society* 21 (1973), pp. 526–534.
- [3] Jesse Engel, Lamtharn (Hanoi) Hantrakul, Chenjie Gu, and Adam Roberts. "DDSP: Differentiable Digital Signal Processing". In: *International Conference on Learning Representations*. 2020.
- [4] Lenny et al. Renault. "DDSP-Piano: A Neural Sound Synthesizer Informed by Instrument Knowledge". In: *Journal of the Audio Engineering Society* (2023).
- [5] Antoine Lavault. "Generative Adversarial Networks for Synthesis and Control of Drum Sounds". PhD thesis. Sorbonne Université, 2023.
- [6] Curtis Hawthorne, Andriy Stasyuk, Adam Roberts, Ian Simon, Cheng-Zhi Anna Huang, Sander Dieleman, Erich Elsen, Jesse Engel, and Douglas Eck. "Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset". In: *International Conference on Learning Representations* (2019).
- [7] Antoine Lavault, Axel Roebel, and Matthieu Voiry. "STYLEWAVEGAN: STYLE-BASED SYNTHESIS OF DRUM SOUNDS WITH EXTENSIVE CONTROLS USING GENERATIVE ADVERSARIAL NETWORKS". In: *19th Sound and Music Computing Conference*. 2022.