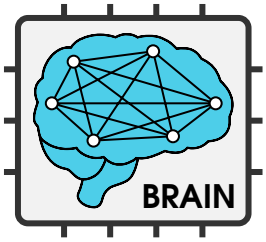

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

anr agence nationale de la recherche

PhD Position: Hybrid Deep Learning (AI) Models for Interpretable Music Analysis

Project Context

This fully funded Ph.D. position is central to the MusAIc project, an ANR JCJC-funded initiative (2026-2029) that aims to develop steerable and interpretable deep learning models for music information retrieval (MIR).

Current deep learning models for MIR essentially fall into two categories:

1. **Low-Rank Factorization Models**: Low-Rank factorisation methods, such as Nonnegative Matrix Factorization (NMF) [1], produce results that are interpretable and steerable, allowing experts to understand and guide their outputs [2, 3, 4, 5]. However, they often struggle with performance and scalability.
2. **Deep Learning Models**: Deep learning models are nowadays the state-of-the-art methods in many MIR tasks, such as automatic transcription [6, 7] and source separation [8], but are often difficult to interpret or control, which limits their use by musicians and musicologists.

The MusAIc project aims to bridge this gap by developing hybrid models [9] that are simultaneously high-performing, interpretable, and steerable.

This internship represents a crucial first step in the project, focusing on enhancing structured models with principles from deep learning.

Ph.D. Thesis Objectives

The objective of this thesis is to develop novel deep learning architectures for music analysis and music information retrieval (MIR) that are interpretable and steerable by design, without sacrificing performance. The research is structured to build upon standard signal processing and deep learning principles. The planned trajectory is as follows:

- **Year 1: Developing Deep Low-Rank Factorization models.** The initial research will focus on integrating principles from deep learning (notably high-scale learning and increased depth in models) to enhance the expressivity and scalability of nonnegative low-rank factorization methods [2, 3, 4, 5, 10]. This involves evaluating deep NMF models [11] and, critically, developing novel algorithms for **deep nonnegative tensor factorizations** (deep NTD and NCP), a key theoretical contribution of the project.

- **Year 2: Transition to Hybrid Architectures.** Building on the expertise from Year 1, the research will shift during the second year towards hybrid models. The first step will involve adapting existing efficient architectures, such as **neural audio codecs** [12, 13], for interpretability. You will investigate how injecting structured priors (*e.g.,* nonnegativity, sparsity) into their latent spaces can lead to musically meaningful and controllable representations. Structured priors will be motivated by results obtained during Year 1.
- **Year 3: Synthesis and By-Design Hybrid Models.** The final year will be dedicated to the core ambition of the project: designing novel hybrid models from the ground up. Leveraging insights from the previous phases (and in collaboration with a postdoctoral researcher focusing on model adaptation) and existing literature [14, 15], you will develop architectures that natively embed properties of steerability and interpretability, moving beyond retrofitting constraints to existing models.

Throughout the thesis, the developed models will be applied to and evaluated on core Music Information Retrieval (MIR) tasks like Automatic Music Transcription, Source Separation, and Structure Analysis.

An additional objective, expected to start during Year 2, will be to develop models jointly *for* and *with* music experts (musicians, musicologists, sound engineers, ...). In particular, meetings with volunteer music experts will be organized in order to propagate knowledge and developments, and to modify future developments in line with their needs.

Candidate Profile

We are seeking an outstanding and motivated candidate with a strong research potential.

- **Required Skills**:
  - A Master's degree in Computer Science, Applied Mathematics, Signal Processing, or a related field;
  - Programming skills in **Python**.
  - A solid mathematical background, especially in **linear algebra** and **optimization**.
  - Ability to work independently, think critically, and formalize creative ideas. A strong interest in fundamental research is required.
- **Desired Skills (ideal – can be acquired during the Ph.D.)**:
  - Previous experience with deep learning frameworks (*e.g.,* PyTorch);
  - Previous experience with high-performance computing (*e.g.,* Slurm jobs management);
  - Previous experience in audio signal processing or Music Information Retrieval (MIR) is a major plus;
  - A personal interest in music is highly valued.

What We Offer

- A **fully funded 3-year Ph.D. position**, and fundings for attending conferences, workshops, and international research stays;
- Integration in a research environment within the BRAIN team at IMT Atlantique;

- Supervision by a team of dedicated researchers, with opportunities for collaboration with leading international experts in both low-rank factorization (Prof. Nicolas Gillis) and deep learning for music (Prof. Gaël Richard);
- Opportunities to contribute to open-source software and publish research results (notably in top-tier conferences such as ICASSP, ISMIR, ICML, NeurIPS, ...).

Practical Details

- **Duration**: 36 months.
- **Start Period**: Flexible, between September and December 2026.
- **Location**: IMT Atlantique, Brest, France.
- **Salary**: To be defined according to institutional and national regulations. The minimum is a gross salary of 2.300€/month.
- **How to Apply**: Please send a CV, your most recent academic transcripts, and a cover letter detailing your interest in the topic (please, do not make a long, generic cover letter; a small but honest and candid cover letter will be valued).

Contact

For any questions or to submit your application, please contact:

Axel MARMORET
Associate Professor, IMT Atlantique (Lab-STICC)
axel.marmoret@imt-atlantique.fr
https://ax-le.github.io

References

[1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[2] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *2003 IEEE Workshop Applications Signal Process. Audio Acoustics (WASPAA)*, pp. 177–180, IEEE, 2003.

[3] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis," *Neural computation*, vol. 21, no. 3, pp. 793–830, 2009.

[4] H. Wu, A. Marmoret, and J. E. Cohen, "Semi-supervised convolutive nmf for automatic music transcription," in *Proc. 19th Sound and Music Computing Conf.*, 2022.

[5] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 3, pp. 550–563, 2009.

[6] R. M. Bittner, J. J. Bosch, D. Rubinstein, G. Meseguer-Brocal, and S. Ewert, "A lightweight instrument-agnostic model for polyphonic note transcription and multipitch estimation," in *ICASSP 2022-2022 IEEE Int. Conf. Acoustics, Speech, Signal Process.*, pp. 781–785, IEEE, 2022.

[7] R. Wu, X. Wang, Y. Li, W. Xu, and W. Cheng, "Piano transcription with harmonic attention," in *ICASSP 2024-2024 IEEE Int. Conf. Acoustics, Speech, Signal Process.*, pp. 1256–1260, IEEE, 2024.

[8] S. Rouard, F. Massa, and A. Défossez, "Hybrid transformers for music source separation," in *ICASSP 2023-2023 IEEE Int. Conf. Acoustics, Speech, Signal Process.*, IEEE, 2023.

[9] G. Richard, V. Lostanlen, Y.-H. Yang, and M. Müller, "Model-based deep learning for music information research: Leveraging diverse knowledge sources to enhance explainability, controllability, and resource efficiency," *IEEE Signal Process. Mag.*, 2025.

[10] A. Marmoret, J. Cohen, N. Bertin, and F. Bimbot, "Uncovering audio patterns in music with nonnegative Tucker decomposition for structural segmentation," in *ISMIR*, pp. 788–794, 2020.

[11] V. Leplat, L. T. K. Hien, A. Onwunta, and N. Gillis, "Deep nonnegative matrix factorization with beta divergences," *Neural Computation*, vol. 36, no. 11, pp. 2365–2402, 2024.

[12] N. Zeghidour, A. Luebs, A. Omran, J. Skoglund, and M. Tagliasacchi, "Soundstream: An end-to-end neural audio codec," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 30, pp. 495–507, 2021.

[13] A. Défossez, J. Copet, G. Synnaeve, and Y. Adi, "High fidelity neural audio compression," *arXiv preprint arXiv:2210.13438*, 2022.

[14] J. Parekh, S. Parekh, P. Mozharovskyi, G. Richard, and F. d'Alché Buc, "Tackling interpretability in audio classification networks with non-negative matrix factorization," *IEEE/ACM Trans. Audio, Speech, Language Process.*, 2024.

[15] M. Lebourdais, T. Mariotte, A. Almudévar, M. Tahon, and A. Ortega, "Explainable by-design audio segmentation through non-negative matrix factorization and probing," in *Interspeech 2024*, 2024.