

Master 2 Internship offer: Large-Scale Learning with NMF for Music Transcription

Project Context

This internship is part of the [MusAlc](#) project, an ANR JCJC-funded initiative (2026-2029) that aims to develop steerable and interpretable deep learning models for music information retrieval (MIR). The project also includes a Ph.D. position, expected to start in late 2026, which could hence follow the proposed internship.

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 MusAlc project aims to bridge this gap by developing hybrid models 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.

Internship Description and Objectives

The primary goal of this internship is to investigate how traditional structured models, specifically non-negative low-rank factorizations (in particular, Nonnegative Matrix Factorization – NMF [1, 2] and Nonnegative Tucker Decomposition – NTD [9, 10]), can be adapted for large-scale learning by incorporating training techniques from the field of deep learning.

Low-rank factorization techniques have proven helpful in music analysis, particularly in music information retrieval (MIR) tasks such as music transcription [2, 3, 4], source separation [5], and structure analysis [11]. Nonetheless, they were not originally designed to process massive datasets efficiently and may hence suffer in comparison with deep learning methods, which scale with data. The goal of the project is therefore to extend and evaluate low-rank factorization techniques when leveraging large datasets for learning relevant and rich representations.

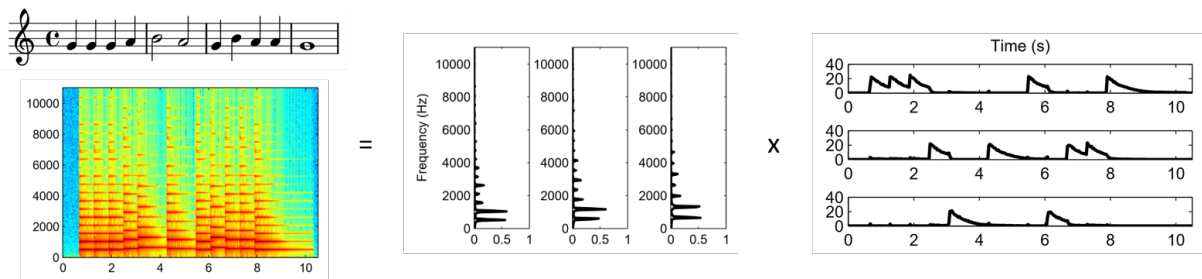


Figure 1: An example of NMF for Music Transcription. The spectrogram (left, also represented with its score) is factorized into two matrices, containing respectively frequency templates (notes) and temporal activations (onsets of each note).

The main tasks will include:

- **Implementation:** Implement batch-wise and epoch-based training procedures for NMF and NTD. This will be done in Python, likely leveraging libraries such as TensorLy [12].
- **Evaluation:** Design and conduct experiments on large-scale music datasets such as [13, 14, 15, 16] to compare the performance of these new models against traditional low-rank factorization methods [4, 17] and deep learning baselines [6, 7]. The internship will focus on automatic music transcription, but the developed methods could be extended to other tasks such as source separation and structure analysis.
- **Analysis:** Assess the trade-offs between model performance, computational scalability, and the interpretability of the results.

Candidate Profile

We are seeking a highly motivated Master's 2 student in Computer Science, Applied Mathematics, Signal Processing, or a related field.

- **Required Skills:**
 - Programming skills in **Python**;
 - Knowledge and skills in **machine & deep learning**;
 - Interest and/or curiosity in research.
 - **Desired Skills (ideal – can be acquired during the internship):**
 - Previous experience with deep learning frameworks (e.g., PyTorch);
 - Previous experience with high-performance computing (e.g., Slurm jobs management);
 - Knowledge of audio signal processing;
 - A demonstrated interest in music or MIR is a significant plus.
-

What We Offer

- Integration in a research environment within the [BRAIN team](#) at IMT Atlantique.
- Guidance from experts in low-rank factorizations, deep learning, and audio signal processing.

- Opportunities to contribute to open-source software and publish research results.
 - This internship is designed as a potential gateway to a **fully funded Ph.D. thesis** within the MusAlc project (starting late 2026).
-

Practical Details

- **Duration:** 6 months.
 - **Start Period:** Flexible, between February and May 2026.
 - **Location:** IMT Atlantique, Brest, France.
 - **Gratification:** To be defined according to institutional and national regulations (legal minimum).
 - **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] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.
- [10] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

- [11] 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.
- [12] J. Kossaifi, Y. Panagakis, A. Anandkumar, and M. Pantic, "Tensorly: Tensor learning in python," *Journal Machine Learning Research (JMLR)*, vol. 20, no. 26, 2019.
- [13] V. Emiya, N. Bertin, B. David, and R. Badeau, "Maps-a piano database for multipitch estimation and automatic transcription of music," 2010.
- [14] C. Hawthorne, A. Stasyuk, A. Roberts, I. Simon, C.-Z. A. Huang, S. Dieleman, E. Elsen, J. Engel, and D. Eck, "Enabling factorized piano music modeling and generation with the maestro dataset," *arXiv preprint arXiv:1810.12247*, 2018.
- [15] E. Manilow, G. Wichern, P. Seetharaman, and J. Le Roux, "Cutting music source separation some slakh: A dataset to study the impact of training data quality and quantity," in *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 45–49, IEEE, 2019.
- [16] Q. Xi, R. M. Bittner, J. Pauwels, X. Ye, and J. P. Bello, "Guitarset: A dataset for guitar transcription.," in *ISMIR*, pp. 453–460, 2018.
- [17] E. Benetos, S. Dixon, Z. Duan, and S. Ewert, "Automatic music transcription: An overview," *IEEE Signal Process. Mag.*, vol. 36, no. 1, pp. 20–30, 2018.