INTERNATIONAL AUDIO LABORATORIES ERLANGEN
A joint institution of Fraunhofer IIS and Universität Erlangen-Nürnberg

SOURCE SEPARATION OF PIANO CONCERTOS WITH TEST-TIME ADAPTATION

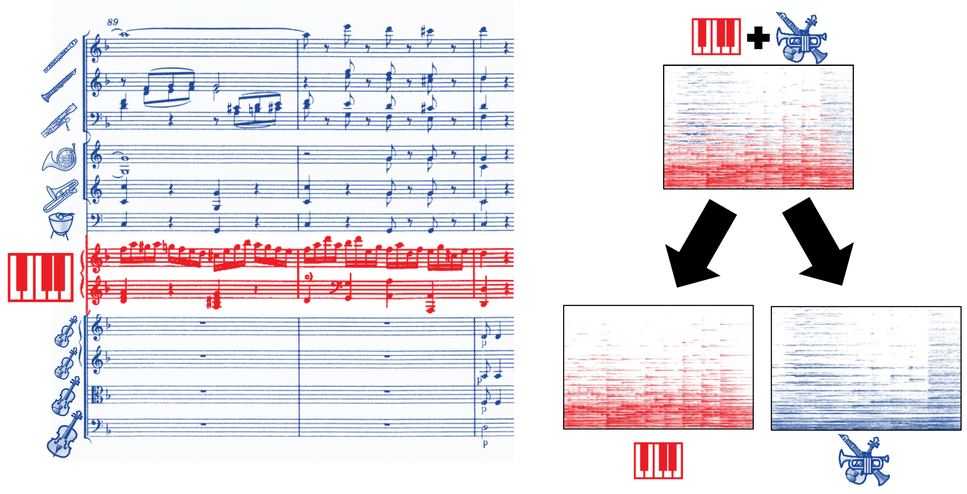
Yigitcan Özer, Meinard Müller

Abstract

Music source separation (MSS) aims at decomposing a music recording into its constituent sources, such as a lead instrument and the accompaniment. Despite the difficulties in MSS due to the high correlation of musical sources in time and frequency, deep neural networks (DNNs) have led to substantial improvements to accomplish this task. For training supervised machine learning models such as DNNs, isolated sources are required. In the case of popular music, one can exploit open-source datasets which involve multitrack recordings of vocals, bass, and drums. For western classical music, however, isolated sources are generally not available. In this work, we consider the case of piano concertos, which is a genre composed for a pianist typically accompanied by an orchestra. The lack of multitrack recordings makes training supervised machine learning models for the separation of piano and orchestra challenging. To overcome this problem, we generate artificial training material by randomly mixing sections of the solo piano repertoire (e.g., piano sonatas) and orchestral pieces without piano (e.g., symphonies) to train state-of-the-art DNN models for MSS. As our main contribution, we propose a test-time adaptation (TTA) procedure, which exploits random mixtures of the piano-only and orchestra-only parts in the test data to further improve the separation quality.

Source Separation of Piano Concertos

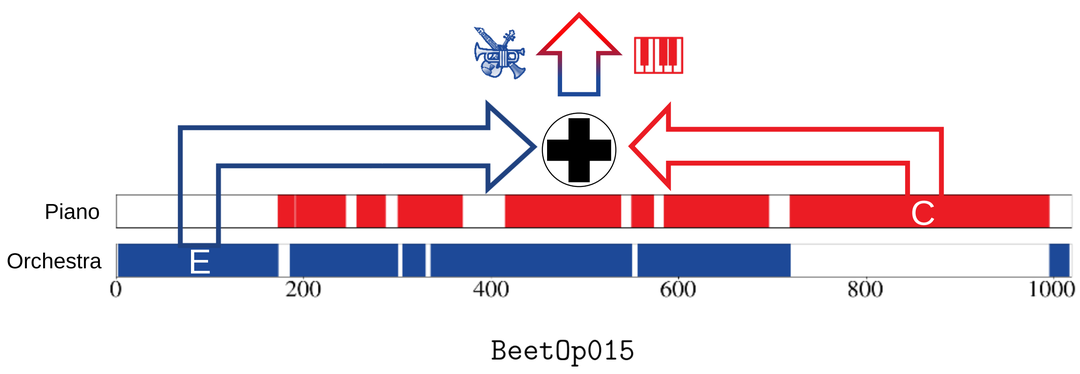
- Task: Lead-accompaniment source separation
  - Piano & orchestra in piano concertos
- Spectral-based music source separation with U-Net [1,2]



Training

- Pre-trained model (PT) trained with artificial mixes
  - Solo piano repertoire (e.g., piano sonatas)
  - Orchestral works without piano (e.g., symphonies)
- Model finetuning with test-time adaptation (TTA) [3]
  - Seven annotated piano concertos
  - Random mixes sampled from the test data
    - Piano-only parts (e.g., cadenza)
    - Orchestra-only parts (e.g., exposition)
  - These piano concertos share the same harmonic and acoustic properties, since they originate from the same recording.

Piano-Orchestra Mixture



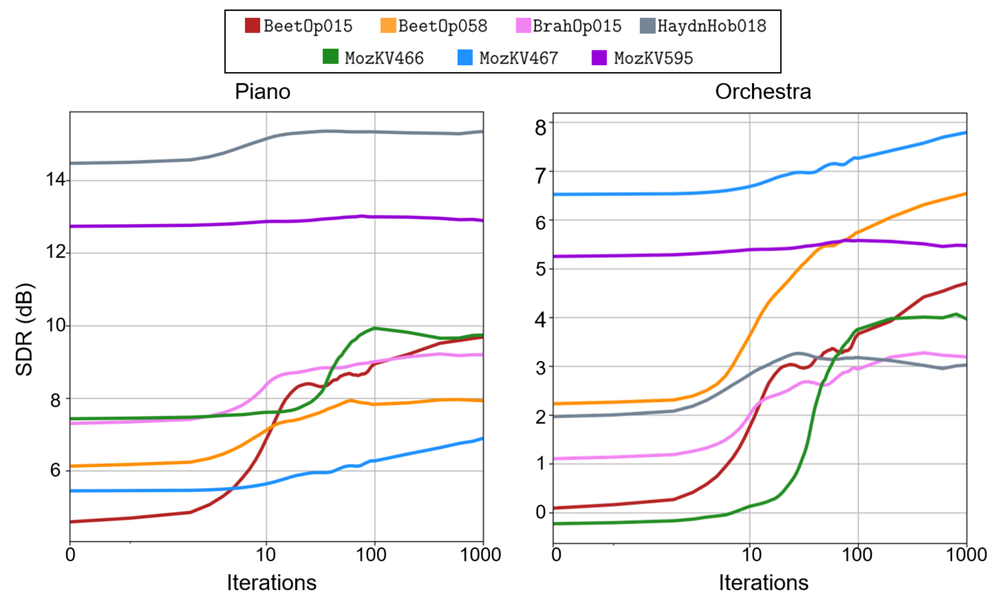
BeetOp015

- Test data: 30-second random mixes sampled from the annotated piano concertos

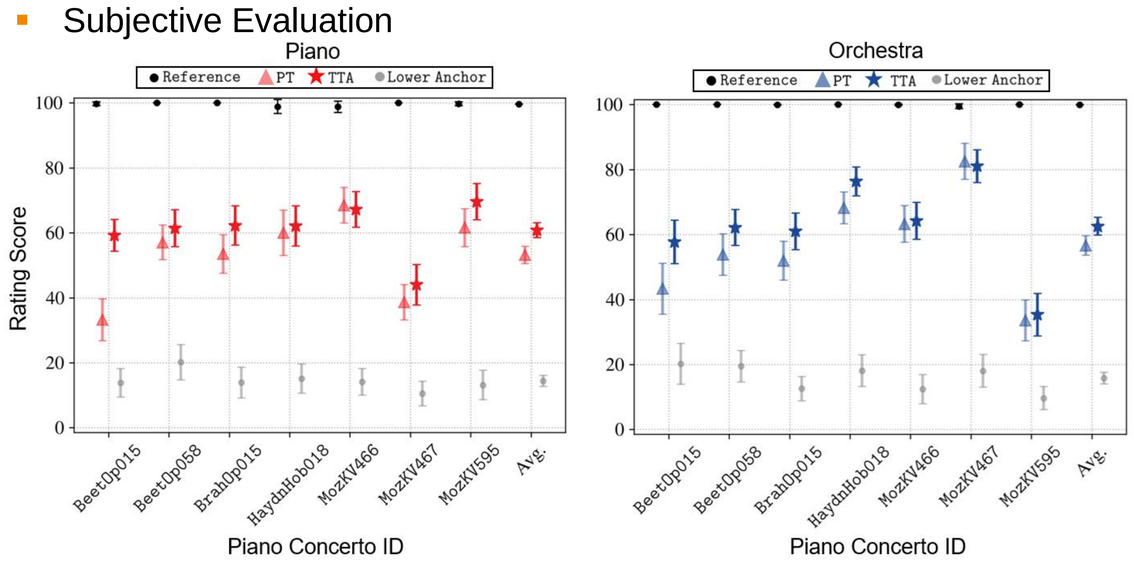
https://www.audiolabs-erlangen.de/resources/MIR/2022-PianoSep

Results

- Quantitative Evaluation: evolution of SDR values per iteration



- Subjective Evaluation



- TTA, in particular, enhances the separation quality of historical recordings.

References

[1] R. Hennequin, A. Khlif, F. Voituret, and M. Moussallam, “Spleeter: a fast and efficient music source separation tool with pre-trained models,” Journal of Open Source Software, 2020.

[2] A. Jansson, E. J. Humphrey, N. Montecchio, R. M. Bittner, A. Kumar, and T. Weyde, “Singing voice separation with deep U-net convolutional networks,” in Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Suzhou, China, 2017.

[3] Y. Sun, X. Wang, L. Zhang, J. Miller, M. Hardt, and A. A. Efros, “Test-time training with self-supervision for generalization under distribution shifts,” in Proceedings of the International Conference on Machine Learning (ICML), 2020.

Acknowledgements

This work was supported by the German Research Foundation (DFG MU 2686/10-2). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institut für Integrierte Schaltungen IIS.