

Using the Sync Toolbox for an Experiment on High-Resolution Music Alignment

Yigitcan Özer, Michael Krause, Meinard Müller



Abstract

Music synchronization aims to automatically align multiple music representations such as audio recordings, MIDI files, and sheet music. For this task, we have recently published the Sync Toolbox [1], an open-source Python package for efficient, robust, and accurate music synchronization. This work combines spectral flux used as onset features with conventional chroma features to increase the alignment accuracy. We conduct some experiments within the Sync Toolbox framework to show that our approach preserves the accuracy compared with another high-resolution approach while being computationally simpler.

Dataset

Schubert Winterreise Dataset (SWD) [5] comprises several representations of the song cycle Winterreise D911 (Op. 89).

We conduct our experiments using:

- 24 songs composed for solo voice with piano accompaniment
- Music recordings by Gerhard Hüscher and Randall Scarlata
- Measure annotations

Sync Toolbox

The Sync Toolbox [1] is a recent open-source Python package that provides all the components of a music synchronization pipeline to produce state-of-the-art alignment results regarding efficiency and accuracy.

Its algorithmic core is based on the high-resolution multiscale dynamic time warping (DTW) approach from [2] and uses

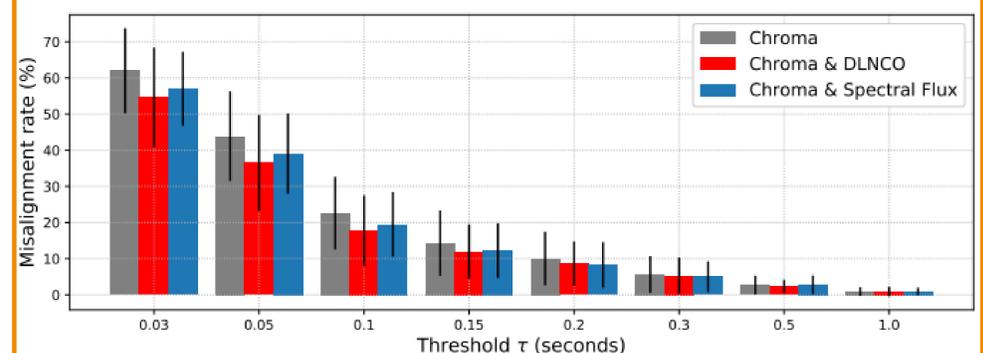
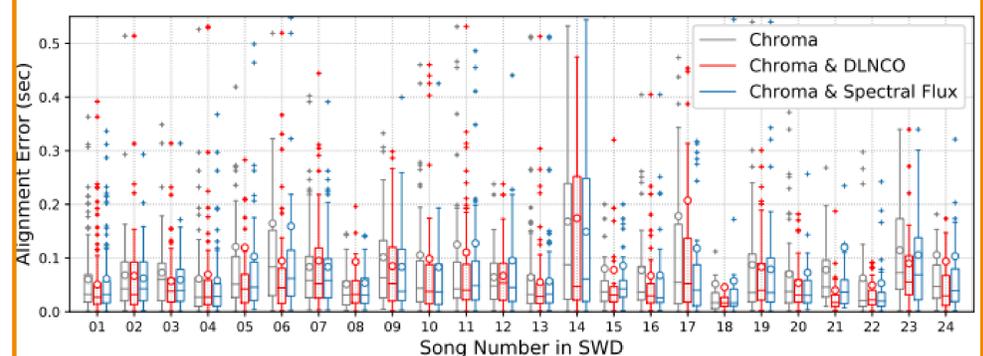
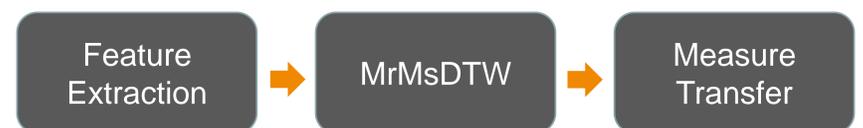
- Chroma features
- High-resolution chroma onset features from [3]
- Spectral Flux from [4]

synctoolbox



<https://github.com/meinardmueller/synctoolbox>

Performance Evaluation



Comparison of alignment with the following features / feature combinations:

- Chroma
- Chroma + DLNCO
- Chroma + Spectral Flux

References

- [1] M. Müller, Y. Özer, M. Krause, T. Prätzlich, and J. Driedger, "Sync Toolbox: A Python package for efficient, robust, and accurate music synchronization," *Journal of Open Source Software (JOSS)*, vol. 6, no. 64, pp. 3434:1–4, 2021.
- [2] T. Prätzlich, J. Driedger, and M. Müller, "Memory-restricted multiscale dynamic time warping," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Shanghai, China, March 2016, pp. 569–573.
- [3] S. Ewert, M. Müller, and P. Grosche, "High resolution audio synchronization using chroma onset features," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 1869–1872.
- [4] P. Grosche, M. Müller, and S. Ewert, "Combination of onset-features with applications to high-resolution music synchronization," in *Proceedings of the International Conference on Acoustics (NAG/DAGA)*, 2009, pp. 357–360.
- [5] C. Weiß, F. Zalkow, V. Arifi-Müller, M. Müller, H. V. Koops, A. Volk, and H. Grohgan, "Schubert Winterreise dataset: A multimodal scenario for music analysis," *ACM Journal on Computing and Cultural Heritage (JOCCH)*, vol. 14, no. 2, pp. 25:1–18, 2021.

Acknowledgments

This work was supported by the German Research Foundation (DFG MU 2686/7-2, MU 2686/11-1). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institut für Integrierte Schaltungen IIS.