The Journal of Open Source Software

Sync Toolbox: A Python Package for Efficient, Robust, and Accurate Music Synchronization

Meinard Müller[1], **Yigitcan Özer**[1], **Michael Krause**[1], **Thomas Prätzlich**[1], **and Jonathan Driedger**[1]

1 International Audio Laboratories Erlangen

DOI: 10.21105/joss.03434

Software

- Review
- Repository
- Archive

Editor: Fabian-Robert Stöter

Reviewers:

- @lutzhamel
- @magdalenafuentes

Submitted: 10 June 2021
Published: 27 August 2021

Summary

Music can be described and represented in many different ways, including sheet music, symbolic representations, and audio recordings (Müller, 2015). For each of these representations, there may exist different versions (e.g., recordings performed by different orchestras and conductors) that correspond to the same musical work. Music information retrieval (MIR) aims at developing techniques and tools for organizing, understanding, and searching this information in a robust, efficient, and intelligent manner. In this context, various alignment and synchronization procedures have been developed with the common goal to automatically link several types of music representations, thus coordinating the multiple information sources related to a given musical work. In the design and implementation of synchronization algorithms, one has to deal with a delicate tradeoff between efficiency, robustness, and accuracy—requirements leading to various approaches with many design choices. In this contribution, we introduce a Python package called *Sync Toolbox*, which provides open-source reference implementations for full-fledged music synchronization pipelines and yields state-of-the-art alignment results for a wide range of Western music. Using suitable feature representations and cost measures, the toolbox's core technology is based on *dynamic time warping* (DTW), which brings the feature sequences into temporal correspondence. To account for efficiency, robustness and, accuracy, our toolbox integrates and combines techniques such as multiscale DTW (MsDTW) (Müller et al., 2006; Salvador & Chan, 2004), memory-restricted MsDTW (MrMsDTW) (Prätzlich et al., 2016), and high-resolution music synchronization (Ewert et al., 2009). While realizing a complete system with presets that allow users to reproduce research results from the literature, our toolbox also provides well-document functions for all required basic building blocks for feature extraction and alignment. Furthermore, the toolbox contains example code for visualizing, sonifying, and evaluating synchronization results, thus deepening the understanding of the techniques and data.

Statement of Need

The task of finding an alignment between two feature sequences has received large research interest in the past, in the context of MIR and beyond. In the music domain, alignment techniques are central for applications such as score following, content-based retrieval, automatic accompaniment, or performance analysis (Arzt, 2016; Müller, 2015). Beyond these classical applications, alignment techniques have gained in importance in view of recent data-driven machine learning techniques. In particular, music synchronization has shown its potential for facilitating data annotation, data augmentation, and model evaluation. To be more specific, for certain types of music one often has a score-like symbolic representation that explicitly encodes information such as note events, measure positions, lyrics, and other types of metadata.

Furthermore, music experts often provide their harmonic, structural, or rhythmic analyses using such symbolic reference representations. Music synchronization techniques then allow for (semi-)automatically transferring these manually generated annotations from the reference to other symbolic or audio representations. This is beneficial in particular for music, where one has many recorded performances of a given piece. Thus, using music synchronization techniques, one may simplify the annotation process and substantially increase the number of annotated training and test versions. For example, in (Zalkow et al., 2017), a multi-version approach for transferring measure annotations between music recordings (Wagner operas) is described. The "Schubert Winterreise Dataset" yields another example where automated techniques were applied to transfer measure, chord, local key, structure, and lyrics annotations (Weiß et al., 2021). Including nine performances (versions) of Schubert's song cycle, this cross-version dataset was used in (Weiß et al., 2020) for training and evaluating data-driven approaches for local key estimation, where the different dataset splits across songs and performances provided new insights into the algorithms' generalization capabilities.

Being a central task, there are a many software packages for sequence alignment of general time series. In the audio domain, the Python packages librosa by (McFee et al., 2015) offers a basic DTW-based pipeline for synchronizing music recordings. Since the complexity of alignment techniques such as DTW is proportional to the product of the feature sequences' lengths, runtime and memory requirements become issues when dealing with long feature sequences. Using a fast online time warping (OLTW) algorithm as described by (Dixon & Widmer, 2005), the software[1] (Music Alignment Tool CHest) allows for an efficient alignment of audio files. While being efficient, such online approaches are prone to local deviations in the sequences to be aligned. An efficient yet robust alternative is offered by offline procedures based on multiscale strategies such as MsDTW (Müller et al., 2006; Salvador & Chan, 2004). The recent Python package linmdtw[2] contains an implementation of MsDTW as well as a linear memory DTW variant described in (Tralie & Dempsey, 2020). Another important issue in music synchronization is the temporal accuracy of the alignments, which may be achieved by considering additional local cues such as onset features (Ewert et al., 2009). Improving the accuracy, however, often goes along with an increase of computational complexity and a decrease of overall robustness.

With our *Sync Toolbox*, we offer a Python package that provides all components to realize a music synchronization pipeline that is robust, efficient, and accurate. First, to account for robustness and efficiency it implements the memory-restricted MsDTW approach from (Prätzlich et al., 2016) as its algorithmic core component. Second, to account for accuracy, it integrates the high-resolution strategy from (Ewert et al., 2009) on the finest MsDTW layer. Third, the toolbox contains all feature extractions methods (including chroma and onset features) needed to reproduce the results from the research literature. Fourth, we also provide functions required for quantitative and qualitative evaluations (including visualization and sonification methods). Even though having an overlap to the previously mentioned software (e.g., librosa and linmdtw), the Sync Toolbox provides for the first time an open-source Python package for offline music synchronization that produces state-of-the-art alignment results regarding efficiency and accuracy. Thus, with the publicly available and well-documented Sync Toolbox, we hope to fill a gap between theory and practice for an important MIR task, while providing a useful pre-processing, annotation, and evaluation tool for data-driven machine learning.

Design Choices

When we designed the Sync Toolbox, we had different objectives in mind. First, we tried to keep a close connection to the research articles (Ewert et al., 2009) and (Prätzlich et

[1] http://www.eecs.qmul.ac.uk/~simond/match/
[2] https://github.com/ctralie/linmdtw

al., 2016). Second, we reimplemented and included all required components (e.g., feature extractors, DTW), even though such basic functionality is also covered by other packages such as librosa and linmdtw. This way, along with a specification of meaningful variable preset, the Sync Toolbox provides reference implementations for exactly reproducing previously published research results and experiments. Third, we followed many of the design principles suggested by librosa (McFee et al., 2015), which allows users to easily combine the different Python packages. The code of the Sync Toolbox along with an API documentation is hosted in a publicly available GitHub repository.[3] Finally, we included the synctoolbox package into the Python package index PyPi, which makes it possible to install synctoolbox with the standard Python package manager pip.[4]

Acknowledgements

The synctoolbox package builds on results, material, and insights that have been obtained in close collaboration with different people. We would like to express our gratitude to former and current students, collaborators, and colleagues who have influenced and supported us in creating this package, including Vlora Arifi-Müller, Michael Clausen, Sebastian Ewert, Christian Fremerey, and Frank Kurth. We also thank the German Research Foundation (DFG) for various research grants that allowed us for conducting fundamental research in music processing (in particular, MU 2686/7-2, DFG-MU 2686/14-1). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits IIS.

References

Arzt, A. (2016). *Flexible and robust music tracking* [PhD thesis]. Universität Linz.

Dixon, S., & Widmer, G. (2005). MATCH: A music alignment tool chest. *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 492–497. https://doi.org/10.5281/zenodo.1416951

Ewert, S., Müller, M., & Grosche, P. (2009). High resolution audio synchronization using chroma onset features. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 1869–1872. https://doi.org/10.1109/icassp.2009.4959972

McFee, B., Raffel, C., Liang, D., Ellis, D. P. W., McVicar, M., Battenberg, E., & Nieto, O. (2015). Librosa: Audio and music signal analysis in Python. *Proceedings the Python Science Conference*, 18–25. https://doi.org/10.25080/Majora-7b98e3ed-003

Müller, M. (2015). *Fundamentals of music processing* [Monograph]. Springer Verlag. https://doi.org/10.1007/978-3-319-21945-5

Müller, M., Mattes, H., & Kurth, F. (2006). An efficient multiscale approach to audio synchronization. *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 192–197.

Prätzlich, T., Driedger, J., & Müller, M. (2016). Memory-restricted multiscale dynamic time warping. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 569–573. https://doi.org/10.1109/ICASSP.2016.7471739

[3] https://github.com/meinardmueller/synctoolbox
[4] https://pypi.org/project/synctoolbox

Salvador, S., & Chan, P. (2004). FastDTW: Toward accurate dynamic time warping in linear time and space. *Proceedings of the KDD Workshop on Mining Temporal and Sequential Data*.

Tralie, C. J., & Dempsey, E. (2020). Exact, parallelizable dynamic time warping alignment with linear memory. *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 462–469. https://doi.org/10.5281/zenodo.4245470

Weiß, C., Schreiber, H., & Müller, M. (2020). Local key estimation in music recordings: A case study across songs, versions, and annotators. *IEEE/ACM Transactions on Audio, Speech & Language Processing*, *28*, 2919–2932. https://doi.org/10.1109/TASLP.2020.3030485

Weiß, C., Zalkow, F., Arifi-Müller, V., Müller, M., Koops, H. V., Volk, A., & Grohganz, H. G. (2021). Schubert Winterreise dataset: A multimodal scenario for music analysis. *ACM Journal on Computing and Cultural Heritage (JOCCH)*, *14*(2), 25:1–18. https://doi.org/10.1145/3429743

Zalkow, F., Weiß, C., Prätzlich, T., Arifi-Müller, V., & Müller, M. (2017). A multi-version approach for transferring measure annotations between music recordings. *Proceedings of the AES International Conference on Semantic Audio*, 148–155.