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

USING ACTIVATION FUNCTIONS FOR IMPROVING MEASURE-LEVEL AUDIO SYNCHRONIZATION

Yigitcan Özer, Matej Ištvánek, Vlora Arifi-Müller, Meinard Müller

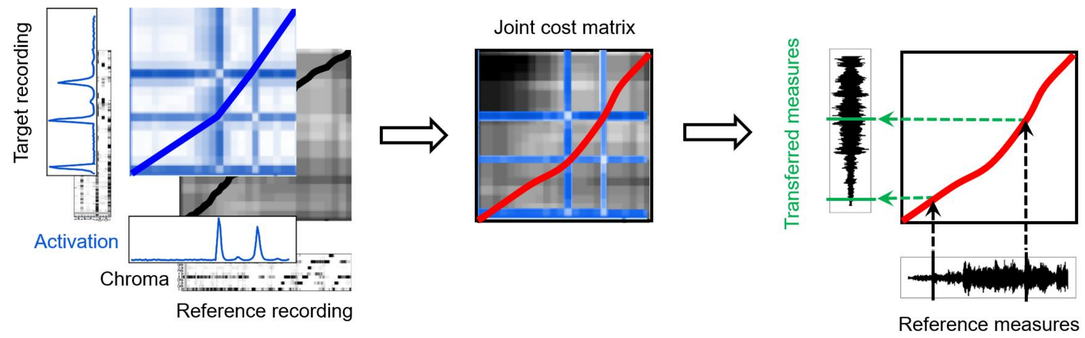
ISMIR 2022 BENGALURU

Abstract

Audio synchronization aims at aligning multiple recordings of the same piece of music. Traditional synchronization approaches are often based on dynamic time warping using chroma features as an input representation. Previous work has shown how one can integrate onset cues into this pipeline for improving the alignment's temporal accuracy. Furthermore, recent work based on deep neural networks has led to significant improvements for learning onset, beat, and downbeat activation functions. However, for music with soft onsets and abrupt tempo changes, these functions may be unreliable, leading to unstable results. As the main contribution of this work, we introduce a combined approach that integrates activation functions into the synchronization pipeline. We show that this approach improves the temporal accuracy thanks to the activation cues while inheriting the robustness of the traditional synchronization approach. Conducting experiments based on string quartet recordings, we evaluate our combined approach where we transfer measure annotations from a reference recording to a target recording.

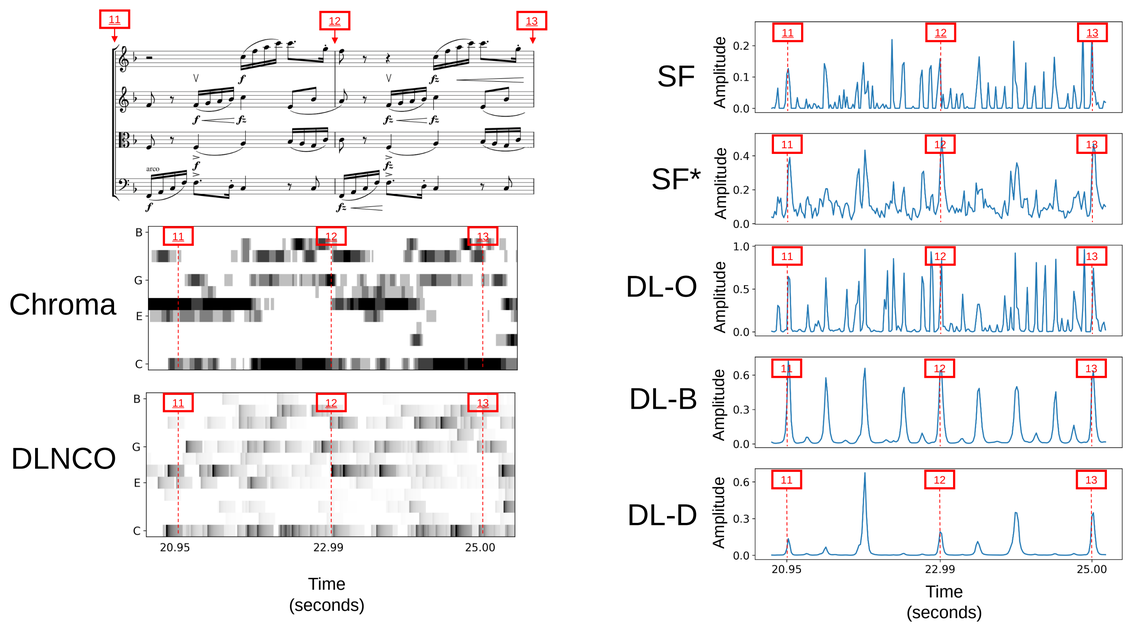
Introduction

- Task: Audio-to-audio synchronization for measure transfer
  - DTW with traditional chroma features for robustness
  - Integration of activation functions for a better temporal accuracy



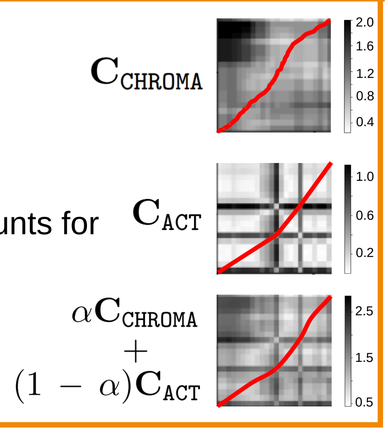
Conventional and DNN-Based Activation Functions

- Activation functions yield the likelihood of a musical event.
- In the ideal case, the value of the activation function is one when an event occurs and zero otherwise.
- DLNCO [1], Spectral Flux (SF) [2], and Superflux (SF*) [3] are conventional methods.
- CNN Onset Detector: DL-O [4]
- RNN Beat Detector: DL-B [5]
- RNN Downbeat Detector: DL-D [6]



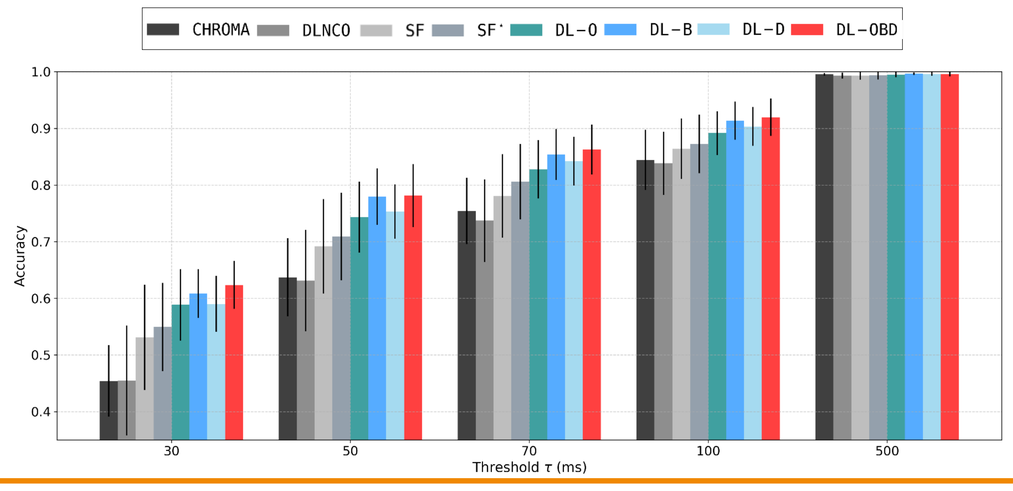
Combined Synchronization Approach

- Introduce three cost matrices:
  - $\mathbf{C}_{\text{CHROMA}}$ based on normalized chroma features,
  - $\mathbf{C}_{\text{ACT}}$ based on activation functions,
  - The sum $\alpha\mathbf{C}_{\text{CHROMA}} + (1-\alpha)\mathbf{C}_{\text{ACT}}$, which accounts for
    - Harmonic and melodic information,
    - Additional activation cues.



Dataset and Experiments

- String Quartet No.12 in F major, Op.96 by Antonín Dvořák
  - 4 movements, 3 different versions
- Comparison of eight different synchronization approaches
  - The avg. accuracy denotes the proportion of correctly transferred measure positions.
  - The integration of DL-OBD, which combines DL-O, DL-B, and DL-D, reveals the best accuracy.



References

[1] 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, 2009.

[2] 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.

[3] S. Böck and G. Widmer, “Maximum filter vibrato suppression for onset detections,” in Proceedings of the International Conference on Digital Audio Effects (DAFx), Maynooth, Ireland, 2013.

[4] J. Schlüter and S. Böck, “Improved musical onset detection with convolutional neural networks,” in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Florence, Italy, May 2014.

[5] S. Böck and M. Schedl, “Enhanced beat tracking with context-aware neural networks,” in Proceedings of the International Conference on Digital Audio Effects (DAFx), Paris, France, 2011.

[6] S. Böck, F. Krebs, and G. Widmer, “Joint beat and downbeat tracking with recurrent neural networks,” in Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), New York City, New York, USA, 2016.

Acknowledgements

This work was supported by the German Research Foundation (DFG MU 2686/10-2), and “Identification of the Czech origin of digital music recordings using machine learning” grant, which is realized within the project Quality Internal Grants of BUT (KInG BUT), Reg. No. CZ.02.2.69/0.0/0.0/19_073/0016948 and financed from the OP RDE. The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institut für Integrierte Schaltungen IIS.

Friedrich-Alexander-Universität Erlangen-Nürnberg

Fraunhofer IIS