

ASSOCIATING PLAYERS TO SOUND SOURCES IN MUSICAL PERFORMANCE VIDEOS

Bochen Li Zhiyao Duan Gaurav Sharma

Department of Electrical and Computer Engineering, University of Rochester, USA

bochenli@rochester.edu, zhiyao.duan@rochester.edu, gaurav.sharma@rochester.edu

ABSTRACT

Music is not just an art of sound. Both the audio and visual aspects of a musical performance, especially their association, are important for expressing players' ideas and engaging the audience. In this paper, we present a first attempt to separate sound sources and associate them to players detected in the video of a multi-instrument musical performance in a score-informed fashion. Our approach enables novel music enjoyment experiences that allow users to choose a player in the video and separate/enhance the corresponding sound source.

1. METHOD

Figure 1 illustrates our proposed approach, which operates in a score informed fashion. The MIDI score of the performance is available to guide both the separation and association processes. On the audio side, Dynamic Time Warping is applied first to find the optimal alignment between audio and score feature sequences. Then we adopt a state-of-the-art score-informed source separation method called Soundprism [1] to separate the sources, after which each score track is associated with a sound source.

On the video side, the natural correlation between the visual motion in performance and auditory note onsets makes it possible to associate the different players in the video to the separated sound sources. This correlation is especially prominent for string instruments, as many notes are started by a bow stroke, that is, many onsets of notes can be matched to onsets of bow strokes. So we extract the bowing motion of string instrument players using an optical flow estimation method [2], and then match bow stroke onsets with note onsets in score tracks to achieve source association. The Hungarian algorithm is then used to find the optimal association among all the bijections from video tracks to sound tracks.

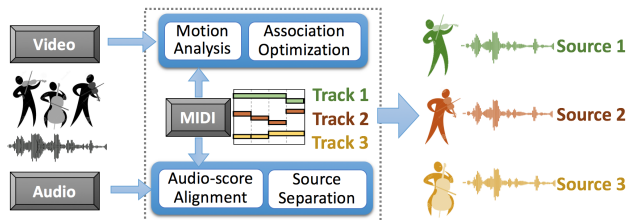


Figure 1. The proposed score-informed audio-video source separation and association system. Each separated audio source is associated with a performer from the video. The association is indicated by the color matching.



Figure 2. Example application of the proposed method. By clicking on a player, a user can isolate/enhance the corresponding sound track of a string quartet.

2. RESULTS AND DEMOS

Figure 2 conceptually demonstrates an application interface of our proposed system. It enables novel music enjoyment experiences that allow users to target an audio source by clicking on the player in the video and separate/enhance it. Formal evaluation on an audio-visual musical performance dataset where pieces were assembled from separately recorded but well coordinated individual tracks shows that the proposed system achieves good source separation and association results.

3. REFERENCES

- [1] Z. Duan and B. Pardo. Soundprism: An online system for score-informed source separation of music audio. *IEEE Journal of Sel. Topics in Sig. Process.*, 5(6):1205–1215, 2011.
- [2] D. Sun, S. Roth, and M. J. Black. Secrets of optical flow estimation and their principles. In *Proc. CVPR*, 2010.

