

Measuring synchronization between movement and music toward a multi-scale approach

Bayd Hamza, Guyot Patrice, Bardy Benoît, Slangen Pierre

EuroMov Digital Health in Motion, Univ. Montpellier, IMT Mines Ales, Ales, France

Abstract

Synchronization between musicians or dancers is typically based on the perception of a common rhythm. Humans also naturally tend to move to the perceived tempo of the music [1]. Recent researches has demonstrated that music is not only an acoustical experience but also engages the whole human body [2]. Due to various technological progresses such as musical rhythm estimation [3] and motion capture systems [4], the evaluation of synchronization performances between a large scope of movements and musical rhythms is today possible [5].

My doctoral thesis focuses on the analysis of synchronizations during movement and music. First, it aims to generate data from motion capture of individuals and groups, in our laboratory as well as in more natural settings. Second, it intends to identify or synthesize audio content with various rhythmic patterns. Third, it aims to analyze these data using different approaches from Artificial Intelligence, including recurrent neural networks and probabilistic graphical models. The final objective is to build architectures and applications that facilitate the human evaluation of synchronization as interactive, corporeal, and in real-time.

In my Ph.D, we consider that musical rhythms appears at different temporal scales (e.g. beats, downbeats, tatum), while the movements appear at different spatial scales or different part of body (e.g. fingers, harms, heartbeats, breathing) [6]. We call *multi-scale* these different temporal and spatial scales. Our objective is to reveal multi-scale synchronizations between movement and music, while analyzing the human body. To reach this goal, I propose with my supervisors to develop algorithms and methods based on computer vision and artificial intelligence (AI) [7, 8].

During the first 6 months of my doctoral thesis, a webcam beat-tracking software has been designed that measure and quantify in real time a synchronization score between natural movements of hands and the musical beats. This application is written in Python, using Mediapipe [9] and OpenCV for image processing and fingers tracking, as well as Librosa for the processing and playback of musical audio. The evaluation of synchronicity between the two time series, the motion time vector and the beat time vector, relies on Dynamic Time Warping (DTW). Hence, the application is able to compute a synchronization score from the similarity distance measured between these two time series [10].

In terms of rendering, the visual interface include two circles. Participants, standing in front of the computer screen, need to control the movements of their fingers by taping in these circles. The application gets timestamps when the fingers hits the circles. At the end of the test, a synchronization score is computed between the musical beats and the movement of the fingers, in percentage format. In future works, the application will calculate a multi-scale synchronization score between several types of movements and rhythmic parameters.

A demo video, available aside this abstract, shows a use-case of the current software application. The screen contains two blue buttons at each corner to choose the music. There are two options: with the left button we can chose audio music that has annotated timings for the respective beats; on the other side we can select the original version of audio music without beats annotations. Then there are two red buttons in the center of the screen to follow the beats with the fingers, and it can be done by moving the index finger in both directions on the specific red button, in order to synchronize with the rhythms of the audio music in the background. A live demo will be presented at the MOCO'22 conference.

Keywords: *Synchronization , Motion capture , Sound synthesis , Multi-scale rhythmic , Dance and music , Human-centered computing Human computer interaction (HCI) Interaction devices Sound-based input / output*

References

- [1] Damm, Loïc, et al. "Why do we move to the beat? A multi-scale approach, from physical principles to brain dynamics." *Neuroscience and Biobehavioral Reviews* 112 (2020): 553-584.
- [2] Zhuang, Wenlin, et al. "Music2dance: Dancenet for music-driven dance generation." arXiv preprint arXiv:2002.03761 (2020).
- [3] Fuentes, Magdalena. Multi-scale computational rhythm analysis: a framework for sections, downbeats, beats, and microtiming. Diss. Université Paris-Saclay (ComUE), 2019.
- [4] Cao, Zhe, et al. "Realtime multi-person 2d pose estimation using part affinity fields." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
- [5] Bellini, Rachele, Yanir Kleiman, and Daniel Cohen-Or. "Dance to the beat: Synchronizing motion to audio." *Computational Visual Media* 4.3 (2018): 197-208.
- [6] Tsuchida, Shuhei, Satoru Fukayama, and Masataka Goto. "Query-by-dancing: a dance music retrieval system based on body-motion similarity." *International Conference on Multimedia Modeling*. Springer, Cham, 2019.
- [7] Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).
- [8] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." *Advances in neural information processing systems* 27 (2014).
- [9] Zhang, Fan, et al. "Mediapipe hands: On-device real-time hand tracking." arXiv preprint arXiv:2006.10214 (2020).
- [10] Cheong, J. H. (2020, December 8). Four ways to quantify synchrony between time series data. <https://doi.org/10.17605/OSF.IO/BA3NY>