Multimodal Synchronization in Musical Ensembles: Investigating Audio and Visual Cues

Sutirtha Chakraborty
Supervisor: Dr. Joseph Timoney

Abstract

Multimodal interaction is a dynamic process and in certain situations where there is a repetitive aspect, it can be modeled as a set of coupled oscillators. This has applications in musical interfaces. A key metric of interactivity is synchronization where events occur with reference to each other. This synchronization is often not instantaneous among participants and different models of how convergence occurs exist. This study will investigate two particular mathematical models–Kuramoto and Swarmalator - used for musical phase synchronization amongst an ensemble, where each musician is represented as an individual oscillator communicating within a closely connected graph network. The research assesses the dynamic response of these models to tempo changes, initially derived from audio feedback alone but then subsequently incorporating extra information via visual cues, such as body sway. We employed the URMP dataset, featuring multimodal music performance data, for the experiments. Results reveal that Kuramoto's model outperforms the Swarmalator approach in predicting the synchronization behaviour of the ensemble for both the audio-only and audio-visual conditions, with the combined audio-visual approach yielding superior results. Additionally, we observed that larger ensembles provided more visual sway information, leading to a higher mean accuracy in Kuramoto's results for the audio-visual model. The findings underscore the significance of considering both auditory and visual cues when studying musical phase synchronization in ensembles and suggest that Kuramoto's model presents a promising avenue for modelling synchronous behaviour in more sophisticated group musical performances.