UNIVERSITY OF WESTMINSTER

IIT

INFORMATICS INSTITUTE OF TECHNOLOGY

in collaboration with

the **UNIVERSITY OF WESTMINSTER, UK**

BEng (Hons) DEGREE PROGRAMME in Software Engineering

A dissertation on

Audio-Visual Based Instrument Extraction in Music

in relation to the module

6COSC012C: Final Year Project

Researched by

Ratnajothy Sangeethanan – 2015084 – w1583030

Supervised by

Dr. Randil Pushpananda

April 2019

Abstract

Music is a form of artistic expression and plays a major role in the entertainment industry. When music is performed by a known group of singers and instrumentalists, human auditory system is capable of distinguishing vocals separately from background accompaniments. This task is effortless for human because auditory system organizes sound into perceptually meaningful elements but is quite challenging for machines. Although many music recordings are publicly accessible with multiple sound objects sharing a track, the necessity to manipulate any desired individual sound components as required will be beneficial to music society.

The proposed solution is based on audio-visual approach which consists of a cross-modal perception that involves interactions between two different sensory modalities. The system allows user to upload an already performed music which exists as a recorded video (unlabeled) and thereafter, its audio is used to predict particular instruments. Also, video annotation takes place to track respective coordinates of instruments in real time. If the camera focus is blurred, there can be a slight performance drop in identifying the right instruments as a time series in video. The prototype as a solution for the audio source separation is currently limited to two instruments which are violin and flute. This work can be also extended and enhanced in future for other classes of instruments.

Keywords: Audio-Visual System, Video Processing, Signal Processing, Convolutional Neural Network