INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

Cross-Lingual Singing Voice Conversion with VAW-GAN

A dissertation by

Mr. Kushan Bhareti

w1715748/2017271

Supervised by

Mr. Torin Wirasingha

Submitted in partial fulfillment of the requirements for the B.Eng. in Software Engineering degree at the University of Westminster.

July 2022

Abstract

Singing is not an ability that all humans have grasped at a level to perform in a professional manner. While most people involved in the entertainment industry are accustomed to singing, there can be instances where an actor does not meet the required criteria to fit into a role that requires singing. In such instances, a designated professional or a professional who has a similar voice is given the task to provide a replacement voice. While this can incur a higher cost in production, a number of dissimilarities will be encountered for which an additional effort will need to be given during post-production.

Singing Voice Conversion is a technology that uses a trained model to complete the conversion process to replace the singing voice of an individual with another, incurring minimal cost and effort. While research has been done on the domain over a number of years, a commercial product has not yet been published. Thus, a requirement exists to improve the existing technologies in relation to the domain. While both traditional and modern approaches have been taken towards achieving high-quality Singing Voice Conversion, it can be noted that the use of Generative Adversarial Networks has proved to have the best results, thus becoming the technological domain of the research.

Testing and evaluation for the project were done using subjective evaluation which required human participants to give feedback and objective evaluation which evaluated the quality of the outputs through pre-defined functions.

Keywords: Singing Voice Conversion, Voice Conversion, Deep Learning, Generative Adversarial Networks