ALTERATION OF LATENT REPRESENTATIONS IN AUDIO/SPEECH STYLE TRANSFER

by

Remizova Anna

A Thesis

Submitted in Partial Fulfillment of the

Academic Advisor:

Dr. Andrey Grabovoy

Bachelor of Science

Major: Applied mathematics and Computer Science

Moscow Institute of Physics and Technology

May 2024

Abstract

Motivation: Style transfer in audio is the field of computer science which is not yet thoroughly studied and, therefore, is of great interest. Moreover, it finds its applications in many practical tasks like dubbing movies, investigating voice abilities and music writing without any effort. In this paper we've been working on the task of voice/music style transfer. While there are many approaches to the task: Variational Auto Encoder(VAE), Gaussian Mixture Models(GMM), Linear Homodomain Transformations, it can be noticed that Image Style transfer is one of the most developed branches of Machine learning. That's why the image-like approach to the audio style transfer is altered in this paper.

Keywords: Machine Learning, Audio Style transfer, Speech Style transfer, Image Style Transfer, Deep Convolutional NN, Gram Matrix, Linear Transformation

Table of Contents

Chapter 1

Introduction

Audio style transfer is a technique to captivate style of one source of sound and content of another. This can be applied to music, for instance, a piano composition may sound in a different pitch or even different instrument while keeping the structure of the composition.