Robust Generative AI pipeline for voice generation.

D1.2 progress submission

Introduction.

The generation of voice is a highly complex problem that is deeply intertwined with the sequential processing of time series data. Voice generation, also known as speech synthesis, involves the creation of human-like speech from textual input. This field encompasses various subfields such as text-to-speech (TTS) synthesis and voice cloning, which have wide-ranging applications in industries like entertainment, telecommunications, and assistive technology.

Currently, the existing models for voice generation are plagued by their lack of modularity and excessive complexity, making them difficult to understand and extend. As a result, researchers and practitioners in the field are faced with the challenging task of investigating and exploring alternative approaches and solutions. In this research project, we aim to address this problem by thoroughly examining existing methodologies and developing a modular pipeline.

Current status.

The current stage of our work involves the comprehensive analysis of existing methodologies, wherein we have successfully examined and assessed various approaches that could potentially enhance the performance of our model. In order to expand our understanding and ensure a well-rounded perspective, we have specifically chosen to base our research on three meticulously selected scholarly articles. These articles, relevant to our research objectives, have been cataloged and made readily accessible via our dedicated repository on GitHub under the "Articles" folder.

The next crucial aspect in our work endeavor pertains to the structuring of our proposed model. Currently, we are devotedly engrossed in the reverse engineering process, unraveling the intricate design of the RVC model. Our meticulous approach involves extracting and compartmentalizing the model and its corresponding preprocessor into distinct and discrete blocks. However, due to the complexity inherent within the infrastructure of the RVC model, we have deemed it vital to allocate additional temporal resources towards meticulous disassembling. This process entails eliminating extraneous modules and expertly reconfiguring the overall architecture of the model to suit our specific research objectives.

Acknowledgements and Team.

Our research team comprises three members:

- Leon Parepko (l.parepko@innopolis.university) - Advanced Machine Learning, scientific research.
- Polina Lesak (p.lesak@innopolis.university) - Machine Learning engineer
- Darina Merzakreeva (d.merzakreeva@innopolis.university) - Machine Learning engineer

For additional information about our work, please visit: https://github.com/Leon-Parepko/Audio-Generation