

# Voice Cloning Project - Documentation

---

## Project Overview

Start Date: 13/09/2024

Current Date: 25/09/2024

Project Title: Voice Cloning with Fine-tuning and Lip-syncing Integration

Teams Involved: Voice Conversion (SVC, RVC), Lip-syncing (Wave2lip)

Objective: To develop a voice cloning model using fine-tuning techniques that will enable cloning of celebrity voices and dynamic integration of lip-syncing features.

## Project Phases

### Phase 1: Initial Research and Model Exploration

Timeframe: 13/09/2024 - 14/09/2024

In this initial phase, the team was tasked with conducting research on models suitable for fine-tuning voice cloning. The following models were explored:

- FastSpeech
- Tacotron 2
- YourTTS
- CoquiTTS
- TensorflowTTS
- StyleTTS
- Tortoise TTS

Outcome: Each team member presented their insights on the models. The focus was on identifying models that could be efficiently fine-tuned for voice cloning, emphasizing their strengths and limitations.

## **Phase 2: Model Selection and Team Division**

Date: 14/09/2024

After thorough discussions, the team was divided into two groups based on model selection for fine-tuning:

Team Allocations:

1. YourTTS Team: Badri, Soheli, Praful
2. FastSpeech Team: Sumit, Vidhan, Kaushik, Shivam

Proof of Discussion (14/09/2024):

A team meeting concluded with task allocation, where each group was assigned the responsibility of researching and planning fine-tuning strategies for their respective models.

## **Phase 2: Dataset Collection and Tool Identification**

Timeframe: 15/09/2024 - 19/09/2024

1. 15/09/2024: GitHub repositories were set up for both YourTTS and FastSpeech models to streamline version control.
2. 16/09/2024: The dataset collection process began. The focus was on gathering clean and varied voice samples for training.
3. 17/09/2024: Tools for Text-to-Speech (TTS) conversion were explored, as the project involves both voice-to-voice and text-to-voice functionalities.
4. 18/09/2024: Regular team meetings were held to monitor project progress.
5. 19/09/2024: So-VITS-SVC model was chosen for fine-tuning by Badri and Soheli.

Parallel efforts began with the RVC model by Sumit. Lip-syncing integration started by Vidhan and Kaushik.

## Phase 3: Model Fine-tuning and Integration

Timeframe: 20/09/2024 - 25/09/2024

1. 20/09/2024: Separate meetings were held for Team A and Team B.
2. 21/09/2024: Teams shared working files to ensure transparency and monitor progress.
3. 22/09/2024: A mild discussion about improvements in the project took place.
4. 25/09/2024: Team members presented their outputs for So-VITS-SVC and RVC models. The lip-sync team demonstrated the integration of their feature. Tasks were assigned for cloning multiple voices.

## Current Status and Next Steps

Fine-Tuning Progress:

- So-VITS-SVC Model: Badri and Soheli are working on fine-tuning with multiple voices (English and Hindi).
- RVC Model: Sumit has made significant progress with a multilingual voice conversion model using RVC.
- Lip-Syncing: Kaushik and Vidhan are working with Wave2Lip for lip synchronization with cloned voices.
- Shivam: Handling audio cleaning and managing dataset preparation for fine-tuning.

Highlights:

- Daily meetings scheduled at 4 PM to ensure alignment with project goals.
- GitHub repositories were created for version control management.

## **Conclusion**

The project is progressing well with daily meetings ensuring transparency and alignment on goals.

We successfully transitioned from model exploration to fine-tuning and integration of So-VITS-SVC and RVC for voice cloning.

The integration of lip-syncing with the voice cloning models will ensure a dynamic final product.

Review Date – 25/9/2024

Drafted By – BadriNarayanan S