**Experiment No.1**: Conduct feasibility study related to Voice Clone Text to Speech Software.

The objective of feasibility study is to determine whether proposed system is feasible. The feasibility is to determine in four aspects. These are:-

**TECHNICAL FEASIBILITY:** The project can be operated using the microphone which takes input to the system. The system operates on the recording and calculates the pitch, wavelength, and other aspects of the voice.

**FINANCIAL FEASIBILITY**: The project is based on microphone and system hardware and software which are affordable and basic needs of software development making it financially feasible to implement.

**OPERATIONAL FEASIBILITY:** This system makes person available with the recordings for his or her voice by just making use of microphone

**SCHEDULE FEASIBILITY:** Schedule feasibility is measure of how reasonable project time table is the system is found schedule feasible because the system is designed in such a way that it will finished the prescribed time.

**2.1 Need and Significance**

* They could enable new applications, such as transferring a voice across languages for more natural speech-to-speech translation, or generating realistic speech from text in low resource settings.
* Needed for people who have accidently lost their voice. It will restore their ability to communicate naturally.
* Will be a leading development for the other technologies to be build on Siri in users own voice, saving distinct languages through native speaker’s input, or saving the voice of loved ones.
* The goal of this work is to build a Text To Speech(TTS) system which can generate natural speech for a variety of speakers in a data efficient manner.
* Provide a secured and safe voice cloned speech.
* This system will help to build test to speech in a natural voice.
* Help the users to have the recording of their voice for number amount of data without actually recording the audio file.

**Experiment No.2**: Preparation of software requirements specification document related to Voice Clone Text to Speech Software.

1. **INTRODUCTION**

The goal of this work is to build a Text-to-Speech(TTS) system which can generate natural speech for a variety of speakers in a data efficient manner. We specifically address a zero-shot learning setting, where an un-transcribed reference audio from a target speaker is used to synthesize new speech in that speaker’s voice, without updating any model parameters.

Interestingly, speech naturalness is best rated with subjective metrics; and comparison with actual human speech leads to the conclusion that there might be such a thing as "speech more natural than human speech". In fact, some argue that the human naturalness threshold has already been crossed. Datasets of professionally recorded speech are a scarce resource.

* 1. PURPOSE: The main purpose of the project development is to clone human’s voice and have any number of lines read into the speech of that input voice.
  2. INTENDED AUDIENCE: The marketing staff, users, testers, developers, project managers are all the part of the audience. Anyone that belongs to any age will be the audience of the project.
  3. PROJECT SCOPE: The scope of the project is that it can be further used as generation of distinct voices of native speakers, can be used in robots to have natural voices of any respective human, can be modified to have security implementations along with voice cloning mechanism.
  4. REFERENCES: Baidu’s voice cloning AI, Lyrebird web application.

1. **OVERALL DESCIPTION**
   1. PROJECT FUNCTIONS: Sound will be recorded using microphone as the input of the project. The sound will be analysed in the form of wavelength, frequency, pitch of the speaker, locations to which the person belongs, native language of the person and many more independent and dependent variables will be notified. After this the dataset will be maintained and then training set will help the machine coding built to get trained about the data. The machine then finds correlations among the datasets. The machine learns the voice or speech of the person with respect to the various dependent and independent variables. Then the input taken is given to the system and that is processed according to the data given at the time of training a system. Output is generated according to the training given to the code of the project and the dataset provided.
   2. ASSUMPTIONS AND DEPENDENCIES: The project depends on the datasets for training the project and that dataset contains dependent and independent variables which according to our project are frequency of voice, pitch, wavelength, region of birth, accent of user and many more.
2. **EXTERNAL INTERFACE REQUIREMENTS**
   1. USER INTERFACE:

* Windows application based of GUI (Graphical user interface)
  1. SOFTWARE INTERFACE
* Operating system (Windows)
* python 3.7 or greater
* TensorFlow >= 1.1
* NumPy >= 1.11.1
* LibROSA == 0.5.1

3.3 HARDWARE INTERFACE

* Computer Storage
* Ram>=4GB
* i5 Processor

3.4 COMMUNICATION INTERFACE

* Microphone
* Speaker

1. **SYSTEM FEATURES**

The system generated the cloned voice of the human which the factors of naturalness and speaker’s similarity of speech.

4.1 SPEECH NATURALNESS: As human being have natural sense of speaking, the machines are also required to produce that sense to make the voice generated from the machine look normal. It make the voice more natural to the listener and this will help to have the other person feel that the voice produced is the natural voice of the human only whose input was given.

4.2 SPEAKER SIMILARITY: There may exist a similarity in two people voice whose accent are same. The similarity of the voices will be a part of the project as the similarity in the speech can help the development. As the systems will be trained with the huge amount of data that will help the code to result into an efficient and more accurate voice generation. More the dataset of one particular accent, place, region etc people will be provided the better will the system created.

1. **NON-FUCTIONAL REQUIREMENTS**
   1. PERFORMANCE REQUIREMENTS: The cloning of human’s voice should be fast and effectively generated. The time used in the learning the input must be less. The percentage of similarity of cloned voice and natural voice at least have 50-70% similarity.
   2. SAFETY AND SECURITY REQUIREMENTS: As anyone can use the cloned voice by just having a sample of that person’s voice. The safety can be provided by having encrypted value added with the voice file to get aware that the audio is security and is cloned. It helps to provide whether the audio is original or the cloned audio recording.

**Experiment No.3**: Preparation of software configuration management and risk management related to Voice Clone Text to Speech Software.

**SOFTWARE CONFIGURATION MANAGEMENT**

1. PLANNING: A configuration management plan details how you will record, track, control, and audit configuration. This document is often part of the project quality management plan.
2. IDENTIFICATION: All configuration requirements on a project should be identified and recorded. That includes functionality requirements, design requirements, and any other specifications. The completion of this process results in the configuration baseline for the project.
3. CONTROL: As the project scope is altered, the impact to the configuration must be assessed, approved, and documented. This is normally done within the project change control process.
4. STATUS ACCOUNTING: As the project scope is altered, the impact to the configuration must be assessed, approved, and documented. This is normally done within the project change control process.
5. AUDIT: This includes any tests to prove that the product conforms with the configuration requirements. Let’s say you built a report that must run within 10 seconds. The audit tests to see if the finished report actually runs that fast. Often, audits and checks will be built in at the completion of major project phases. This is so you can identify issues early.

**RISK MANAGEMENT**

1. IDENTIFY THE RISK: Firstly, we have to identify which risk are their in the development of the project. It is the most difficult as well as crucial part of our project as if the risks are not identified then it is not possible to resolve them or manage them properly. According to our project the main risk is that anyone can clone voice of anyone and perform any malicious tasks.
2. ANALYSE IT: Then we have to analyse the risk which includes ‘How the risk was generated?’, ‘What will be its harmful aspects?’, ‘Which parts of the system it affects the most?’, ‘What are the consequences of the risk?’ and ‘Which part of the society will suffer from the risk?’. As taking the AI voice clone system, it will affect the society in various factors as duplicating anyone’s voice intentionally and make it use as an evidence to any legal issue.
3. PRIORITIZE: Prioritising is the step that explains which risks are to be solved first and which will help the best in solving the system. Priority should be given to each risk so that we can check which risk should be immediately resolved and which will have no effect even if kept un-altered.
4. OWNERSHIP: Assigning the risk to the different team members so that they can solve the risk to have the best system without the problems in the particular part. This will help people to be aware about the part they are to concentrate and will resolve it in more effective and efficient way.
5. RESPOND: All that planning you’ve done is going to get implicated. First you need to know if this is a positive or negative risk. Is it something you could exploit for the betterment of the project?. For each major risk identified, you create a plan to mitigate it. You develop a strategy, some preventative or contingency plan. You then act on the risk by how you prioritized it. You have communications with the risk owner and, together, decide on which of the plans you created to implement to resolve the risk.
6. MONITOR: Whoever owns the risk will be responsible for tracking its progress towards resolution. But you will need to stay updated to have an accurate picture of the project’s overall progress to identify and monitor new risks.

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk Information Sheet** | | | |
| Risk ID : 01 | Date: 4 Mar, 2020 | Prob: 80% | Impact: Very High |
| **Description:**  The recording of any human being can be used as input and output can be used as an evidence to any un-spoken words of that human being. | | | |
| **Refinement/ Context:**  Sub condition 1: If call recording is done by culprit and used as input to the software.  Sub condition 2: Telephone cloning by any hacker.  Sub condition 3: Software takes recorded files as input. | | | |
| **Mitigation/ monitoring:**   1. Software should not accept the prior recording as the input to the software. 2. The input should be only accepted from the microphone. 3. The output recording of cloned voice should contain encrypted knowledge of the user who has given input. | | | |
| **Management/ Contingency plan/ Trigger:**  Time span for the project is 3 months. The input from microphone can be made only option at first but, it can be modified later on in last month. The user for taking the input must be verified from the Google account or any authorized source. | | | |
| **Current status:**  4 Mar, 2020: Mitigation Steps Initiated. | | | |
| **Originator:** Gursimar | | **Assigned:** Sandeep | |