UE21CS320A – Capstone Project Approval

Project Title : Generating Clear speech from Speech impaired Audio
Project ID : PW24_RS_03
Project Guide : Dr. Ramamoorthy Srinath
Project Team : Roseline Jerry A
T P Shriambhikesh
Tanmay Praveen Udupa
Vandana S

Outline

- Introduction
- Problem Statement
- Scope
- Feasibility study
- Applications/Use cases
- Expected Deliverables
- Capstone (Phase-I & Phase-II) Project Timeline

Introduction

- Speech disability accounted for 7.6% [1] of the total disabilities in India.
- According to National Institute on Deafness and Communication (NIDCD), in 2016 approximately 7.5 million people have language impairments.[2]
- There are more than 250,000,000 people with non standard speech needing accessible speech technologies. [3]
- Models are usually trained on specific disorders, not generic.

[1] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6730860/ & https://informahealthcare.com/doi/abs/10.1080/13682820110116776
[2] https://www.ijert.org/research/diagnosis-of-disordered-speech-using-automatic-speech-recognition-IJERTCONV8IS11029.pdf
[3] https://sites.research.google/relate/

Problem Statement

The project addresses the challenge of impaired speech communication, where various speech impairments hinder clarity and understanding.

Individuals with speech difficulties, whether due to medical conditions or other factors, encounter barriers in effective communication.

The project emphasizes the need for a solution that enhances different types of impaired speech outputs, ensuring clearer and more understandable communication.

Scope

The project aims to develop an application to generate “normal” speech from speech with impairments.

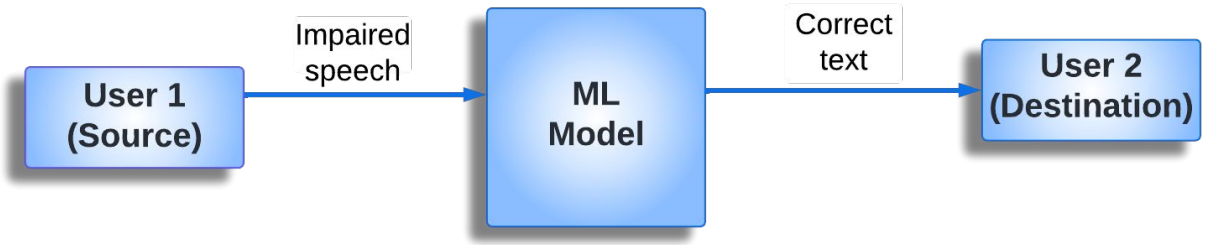
Key components of the scope:

- Develop and enhance a generic speech synthesis system for individuals with speech disorders.
- Provide clear and natural-sounding speech output in the user's original voice.
- Support real-time adaptation to changing speech patterns and contexts.
- Facilitate user-friendly interactions and customization options.
- Integrate feedback mechanisms to continuously improve the model's performance.
- Ensure adherence to ethical standards, privacy, and security considerations.
- Collaborate with speech therapists and healthcare professionals to align with therapeutic goals.

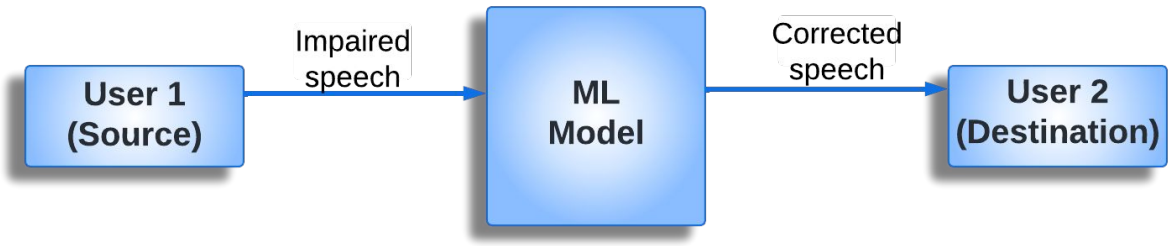
Feasibility study

- **Data collection:** Availability of a diverse dataset which includes conversations and their correct translation.
- **NLP Techniques:** these techniques can be used to preprocess, analyse, extract features from the dataset which is to be used in the model.
- **Sentiment Analysis:** Determining positive or negative response
- **Model Training:** Training a model to accurately predict the destination user's response based on the source user's message is feasible using supervised learning techniques.
- **Number of Speakers Involved:** Two/multiway impaired speech conversation

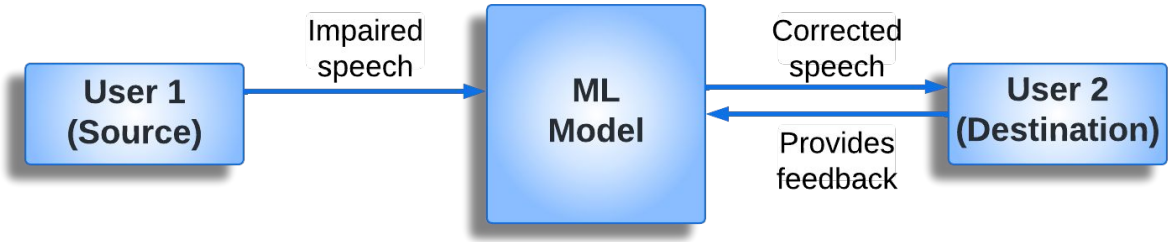
High level Scenario : 2 Speakers



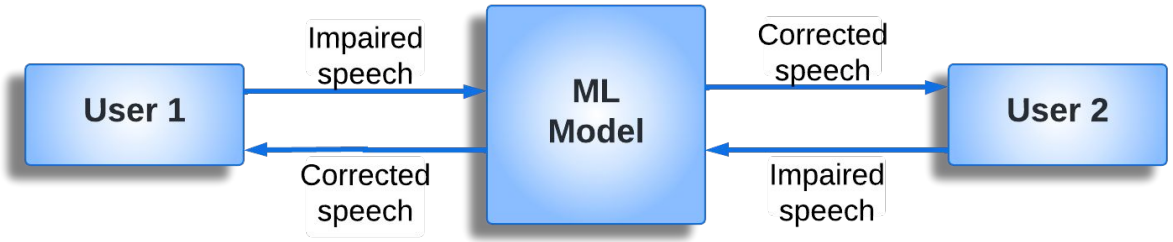
Communication between 2 users where one has impaired speech and the speech is converted to text



Communication between 2 users, where one has impaired speech so its converted to corrected speech

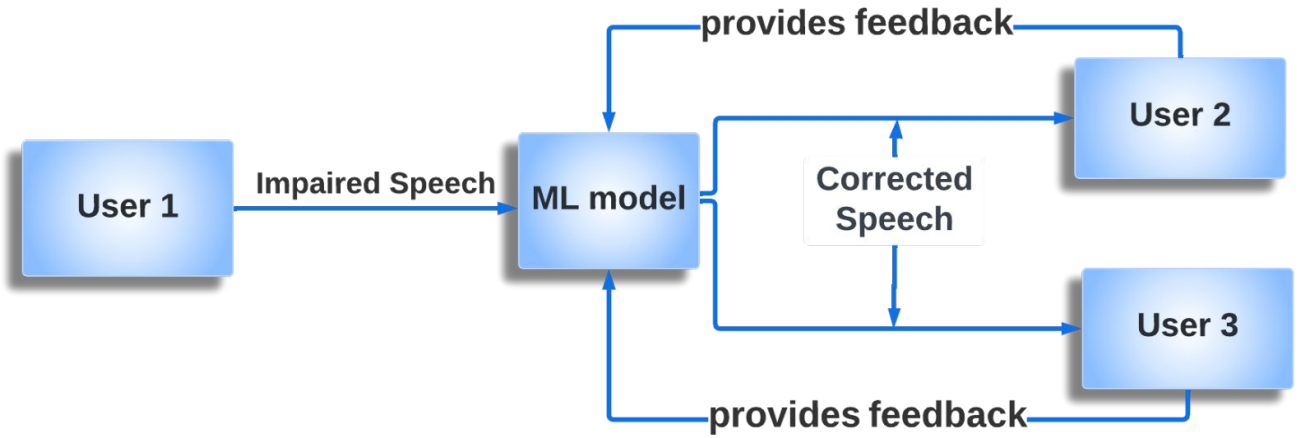


Impaired speech converted to corrected speech after which the User receiving corrected speech provides Feedback

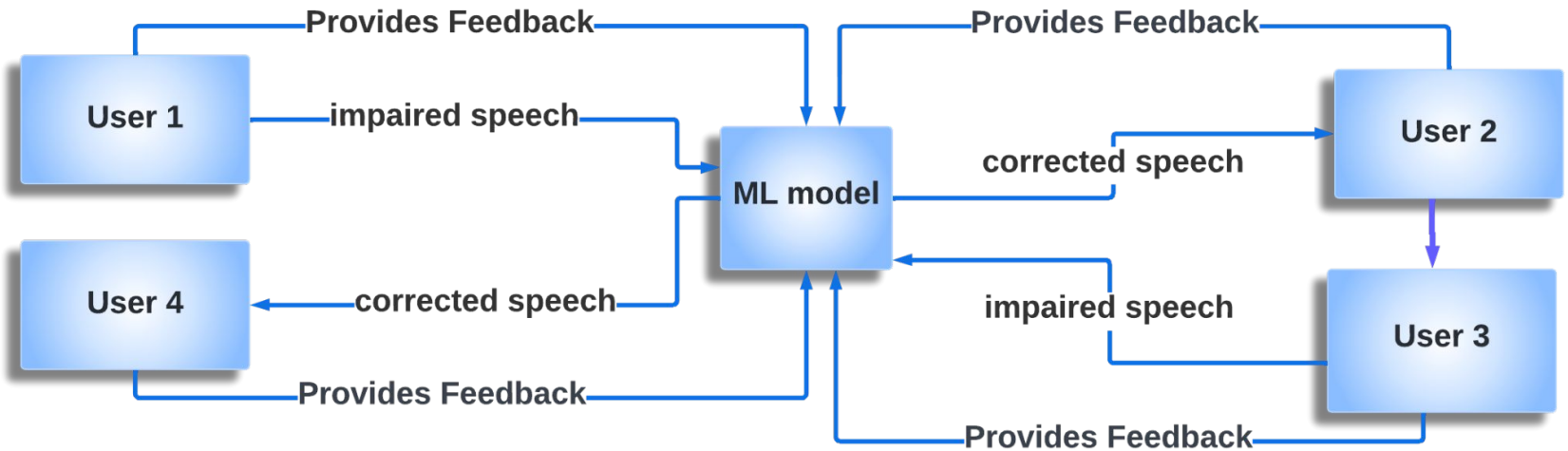


Communication between two impaired speakers

High level Scenario : Multiple-Speakers (Future Scope)



One impaired speaker communicating with 2 other normal speakers.



Multiple impaired speakers communicating with other speakers in a group setting. Example: Group meeting , Job panel etc

Challenges

- **Data Limitations:** Limited data can lead to underfitting or overfitting, which would affect the model's accuracy.
- **Ethical Considerations:** It should be ensured that privacy and consent in data collection is crucial. Ethical guidelines must be followed and the users must be informed about the purpose of their data.
- **Real world noise:** Preprocessing techniques such as noise reduction, can be used to improve the model's performance.
- **Subjective nature of clarity:** Clarity of speech often varies on user's perspective.

Applications/Use cases

- **Communication Aid:** People with speech impairments like stammering or old people who have difficulty in articulating speech can use this application to communicate with other people more effectively.
- **Education:** This application can be used to support students with speech impairments, so that they can participate actively in classroom discussions.
- **Employment Setting:** Provides support to help the speech impaired candidates effectively convey their thoughts to the interviewer.
- **Courtrooms:** People who are facing speech challenges can fully participate in legal proceedings.
- **Voice enabled technologies:** This application makes it easier for speech impaired people to use the voice enabled feature in applications.

Dataset Survey

- VCTK-Corpus | Kaggle
  Includes speech data uttered by 110 English speakers with various accents.

- EasyCall Corpus
  This dataset currently consists of 16683 audio recordings from 21 healthy and 26 dysarthric speakers.

- Mobile Device Voice Recordings at King's College London (MDVR-KCL)
  This dataset both contains recordings of both early and advanced Parkinson's disease patients and healthy controls.

- LSVT Voice Rehabilitation
  This dataset contains 126 samples from 14 subjects with Parkinson's Disease.

Expected Deliverables

- Capstone-I deliverables

- Project Plan
- Project Literature Survey
- Project Requirements Specification
- High Level Design Document of the Project
- Timeline and workflow outline
- Speech Enhancement Algorithm
- Cleaned and Pre-processed Dataset
- Preliminary Proof-of-Concept

Expected Deliverables

- Capstone-II deliverables

- Low-level Design Document
- Generative Model for Speech Enhancement
- Multi-agent Speech Correction Model
- Real-time Speech Processing Application
- Application with a User-Friendly Interface

Capstone (Phase-I & Phase-II) Project Timeline

Stage	Activity	Est. Duration	Start date	End date	Deliverables
Research Design and Planning	Finalize research statement	1 week	05/01/2024	21/01/2024	Research Proposal
	Feasibility and requirement analysis	1 week	12/01/2024	19/01/2024	Requirement Specification
Literature Review	Search and categorize relevant literature	4 weeks	05/01/2024	03/02/2024	Reference list
	Prepare draft literature review	6 weeks	05/02/2024	18/03/2024	Literature review and analysis
Data Collection	Finalize sampling plan	2 weeks	12/02/2024	26/02/2024	Data collection and sampling plan
	Carry out data collection	4 weeks	01/03/2024	28/03/2024	Raw data
	Prepare dataset	1 week	29/03/2024	05/04/2024	Draft dataset
Data Analysis	Clean and annotate data	1 week	06/04/2024	13/04/2024	Preprocessed dataset
	Analyze data behaviour	1 week	14/04/2024	21/04/2024	Data analysis report
Speech Processing Model	Decide finer details of the neural network architecture	2 weeks	22/04/2024	05/05/2024	Low level design document
	Create prototype of the algorithm	4 weeks	25/06/2024	25/07/2024	Proof of concept
	Fine tune the model	2 weeks	26/07/2024	10/08/2024	Generative speech enhancement model
	Expand to include multiple scenarios	4 weeks	11/08/2024	11/09/2024	Multiagent speech processing model
Paper writing	Draft the findings	2 weeks	12/09/2024	26/09/2024	First draft
	Seek feedback from Guide and iteratively improve	1 week	27/09/2024	10/10/2024	Final draft
	List conferences, journals and submission details	1 week	11/10/2024	20/10/2024	Submission Plan

References

- https://www.isca-archive.org/interspeech_2023/hermann23_interspeech.html
- Parrotron: An End-to-End Speech-to-Speech Conversion Model and its Applications to Hearing-Impaired Speech and Speech Separation
- Parrotron: New Research into Improving Verbal Communication for People with Speech Impairments
- Personalized ASR Models from a Large and Diverse Disordered Speech Dataset – Google Research Blog
- Detection of voice impairment for parkinson's disease using machine learning tools | IEEE Conference Publication
- How Artificial Intelligence (AI) Can Help Children with Speech, Hearing, and Language Disorders: 40 Free Resources - Columbia Engineering Boot Camps
- Machine learning assistive application for users with speech disorders - ScienceDirect
- [1807.10948] Articulatory Features for ASR of Pathological Speech
- Speech synthesis technologies for individuals with vocal disabilities: Voice banking and reconstruction

Thank
You