

CSE712

# Project Idea Presentation

*by*

Sajjad Hussain  
21166057  
Group-19

# Natural Speech Synthesis using Modular Neural Network

## Speech Generation

- ChatBots, ReadOutLoud
- Synthetic, Robotic
- Conversational Speech
  - Nuanced, Non-standard Phonetics
  - Context, Emotion

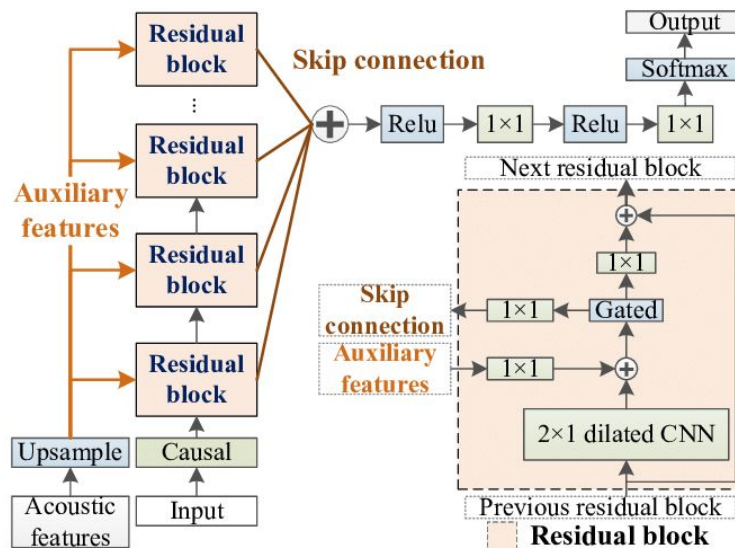
## Objective

- Casual Conversation
- Modular Neural Network
- Imitate Human Brain Architecture

# Background

WaveNet (van den Oord, 2016)

- Convolutional Neural Network
- Realistic Noise
- Content Swapping
- Modes of delivery
- Google Assistant



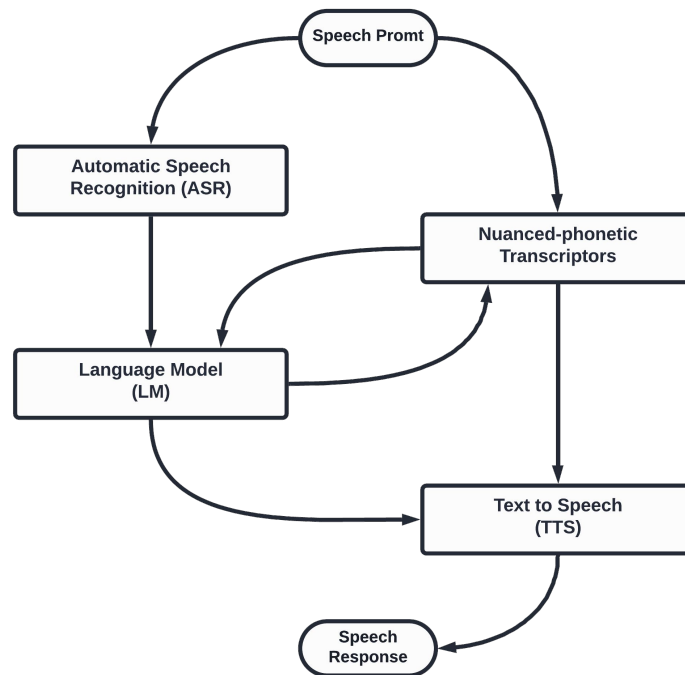
# Proposition – Modular Neural Network Framework

## Task Breakdown

- Prompt recognition
  - Automatic Speech Recognition (ASR)
- Expression or nuance analysis
  - Nuanced-phonetic Transcriptors
- Natural language processing
  - Language Models (LM)
- Response generation
  - Text To Speech (TTS)

## Key Features

- Analyze speaker's voice attributes
  - Modulation, speed, pitch, mood, etc.
- Inter-Model communication
  - Interfacing



# Challenges

## Modular NN

- Inter-Model Interfacing
- Modifications
- Complexity Reduction
- Nuanced Phonetic Transcriptors
  - Unsupervised Training
  - Variations

## Dataset

- Whole Conversations
- Nuance Annotation

# Conclusion

## Use-cases

- Speech Tutoring
- Digital PA
- Complaint Management
- Audio therapy for Alzheimer's

## Future Expansions

- AI Scaling
  - Augmentation
  - Aggregation
  - Synthetic Brain

## Summary

- Modular Neural Network
- Casual/Conversational Speech Generation
  - Nuance
  - Context

Thanks