Project Title

Devil's Tone 𝄞

Submitted by

Ayesha Islam Qadri **2020-CE-36**

Fatima Sohail Shaukat **2020-CE-37**

Ayesha Shafique **2020-CE-39**

Submitted to

Sir Raja Muzammil Munir

Course

CMPE-332L **Database Systems**

CMPE-341L **Artificial Intelligence**

Semester

5th

Date:

11th October 2022

Department of Computer of Engineering

University of Engineering and Technology, Lahore

Table of Contents

Problem Statement

A text-to-speech (TTS) system converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech. The reverse process is speech recognition. The text to speech synthesis is used to convert arbitrary input text into intelligible and natural-sounding speech. It is used in many applications such as vocal monitoring systems for blind people, web browsers, mobile phones, PCs, Laptops, and so on.

One of the biggest pitfalls of text-to-speech utilities is that people always say that they never sound like human beings. They just aren't natural, and often have a robotic tone to their voice that makes it difficult to listen to for a lengthy time. What if we could convert our lengthy boring notes into a beautiful melody? That much be better.

Abstract

Whether you realize it or not, AI makes the world a better place by allowing people to live more comfortably, but can AI also entertain people with text-to-speech singing? Maybe some people are familiar with ai text to speech or voice-over apps, as these technical features have been used in many video content like TikTok or YouTube Shorts. Just think about how good it can be to the extent that it can fool people, whether the singer is an AI or a human.

If you watch the famous worldwide audition program, "The Voice", the judges only focus on the voice and turn their chairs when they listen to the right voice that they'd want on their team. While watching the TV show, imagine putting an AI with a hologram instead of a human performer to see how the judges would react.

Objective

In this project, our main aim is to convert text into a song. Music is an effective stress reducer in both healthy individuals and people with health problems. Furthermore, songs are easier to memorize. One can learn the lyrics of a song by just hearing them twice or thrice. Wouldn't it be a revolutionary idea to turn the hectic syllabus into a mesmerizing song?

Scope

From a musician or songwriter who wants to hear their songs in someone else's voice, to a large company that wants to create a virtual influencer or celebrity, this technology can be used in many ways. With singing text to speech, a writer can have an AI sing wherever they are and see how their lyrics can be delivered.

Virtual influencers can finally have their unique voice and the irony of a singer-songwriter not having a voice would no longer occur. For example, since a singer-songwriter sounds like such a cool job and a free soul, AI models tend to be a singer-songwriter in many places.

It seems like the ability to sing is a big deal for both an AI and a human. And we know that a real singer-songwriter with his unique voice and virtual face would be a thing in a world where we interact with everyone online at the same time.

Description

This AI-based project will integrate text in such a way that it takes input in text form from the user and convert it to the soundtrack. We will use the singing voice synthesis technique in our project.

Singing Voice Synthesis

Singing voice synthesis (SVS) is a method of generating a singing voice from musical scores with lyrics using computer models. [1]

More about SVS

Singing synthesis has been developing since the 1950s and, like text-to-speech, revolves around two paradigms: statistical parametric synthesis, using statistical models to reproduce the features of a voice, and unit selection, when snippets of vocal recordings are recombined on the fly. Thanks to recent advances in technology, maestros can listen to a song immediately after composing it, with no recording necessary. In recent years, the following technologies have been used to achieve SVS:

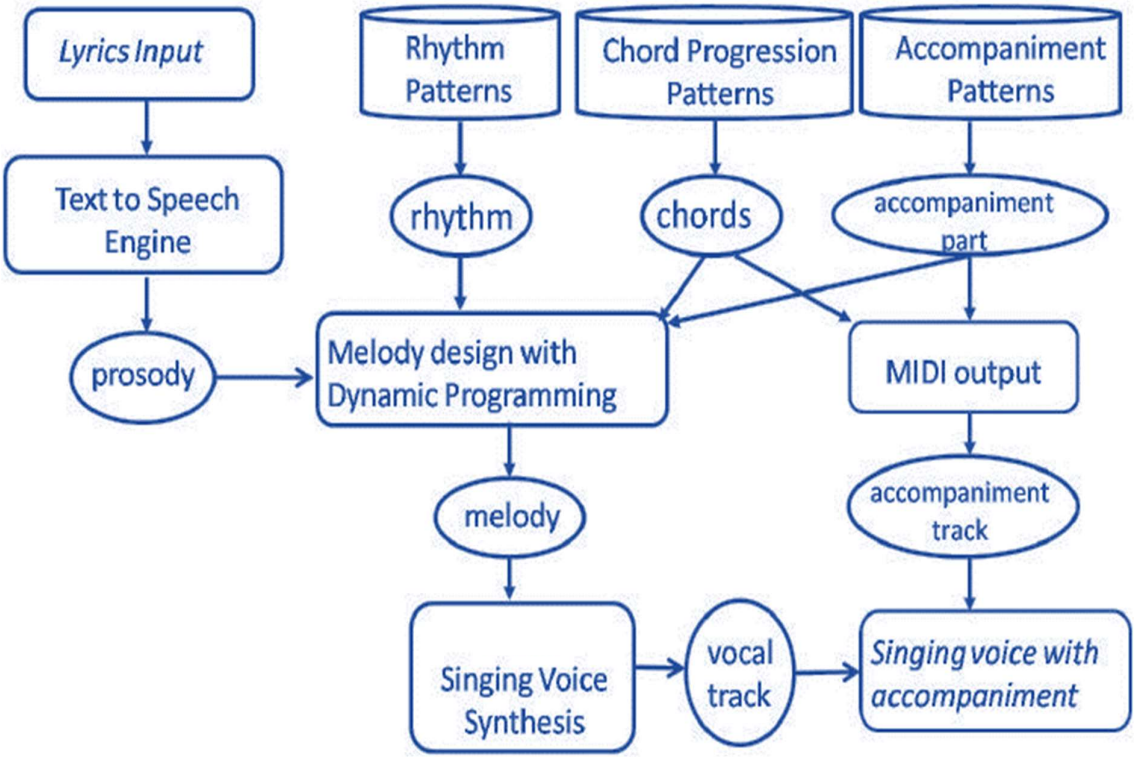
- Generic Deep Neural Networks (DNN)
- Convolutional Neural Networks (CNN)
- Recurrent Neural Network with Long-Short Term Memory (LSTM)
- Generative Adversarial Networks (GAN)

Applications

Singing voice synthesis technology allows musicians and singers to instantly know how their written music will sound. It's no longer necessary to go through the process of recording a piece of music and invest all the time, money, and resources that go into it. And no need to hire a team to assist with recording sessions.

Another critical use case is creating music for games and other projects that demand high degrees of audio support. Recording songs with real artists is extremely expensive for video game producers.

Block Diagram



Key Terms used in Block Diagram

- **Prosody**: The study of all the elements of language that contribute toward acoustic and rhythmic effects. [2]
- **Chords**: A chord, in music, is any harmonic set of pitches/frequencies consisting of multiple notes that are heard as if sounding simultaneous. [3]
- **Melody**: A pleasing series of musical notes that form the main part of a song or piece of music. [4]
- **Accompaniment**: Accompaniment is the musical part that provides rhythmic and/or harmonic support for the melody or main themes of a song or instrumental piece. [5]
- **MIDI**: MIDI, in full musical instrument digital interface, technology standard allowing electronic musical instruments to communicate with one another and with computers. [6]
- **Rhythm**: Rhythm may be defined as how one or more unaccented beats are grouped into an accented one. [7]

References

[1] https://typecast.ai/learn/is-text-to-speech-singing-the-future/

[2] https://www.britannica.com/art/prosody

[3] https://en.wikipedia.org/wiki/Chord_(music)

[4] https://www.britannica.com/dictionary/melody

[5] https://artsandculture.google.com/entity/accompaniment/m0257v7?hl=en

[6] https://www.britannica.com/art/MIDI-music-technology

[7] https://en.wikipedia.org/wiki/Rhythm