

TACOSI: A SINHALA TEXT TO SPEECH SYSTEM WITH NEURAL NETWORKS

THARUKA KASTHURIARACHCHIGE

A dissertation submitted in partial fulfilment of the requirement for Master of Science in Big
Data Analytics

**Department of Computing
Informatics Institute of Technology, Sri Lanka
in collaboration with
ROBERT GORDON UNIVERSITY, UK**

2022

Abstract

Text to Speech Synthesis (TTS) refers to a software program that converts text into speech and has a long history since the 12th century. This technology can be utilized for various purposes such as content reading, vehicle navigation, and announcement reading. More importantly, TTS can assist visually impaired people to use digital devices and managing their day-to-day life. TTS-related products such as speech-enabled websites, devices that assist speech-disabled individuals in communicating (AAC Devices), and digital talking books have already hit the market.

Sinhala is the first language of Sri Lanka and speaks by over 16 million people which are the major ethnic group in the country. Even though some language-independent TTS programs support Sinhala, they don't provide very strong performance. Few researchers have tried to develop Sinhala-dependent TTS systems using traditional methods, but they required time-consuming, manually generated features, and didn't produce natural-sounding speech. However, since 2016, deep learning has made significant advancements in TTS applications for various languages, including English and Chinese. This study's motivation came from the lack of Sinhala TTS tools with natural sounds. This study proposes TacoSi, a Sinhala language TTS tool based on the Tacotron algorithm. Based on 10 respondents' ratings for 10 voice recordings, the proposed technique achieved a 4.39 MOS. TacoSi's intelligibility has been tested by SUS sentences-based technique and was able to achieve 84% of intelligibility.