# Evaluation of Existing Text-to-Speech Systems for the Tamil Language

**Ahrane Mahaganapathy and Kengatharaiyer Sarveswaran**
**Department of Computer Science, University of Jaffna**
ahrane@univ.jfn.ac.lk

Deutscher Akademischer Austauschdienst
German Academic Exchange Service

## 1. Introduction

**Text-to-speech (TTS)** systems transform text to spoken language.

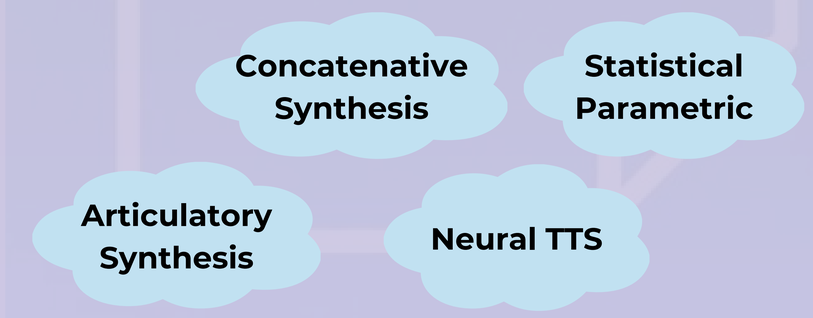
Text-to-speech conversion is not only modelling correct pronunciation as speech conveys elements such as expressiveness ( stress, intonation) and emotions.

There are more than 20 existing open-source ( e.g. Bhashini) and commercial ( e.g. ElevenLabs) Tamil TTS systems.

**Research Question:** How well do existing TTS systems synthesize expressiveness in Tamil speech?

## 2. Literature Review

### Speech Synthesis Techniques




### Evaluation Metrics

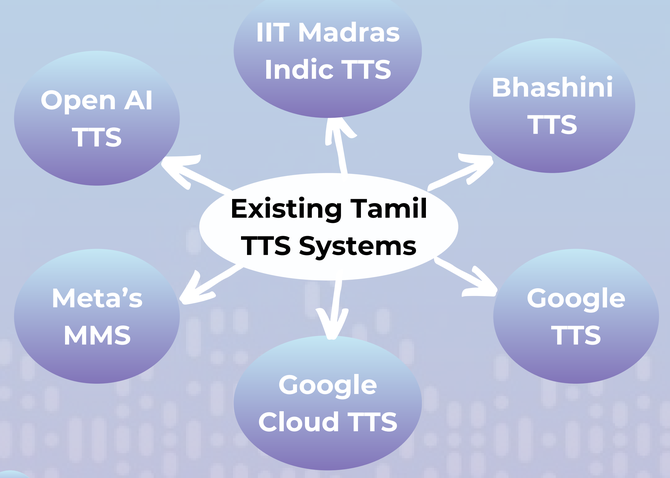
**Mean Opinion Score (MOS):** A subjective score where listeners rate the quality of speech on a scale from 1 to 5.

**Comparative Mean Opinion Score (CMOS):** A subjective score that allows listeners to compare two speech samples and rate them on a scale from -3 to +3.

**Word Error Rate (WER)**: (Yet to be done)
An objective score that measures the accuracy of synthesized speech by comparing it to a reference transcript.

## 3. Methodology for Evaluation

### Tamil TTS Systems evaluated



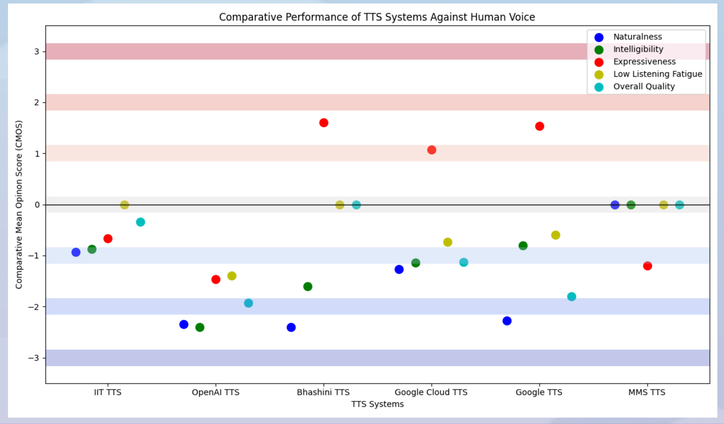

**Step 1** **Speech Data Collection**

- The human speeches were taken from published sources, including the book *Oru Yogiyin Suyasarithai* and audio narrations from *Ezhuna* Media. These speech data cover various domains, include loanwords, and show various expressions.
- TTS speeches were synthesized using the compiled corpus.

**Step 2** **Evaluation Methodology**

- The study involved 16 participants.
- Information such as basic demographic details, auditory-related issues and experience with TTS technologies was collected.
- The speech clips (both human and synthesized) were compared and evaluated two at a time by the participants.
- The speech clips were rated based on dimensions: naturalness, intelligibility, expressiveness, listening fatigue, and overall quality.
- A rating scale from -3 to +3 was used to express preference.
- The Comparative Mean Opinion Score (CMOS) was calculated by averaging the ratings for each pair across these dimensions.

## 4. Results




## 5. Challenges

- Collection of speech data across different domains, having loan words and conveying different expressions.
- Challenging for participants to understand and evaluate aspects such as stress intonation
- Objective method to evaluate expressiveness in speech.
- The results depends on the quallity of human speech evaluated.

## Selected References

1. Jalin, A. F., & Jayakumari, J. (2017). Text to speech synthesis system for Tamil using HMM. IEEE International Conference on Circuits and Systems (ICCS), pp. 447-451.
2. Arulprakash, A., Synthiya, M., Vijila, T., & Rajabhusanam, C. (2023). Tamil speech synthesizer app for Android: Text processing module enhancement. Indian Journal of Science and Technology, 16(7), 485-491.
3. ITU-T (1996). Methods for subjective determination of transmission quality. Recommendation P.800, International Telecommunication Union. Accessed: 2024-08-27.

## Acknowledgement

This research study was carried out under the DigSAL project, supported by the German Academic Exchange Service (DAAD) and funded by the Federal Ministry for Economic Cooperation and Development (BMZ) through SDG Partnerships.