Step-by-Step Process of Building Voices for Under Resourced Languages using MARY TTS Platform

Manuri Senarathna[1], Koliya Pulasinghe[1], and Shyam Reyal[1]

[1]Faculty of Computing Sri Lanka Institute of Information Technology, Sri Lanka.
manuri.s@sliit.lk, koliya.p@sliit.lk, shyam.r@sliit.lk

Abstract

This paper presents a comprehensive guide for creating synthetic voices to support under resourced languages for the MaryTTS platform. Although researchers have extensively contributed in the domain of speech synthesis, the lack of a thorough documentation hinders the voice building process for languages not yet supported by MaryTTS, complicating the implementation process for users with inadequate knowledge in the field of Text-to-Speech (TTS). The step-by-step process discussed in this study is further demonstrated with the creation of a synthetic voice for the Sinhala language, with unit selection as the voice building approach. A Sinhalese voice was generated with an intelligibility score of 91.7% upon evaluation with Diagnostic Rhyme Test (DRT). Comparison with ground truth data proved a close approximation to human speech where the intelligibility score was identified as 97.9%, when tested with the same participants. The Mean Opinion Score (MOS) revealed a naturalness level of 2.993, indicating a moderately high speech quality for the proposed system in comparison with the ideal score of 4.972.

Keywords

Text-To-Speech, Unit Selection, Low-Resourced Languages, Mean Opinion Score, Diagnostic Rhyme Test