

Bridging Linguistic Gaps: A Review of AI-Driven Speech-to-Speech Translation for Sinhala and Tamil in Sri Lanka

Imesha Dilshani^{1*}, Madusha Chandrasena²

¹*Department of Computer Systems Engineering, University of Kelaniya, Dalugama, Sri Lanka, dilshan-cs19063@stu.kln.ac.lk*

²*Department of Computer Systems Engineering, University of Kelaniya, Dalugama, Sri Lanka, madushac@kln.ac.lk*

Sri Lanka maintains two official languages which constitute Sinhalese and Tamil making it linguistically diverse. Almost all communication requires effective dialogue between Sinhalese-speaking and Tamil-speaking communities especially when operating through real-time speech-to-speech translation. The current version of Speech-to-Speech Translation (S2S) solutions serves a useful purpose yet faces three major limitations, including internet dependence, performance difficulties in loud environments, and unnatural Text-to-Speech (TTS) outputs. The Automatic Speech Recognition (ASR) systems from CallTran along with Android-based solutions through Google APIs and PocketSphinx, struggle with flexible operations when processing different accent varieties. Furthermore, the Machine Translation (MT) system performs poorly in achieving semantic relevance due to the scarcity of parallel corpora. The combination of ASR, MT and TTS systems produces performance delays and misinterpretation issues, which interfere with real-time functionality. This review examines current models, highlights theoretical and practical gaps, and proposes directions for future research, followed by a comparison of existing approaches. The research requires attention to three essential gaps, including bilingual dataset annotation tasks alongside offline functionality and natural voice synthesis development. We propose future research directions to establish massive bilingual datasets as well as implement noise-resistant ASR models using self-supervised approaches such as Whisper and Wave-to-Vector v2 (Wav2Vec2) and the fine-tuning of multilingual MT models like Multilingual Bidirectional and Auto-Regressive Transformer (mBART) for Low Resource Sinhala-Tamil Language translation systems. Additionally, TTS models like Tacotron, FastSpeech and Coqui TTS should be optimized for prosody and intonation.

Keywords: *automatic speech recognition, low resource speech-to-speech translation, machine translation, Sinhala-Tamil translation, text-to-speech*