

Pengembangan Sistem Penerjemah Suara *Real-Time* Dengan Integrasi Model ASR, NMT, Dan TTS Untuk Terjemahan Bahasa Indonesia Ke Bahasa Inggris. (*Development of a Real-Time Indonesian-to-English Speech Translation System with ASR, NMT, and TTS Integration*)

Supervised by Khen Dedes, S.Tr.Kom., M.T

Yuwandana
Study Program of Informatics Engineering
Majoring in Information Technology

ABSTRACT

This research aims to design and implement a real-time speech translation system based on an Android application by integrating three artificial intelligence models: Automatic Speech Recognition (ASR) using Whisper Small, Neural Machine Translation (NMT) using Opus-MT, and Text-to-Speech (TTS) using Piper TTS. The background of this research is Indonesia's low English proficiency ranking of 80th out of 113 countries according to the EF English Proficiency Index (EPI) 2024, which poses significant barriers in the education, business, and tourism sectors. The system was developed with a sequential ASR–NMT–TTS pipeline, where voice input is converted to text, translated into the target language, and then synthesized back into speech. The interface was built using the Flutter framework, while all models were fine-tuned using the Mozilla Common Voice dataset and the OPUS parallel corpus with an 80/20 train-test split. Evaluation results show that the fine-tuned Whisper Small ASR model achieved a transcription accuracy of 87.17% with a Word Error Rate (WER) of 12.83%. The Opus-MT NMT model obtained BLEU scores of 45.94 for Indonesian-to-English and 40.74 for English-to-Indonesian translation. The Piper TTS model achieved Mel Cepstral Distortion (MCD) values of 6.69 dB (Indonesian) and 5.78 dB (English). Overall, the integrated system was capable of executing the full translation pipeline with an average latency of approximately 3.2 seconds per sentence. Black Box Testing confirmed that all system features functioned as expected, while a User Acceptance Test (UAT) conducted with 30 respondents yielded average scores of 91% for design variables, 86% for system variables, and 88% for model system variables. The system has been proven effective in overcoming language barriers and opens opportunities for future development such as the addition of regional language support and latency reduction through streaming techniques.

Keywords: Translation, ASR, NMT, TTS, Automatic Speech Recognition, Neural Machine Translation, Text-to-Speech, Indonesian, English