IMPLEMENTASI SINTESIS UCAPAN BAHASA INDONESIA MENGGUNAKAN MODEL *TEXT-TO-SPEECH* TACOTRON 2 DAN HIFI-GAN

Nama Mahasiswa	: Angela Catherina
NIM	: 11211013
Dosen Pembimbing Utama	: Bima Prihasto, S.Si., M.Si., Ph.D.
Dosen Pembimbing Pendamping	: Boby Mugi Pratama, S.Si., M.Han.

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan model Text-to-Speech (TTS) berkualitas tinggi untuk Bahasa Indonesia dengan memanfaatkan arsitektur Tacotron 2 sebagai *mel-spectrogram synthesizer* dan HiFi-GAN sebagai *vocoder*. Dataset yang digunakan berupa audiobook berbahasa Indonesia yang dikumpulkan dan disusun oleh peneliti dalam format serupa dengan LJSpeech. Model Tacotron 2 dilatih untuk mengubah teks menjadi mel-spectrogram, sedangkan HiFi-GAN digunakan untuk menghasilkan sinyal audio dari mel-spectrogram tersebut. Pelatihan dilakukan menggunakan *toolkit open-source* SpeechBrain yang memungkinkan modifikasi arsitektur model, termasuk perubahan pada *attention mechanism*. Evaluasi kualitas suara dilakukan melalui metode *Mean Opinion Score* (MOS) dan *cross-similarity matrix*. Evaluasi performa model dilakukan melalui analisis *attention weight* dan nilai *loss function* model. Hasil penelitian menunjukkan bahwa model variasi dengan modifikasi *content-based attention* tanpa penggunaan *phonemizer* menghasilkan nilai MOS keseluruhan tertinggi sebesar 4,09. Sementara itu, model dengan modifikasi *content-based attention* dan penggunaan *phonemizer* mencatatkan nilai kemiripan *embedding* tertinggi terhadap suara asli (0,916) berdasarkan analisis *cross-similarity*. Temuan ini menunjukkan bahwa modifikasi arsitektur dapat meningkatkan performa model, serta bahwa pendekatan evaluasi subjektif dan objektif saling melengkapi dalam menilai kualitas sistem TTS. Penelitian ini membuktikan bahwa arsitektur Tacotron 2 dan HiFi-GAN dapat diimplementasikan secara efektif untuk sintesis suara Bahasa Indonesia dengan hasil yang kompetitif.

Kata kunci :
Sintesis Suara, *Text-to-Speech* (TTS), Bahasa Indonesia, Tacotron 2, *attention mechanism*