"PERANCANGAN SISTEM PENGENALAN UCAPAN MULTIBAHASA (ARAB & INDONESIA) BERBASIS OPENAI WHISPER & *VOICED-UNVOICED CLASSIFICATION*"

Nama Mahasiswa	: Richo Dwi Saputra
NIM	: 04191072
Dosen Pembimbing Utama	: Mifta Nur Farid, S.T., M.T.
Dosen Pembimbing Pendamping	: Himawan Wicaksono, S.ST., M.T.

ABSTRAK

Speech to Text merupakan perkembangan teknologi *ASR* yang memungkinkan komputer dapat menerima masukan berupa kata yang diucapkan (suara) dan mengubahnya menjadi teks yang dapat dibaca. Dalam perkembangannya, *Speech to Text* belum bisa menghasilkan teks bahasa secara simultan. Oleh karena itu, pada penelitian ini dilakukan perancangan sistem pengenalan ucapan multibahasa dengan menggunakan OpenAI Whisper dengan penambahan *initial prompt* dan *Voiced-Unvoiced Classification* yaitu *VAD*. Sehingga sistem dapat mentranskripsikan ucapan menghasilkan teks bahasa secara simultan yaitu bahasa Arab dan bahasa Indonesia. Selain itu juga, dilakukan pengukuran dan analisis terkait tingkat akurasi yang dihasilkan sistem yaitu *Word Error Rate* (*WER*) yang merupakan variabel terikat pada penelitian ini dan dengan dilakukan perubahan pada variabel bebasnya yaitu *Playback Speed*, *Signal to Noise Ratio* (*SNR*) dan Whisper Model. Setelah dilakukan simulasi, diperoleh hasil pengambilan data dari Dataset berisi 18 tabel yang memuat teks asli pada file audio dan teks hasil transkripsi yang digunakan sebagai data untuk uji akurasi dari sistem transkripsi. Adapun tingkat akurasi rata-rata pada sistem ini yaitu pada perubahan *playback speed* secara urut 75%, 100% dan 125% didapatkan nilai *WER* 67.29 %, 52.87 %, 54.95 % dengan model Small, 50.65 %, 26.81 %, 38.94 % dengan model Medium, 50.98 %, 39.93 %, 32.62 % dengan model Large-V1. Pada perubahan *SNR* secara urut yaitu default, 3 dB, 0 dB dan -3 dB didapatkan nilai *WER* 52.87 %, 72.14 %, 72.06 %, 79.37 % dengan model Small, 26.81 %, 64.29 %, 67.70 %, 71.35 % dengan model Medium, 39.93%, 69.30 %, 60.01 %, 67.93 dengan model Large-V1. Secara keseluruhan didapatkan nilai *WER* terendah yang berarti memiliki tingkat akurasi lebih akurat yaitu pada model Medium dengan perubahan *playback speed* 100% dan *SNR* default sebesar 26.81 %.

Kata Kunci : *Auto Speech Recognition* (*ASR*), OpenAI Whisper, *Speech to Text*, *Voiced-Unvoiced Classification*