

IMPLEMENTASI TEKS MENJADI UCAPAN MENGUNAKAN KORPUS BAHASA INDONESIA BERBASIS TRANSFER LEARNING SV2TTS

Nama Mahasiswa : Muhammad Gozy Al Vaiz
NIM : 11191051
Dosen Pembimbing Utama : Bima Prihasto, S.Si., M.Si., Ph.D.
Pembimbing Pendamping : Bowo Nugroho, S.Kom., M.Eng.

ABSTRAK

Pendekatan *transfer learning* digunakan untuk menghasilkan kualitas ucapan sintesis yang lebih alami dengan menggunakan model pembicara dari basis data *speaker verification*. Korpus bahasa Indonesia digunakan sebagai sumber data untuk pelatihan model. Metodologi penelitian mencakup pengumpulan korpus bahasa Indonesia, *pre-processing data*, pelatihan model *Generalized End-To-End Loss for Speaker Verification (GE2E)*, dan pelatihan model *synthesizer Tacotron2* yang diintegrasikan dengan model *encoder*. Selanjutnya pembuatan model *vocoder auto-regressive* berdasarkan *WaveRNN* yang diintegrasikan dengan model *synthesizer* untuk menghasilkan ucapan. Hasil penelitian menunjukkan bahwa model *encoder GE2E* memiliki *plot cluster* yang baik pada step ke-174500. Model *encoder* mampu mengelompokkan suara yang mirip serta dapat membedakan suara dengan *pitch* rendah dan tinggi. Hasil pelatihan model *synthesizer LSA* memiliki nilai *loss* sebesar 0.30740 pada *step* ke-50000. Hasil pelatihan model *synthesizer LSA* memiliki nilai *loss* sebesar 0.22048 pada *step* ke-50000. Hasil pelatihan model *vocoder* dengan *synthesizer LSA* memiliki nilai *loss* sebesar 4.0005 pada *epoch* ke-350. Hasil pelatihan model *vocoder* dengan *synthesizer forward attention* memiliki nilai *loss* sebesar 3.9152 pada *epoch* ke-350. Hasil evaluasi model secara objektif pada model *vocoder LSA* mendapatkan nilai 0.5333. Hasil evaluasi model secara objektif pada model *vocoder forward attention* mendapatkan nilai 0.6516. Hasil evaluasi model secara subjektif pada model *vocoder LSA* mendapatkan nilai 1.4285. Hasil evaluasi model secara subjektif pada model *vocoder forward attention* mendapatkan nilai 1.5714. Penelitian mengidentifikasi beberapa keterbatasan dalam hal sintesis suara dalam bahasa Indonesia. Jumlah *dataset* berkualitas mempengaruhi kualitas dari model *synthesizer*. Kualitas model *synthesizer* mempengaruhi hasil dari model *vocoder*.

Kata Kunci: *transfer learning, teks, sintesis, ucapan, bahasa, speaker verification*

INSTITUT TEKNOLOGI
KALIMANTAN