

SKRIPSI

IMPLEMENTASI KOMUNIKASI DUA ARAH PADA *SERVICE ROBOT* MENGGUNAKAN MODEL *WAV2VEC 2.0 & TACOTRON 2*



**Disusun Untuk Memenuhi Syarat mendapatkan Gelar Sarjana Teknik Pada
Jurusan Teknik Elektro Fakultas Teknik
Univeristas Sriwijaya**

OLEH:

M. APRIYADI TRI PUTRA

03041282126087

**JURUSAN TEKNIK ELEKTRO
FAKULTAS TEKNIK
UNIVERSITAS SRIWIJAYA**

2025

LEMBAR PENGESAHAN

**IMPLEMENTASI KOMUNIKASI DUA ARAH PADA *SERVICE ROBOT*
MENGUNAKAN MODEL *WAV2VEC 2.0 & TACOTRON 2***



SKRIPSI

Disusun Untuk Memenuhi Syarat Memperoleh Gelar Sarjana Teknik pada Jurusan
Teknik Elektro Fakultas Teknik
Universitas Sriwijaya

Oleh

M. APRIYADI TRI PUTRA

03041282126087

Palembang, 21 Juli 2025

Mengetahui,

Ketua Jurusan Teknik Elektro



Menyetujui,

Dosen Pembimbing

Ir. Hera Hikmarika, S.T., M.Eng.

NIP. 197812072002122002

HALAMAN PERNYATAAN DOSEN

Saya sebagai pembimbing menyatakan bahwa telah membaca dan menyetujui skripsi ini dan dalam pandangan saya skop dan kuantitas skripsi ini mencakupi sebagai mahasiswa sarjana strata satu (S1).

Tanda Tangan :  _____
Pembimbing Utama : Ir. Hera Hikmarika, S.T., M.Eng.
Tanggal : Juli 2025

**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMIS**

Sebagai civitas akademik Universitas Sriwijaya, Saya yang bertanda tangan di bawah ini:

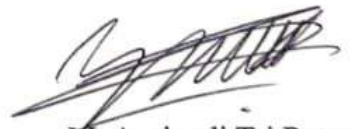
Nama : M. Apriyadi Tri Putra
NIM : 03041282126087
Jurusan : Teknik Elektro
Fakultas : Teknik
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Sriwijaya **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah Saya yang berjudul:

**IMPLEMENTASI KOMUNIKASI DUA ARAH PADA *SERVICE ROBOT*
MENGUNAKAN MODEL *WAV2VEC 2.0 & TACOTRON 2***

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini, Universitas Sriwijaya berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tulisan Saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini Saya buat dengan sebenarnya.

Dibuat di: Indralaya
Pada tanggal: 21 Juli 2025


M. Apriyadi Tri Putra
NIM. 03041282126087

HALAMAN PERNYATAAN INTEGRITAS

Yang bertanda tangan dibawah ini:

Nama : M. Apriyadi Tri Putra
NIM : 03041282126087
Fakultas : Teknik
Jurusan/Prodi : Teknik Elektro
Universitas : Universitas Sriwijaya

Hasil Pengecekan Software *iThenticate/Turnitin*: 2%

Menyatakan bahwa laporan hasil penelitian saya yang berjudul “Implementasi Komunikasi Dua Arah Pada *Service Robot* Menggunakan Model *Wav2vec 2.0 & Tacotron 2*” merupakan hasil karya sendiri dan benar keasliannya. Apabila ternyata dikemudian hari ditemukan unsur penjiplakan/plagiat dalam karya ilmiah ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tanpa paksaan.

Indralaya, 21 Juli 2025


M. Apriyadi Tri Putra

NIM. 03041282126087

KATA PENGANTAR

Puji syukur penulis ucapkan kepada Allah SWT. Atas Berkat dan dan Rahmat-Nya serta dukungan keluarga dan para sahabat penulis dapat menyelesaikan skripsi " Implementasi Komunikasi Dua Arah Pada Service Robot Menggunakan Model Wav2vec 2.0 & Tacotron 2".

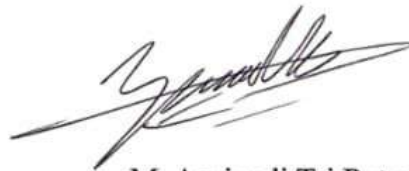
Penulisan skripsi ini disusun untuk memenuhi syarat mendapatkan gelar Sarjana Teknik di Jurusan Teknik Elektro Fakultas Teknik Universitas Sriwijaya. Dalam . Dalam kesempatan ini penulis mengucapkan terimakasih kepada:

1. Orangtua serta keluarga yang selalu memberikan semangat dan memberi dukungan baik secara mental, fisik, maupun finansial.
2. Ibu Ir. Hera Hikmarika, S.T., M.Eng. selaku pembimbing utama tugas akhir ini yang telah memberikan bimbingan, arahan, dan masukan serta memberikan ilmu selama proses penulisan skripsi.
3. Bapak Dr. Bhakti Yudho Suprpto, S.T., M.T., IPM. dan Ibu Dr. Eng. Suci Dwijayanti S.T., M.S., IPM. selaku pencetus dan memberikan bimbingan pada tugas akhir ini serta pengembang ide.
4. Bapak Muhammad Abu Bakar Sidik, S.T., M.Eng., Ph.D., IPU. Ketua Jurusan Teknik Elektro Universitas Sriwijaya dan Ibu Dr. Eng. Suci Dwijayanti S.T., M.S., IPM. selaku Sekretaris Jurusan Teknik Elektro Universitas Sriwijaya.
5. Bapak Dr. Bhakti Yudho Suprpto, S.T., M.T., IPM., Ibu Hera Hikmarika, S.T., M.Eng, Bapak Baginda Oloan Siregar, S.T., M.T, Bapak Irmawan, S.Si., M.T., dan Bapak Rendyansyah, S.Kom., M.T. yang telah memberikan kritik dan saran yang membangun dalam penelitian yang dilakukan agar menjadi lebih baik.
6. Segenap Dosen Jurusan Teknik Elektro Fakultas Teknik Universitas Sriwijaya yang telah memberikan ilmu selama perkuliahan.

7. Teman-teman seperjuangan *Service robot* Ari, Shalput, dan Silfani yang telah berkontribusi baik tenaga, waktu dan pikiran dalam menyelesaikan penelitian ini.
8. Seluruh teman-teman konsentrasi Teknik Kendali dan Robotika Teknik Elektro Universitas Siwijaya Angkatan 2021 selaku rekan kerja yang membantu dalam pengambilan data dan selalu bersemangat dalam pembuatan tugas akhir ini.
9. Teman-teman Klub Robotika UNSRI yang selalu membantu juga menyemangati selama proses penulisan skripsi ini berlangsung.
10. Seluruh pihak yang tidak mampu disebutkan satu-persatu yang telah membantu penulis dalam menyelesaikan tugas akhir hingga meraih gelar Sarjana Teknik.

Di dalam penyusunan skripsi ini, masih terdapat kekurangan karena keterbatasan penyusun, maka dari itu penyusun mengharapkan kritik dan saran yang membangun agar dapat menjadi evaluasi untuk penelitian yang lebih baik bagi penyusun dimasa yang akan datang. Akhir kata, penulis berharap agar skripsi ini menjadi lebih bermanfaat bagi pembaca terutama mahasiswa Jurusan Teknik Elektro Universitas Sriwijaya dan Masyarakat umum.

Palembang, 21 Juli 2025



M. Apriyadi Tri Putra

NIM. 03041282126087

ABSTRAK

Implementasi Komunikasi Dua Arah Pada *Service Robot* Menggunakan Model *Wav2vec 2.0* & *Tacotron 2*

(M. Apriyadi Tri Putra, 03041282126087, 2025, 82 halaman)

Abstrak—Service robot merupakan robot yang dapat melakukan bermacam kegiatan untuk membantu manusia. Dikarenakan service robot ditugaskan untuk membantu manusia, maka sangat diperlukan fitur komunikasi dua arah agar interaksi antara service robot dan manusia menjadi lebih efektif. Komunikasi pada service robot dapat terjadi dengan adanya sistem *speech to text* dan *text to speech* yang ditanamkan. Penelitian ini membahas pengembangan sistem komunikasi dua arah antara manusia dan humanoid robot berbasis suara dalam dua bahasa, yaitu bahasa Indonesia dan bahasa daerah Palembang. Sistem dibangun dengan mengintegrasikan algoritma *speech-to-text* (STT) menggunakan model *Wav2Vec 2.0* dan *text-to-speech* (TTS) menggunakan *Tacotron 2* yang dioptimalkan untuk pengenalan perintah secara real-time. Sistem STT terdiri dari dua model: model khusus yang dilatih menggunakan 25 kalimat perintah terbatas, dan model umum untuk menangani input bebas. Evaluasi dilakukan terhadap 10 partisipan dengan pengujian pada tiga jarak berbeda (30 cm, 60 cm, dan 90 cm) serta berbagai tingkat kebisingan. Hasil pengujian menunjukkan bahwa model STT khusus memiliki rata-rata *Word Error Rate* (WER) sebesar 2.10% untuk bahasa Indonesia dan 6.15% untuk bahasa Palembang, dengan akurasi tertinggi pada jarak 30–60 cm. Model umum menunjukkan WER lebih tinggi, yaitu 21.86% untuk bahasa Indonesia dan 25.76% untuk bahasa Palembang. Sementara itu, sistem TTS berbasis Tacotron 2 berhasil menghasilkan respon suara dengan rata-rata keberhasilan 81.66% untuk bahasa Indonesia dan 82.42% untuk bahasa Palembang. Hasil ini menunjukkan bahwa sistem komunikasi dua arah berbasis suara dalam bahasa lokal dapat diimplementasikan secara efektif pada *service robot* untuk mendukung interaksi yang lebih natural antara manusia dan mesin.

Kata kunci: Sistem komunikasi dua arah, *Speech to text*, *Text to speech*

ABSTRACT

Implementation of Two-Way Communication on a Service Robot Using Wav2Vec 2.0 & Tacotron 2 Models

(M. Apriyadi Tri Putra, 03041282126087, 2025, 82 pages)

Abstract— *Service robots are robots that can perform various tasks to assist humans. Since service robots are designed to help people, the presence of two-way communication features is essential to ensure effective interaction between the robot and humans. Communication on service robots can be realized through the integration of speech-to-text and text-to-speech systems. This study discusses the development of a voice-based two-way communication system between humans and humanoid robots in two languages: Indonesian and the Palembang regional language. The system is built by integrating the speech-to-text (STT) algorithm using the Wav2Vec 2.0 model and the text-to-speech (TTS) algorithm using Tacotron 2, optimized for real-time command recognition. The STT system consists of two models: a specialized model trained using 25 limited command sentences, and a general model designed to handle open-ended input. Evaluation was conducted on 10 participants with testing performed at three different distances (30 cm, 60 cm, and 90 cm) and under varying noise levels. The results show that the specialized STT model achieved an average Word Error Rate (WER) of 2.10% for Indonesian and 6.15% for the Palembang language, with the highest accuracy observed at a distance of 30–60 cm. The general model yielded higher WERs, at 21.86% for Indonesian and 25.76% for Palembang. Meanwhile, the Tacotron 2-based TTS system successfully generated voice responses with an average success rate of 81.66% for Indonesian and 82.42% for Palembang. These results demonstrate that a voice-based two-way communication system in local languages can be effectively implemented in service robots to support more natural interactions between humans and machines.*

Keywords: *Two-way communication system, Speech to text, Text to speech*

DAFTAR ISI

LEMBAR PENGESAHAN	ii
HALAMAN PERNYATAAN INTEGRITAS.....	iii
HALAMAN PERNYATAAN DOSEN.....	iv
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	v
KATA PENGANTAR.....	vi
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penulisan.....	4
1.4 Pembatasan Masalah	4
1.5 Keaslian Penelitian.....	5
BAB II TINJAUAN PUSTAKA	8
2.1 <i>State Of The Art</i>	8
2.2 Teori Pendukung	13
2.2.1 Humanoid Robot.....	13
2.2.2 <i>Natural language Processing</i>	14
2.2.3 <i>Speech to Text</i>	14
2.2.4 <i>Text to Speech</i>	15

2.2.5 Transformer	15
2.2.6 Wav2Vec 2.0	20
2.2.7 Tacotron 2	22
BAB III METODOLOGI PENELITIAN	25
3.1 Studi Literatur	26
3.2 Pengambilan Data	26
3.3 Perancangan Sistem Komunikasi <i>Humanoid Service Robot</i>	28
3.3.1 Desain <i>Humanoid Service Robot</i>	28
3.3.2 Perancangan sistem Hardware dan Software.....	29
3.3.2.1 Perancangan sistem hardware	29
3.3.2.2 Perancangan Sistem Software	32
3.3.3 Perancangan Algoritma Sistem Komunikasi Dua Arah pada <i>Humanoid Service Robot</i>	33
3.3.3.1 Algoritma Pelatihan Sistem Speech to Text Pada <i>Humanoid Service Robot</i>	33
3.3.3.2 Algoritma Pelatihan Sistem Text to Speech pada <i>Humanoid Service Robot</i>	35
3.3.3.3 Algoritma Komunikasi pada <i>Humanoid Service Robot</i> untuk Memberikan Respon.....	36
3.3.3.4 Algoritma Komunikasi pada <i>Humanoid Service Robot</i> untuk Menyapa..	37
3.4 Pengujian Sistem.....	38
BAB IV HASIL DAN PEMBAHASAN.....	40
4.1 Pengumpulan Data Suara	40
4.2 Pembuatan <i>Speech to Text</i>	41
4.2.1 Pra-pemrosesan data	41
4.2.1.1 Dataset Buatan Sendiri	42
4.2.1.2 Dataset common voice	43

4.2.2 Pelatihan Model Wav2Vec 2.0.....	44
4.2.2.1 Bahasa Indonesia.....	45
4.2.2.2 Bahasa daerah Palembang.....	48
4.3 Pembuatan <i>Text to Speech</i>	51
4.4 Hasil Desain Robot untuk Komunikasi Dua Arah	57
4.5 Pengujian <i>Service Robot</i> Untuk Komunikasi Dua Arah	59
4.5.1 Pengujian Bahasa Indonesia	59
4.5.2 Pengujian Bahasa Daerah Palembang	68
4.6 Pengujian Algoritma Menyapa	78
BAB V KESIMPULAN DAN SARAN	81
5.1 Kesimpulan	81
5.2 Saran	81
DAFTAR PUSTAKA.....	83
LAMPIRAN.....	87

DAFTAR GAMBAR

Gambar 2. 1 Sistem speech recognition dari penelitian Human-Robot Interaction pada Robot Bareleng 7 [16]	10
Gambar 2. 2 Proses pengubahan sinyal suara menjadi sintesis suara [18]	12
Gambar 2. 3 Contoh fisik <i>Humanoid Service Robot</i> [21]	14
Gambar 2. 4 Arsitektur dari <i>transformer</i> [25]	16
Gambar 2. 5 Arsitektur <i>Wav2Vec 2.0</i> [26]	20
Gambar 3. 1 <i>Flowchart</i> penelitian	25
Gambar 3. 2 <i>Flowchart</i> sistem komunikasi dua arah	26
Gambar 3. 3 Desain <i>humanoid service robot</i>	28
Gambar 3. 4 Perancangan peletakkan komponen komunikasi	29
Gambar 3. 5 <i>Microphone Rode Video Micro</i>	30
Gambar 3. 6 <i>Dot Matrix 8x32</i>	31
Gambar 3. 7 <i>Speaker audio</i>	31
Gambar 3. 8 Sensor suara KY-037	32
Gambar 3. 9 <i>Flowchart</i> pelatihan model <i>Wav2Vec 2.0</i>	34
Gambar 3. 10 <i>Flowchart</i> pelatihan model <i>Tacotron 2</i>	35
Gambar 3. 11 <i>Flowchart</i> algoritma respon <i>humanoid service robot</i>	36
Gambar 3. 12 <i>Flowchart</i> algoritma robot menyapa	38
Gambar 4. 1 (a) Sampel suara bahasa Indonesia (b) Sampel suara bahasa Palembang (c) Sampel Suara laki-laki dan perempuan dalam bahasa Palembang	41
Gambar 4. 2 Pembuatan file dataset dengan format .csv	42
Gambar 4. 3 Proses pengambilan dataset <i>common voice</i>	44
Gambar 4. 4 Proses penghapusan karakter unik pada dataset	44
Gambar 4. 5 (a) Grafik <i>train loss</i> dari pelatihan model bahasa Indonesia umum, (b) Grafik <i>learning rate</i> dari pelatihan model bahasa Indonesia umum	47
Gambar 4. 6 Grafik pelatihan model bahasa Indonesia khusus dengan <i>learning rate</i> 0,0001 (b) Grafik pelatihan model bahasa Indonesia khusus dengan <i>learning rate</i> 0,0003	48
Gambar 4. 8 (a) Grafik pelatihan model bahasa daerah Palembang khusus <i>learning rate</i> 0,0001, (b) Grafik pelatihan model bahasa daerah Palembang khusus <i>learning rate</i> 0,0003	51

Gambar 4. 9 Format txt untuk dataset Tacotron 2.....	52
Gambar 4. 10 Hasil Pelatihan Tacotron 2 pada epoch 5 dan batch size 6.....	53
Gambar 4. 11 Hasil Pelatihan Tacotron 2 pada epoch 5 dan batch size 16.....	54
Gambar 4. 12 (a) Kalimat respon dataset umum bahasa Indonesia, (b) Kalimat respon dataset umum bahasa daerah Palembang	57
Gambar 4. 13 Posisi komponen pendukung komunikasi dua arah	57
Gambar 4. 14 Tampilan LED Matrix pada saat merekam.....	58
Gambar 4. 15 Proses pengiriman kelas wajah dari pendeteksian	79
Gambar 4. 16 Program menerima data ROS.....	79
Gambar 4. 17 Program menambahkan nama pada respon.	80

DAFTAR TABEL

Tabel 2. 1 Analisis Hasil Self-Coding Neural Network	8
Tabel 2. 2 Hasil Pengujian Pada Ruangan Tenang	9
Tabel 2. 3 Hasil dari pengujian pada penelitian Human-Robot Interaction pada Robot Bareleng 7.....	11
Tabel 3. 1 Kalimat dataset bahasa Indonesia dan bahasa Palembang	27
Tabel 3. 2 Spesifikasi laptop Lenovo LOQ 82XT.....	30
Tabel 4. 1 Parameter <i>feature extractor Wav2Vec 2.0</i>	43
Tabel 4. 2 Spesifikasi <i>Google Colaboratory</i>	45
Tabel 4. 3 Parameter pelatihan model umum dan khusus Bahasa Indonesia.....	45
Tabel 4. 4 Parameter pelatihan model bahasa daerah Palembang umum dan khusus	49
Tabel 4. 5 Parameter pelatihan Tacotron 2	52
Tabel 4. 6 Kalimat respon dataset khusus bahasa Indonesia.....	54
Tabel 4. 7 Kalimat respon dataset khusus bahasa daerah Palembang.....	55
Tabel 4. 8 Beberapa data hasil pengujian sistem bahasa Indonesia pada jarak 30 cm.....	59
Tabel 4. 9 Evaluasi hasil pengujian bahasa Indonesia di jarak 30 cm	61
Tabel 4. 10 Beberapa data hasil pengujian sistem bahasa Indonesia pada jarak 60 cm.....	62
Tabel 4. 11 Evaluasi hasil pengujian bahasa Indonesia di jarak 60 cm.....	64
Tabel 4. 12 Beberapa data hasil pengujian sistem bahasa Indonesia pada jarak 90 cm.....	65
Tabel 4. 13 Evaluasi hasil pengujian bahasa Indonesia di jarak 90 cm	67
Tabel 4. 14 Beberapa data hasil pengujian sistem bahasa daerah Palembang pada jarak 30 cm.....	68
Tabel 4. 15 Evaluasi hasil pengujian bahasa Palembang di jarak 30 cm	70
Tabel 4. 16 Beberapa data hasil pengujian sistem bahasa daerah Palembang pada jarak 60 cm.....	71
Tabel 4. 17 Evaluasi hasil pengujian bahasa Palembang di jarak 60 cm	73
Tabel 4. 18 Beberapa data hasil pengujian sistem bahasa daerah Palembang pada jarak 90 cm.....	74

Tabel 4. 19 Evaluasi hasil pengujian bahasa Palembang di jarak 90 cm	76
Tabel 4. 20 Data durasi proses speech-to-text.....	77

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pertumbuhan pesat teknologi di masa Industri 4. 0 sudah bawa pergantian signifikan dalam bermacam zona, paling utama di dunia industri. Di antara teknologi yang tumbuh, *Artificial Intelligence* (AI) serta robotika menonjol selaku 2 inovasi yang sangat menguntungkan. Kedua teknologi ini tidak cuma tingkatkan efisiensi operasional, namun pula mengganti pola kerja serta mindset industri, dari skala besar sampai kecil[1],[2]. Peran robotika dapat dimanfaatkan untuk membantu pekerjaan manusia yang mana dapat berjalan secara otomatis dengan pengawasan manusia ataupun tanpa keterlibatan manusia[3].

Dalam beberapa tahun terakhir, *humanoid robot* telah berkembang pesat, menjadi semakin canggih dalam kemampuannya mengekspresikan gerakan dengan kepala, lengan, dan kaki mereka. *Humanoid robot* dianggap sangat cocok untuk pengembangan "robot komunikasi" yang akan berfungsi sebagai mitra yang mendukung manusia dalam kehidupan sehari-hari melalui kemampuan interaksinya yang canggih[4],[5]. *Service robot*, khususnya *humanoid robot*, telah menjadi inovasi penting dalam banyak bidang kehidupan sehari-hari, seperti perawatan kesehatan, layanan pelanggan, dan perhotelan. Penelitian menunjukkan bahwa robot ini dapat meningkatkan efisiensi dan memberikan pengalaman yang lebih baik bagi penggunanya, terutama untuk produk dengan fungsi yang beragam. Misalnya, *humanoid robot* lebih efektif dibandingkan karyawan manusia dalam menyampaikan informasi tentang produk dengan nilai praktis tinggi, seperti jam tangan pintar atau peralatan kesehatan. Keunggulan ini menjadikan robot servis sebagai solusi yang tepat untuk memenuhi kebutuhan modern di berbagai industri[6],[7].

Untuk mendukung efisiensi pelaksanaan tugas menggunakan robot service humanoid, pengembangan kemampuan komunikasi dua arah menjadi aspek yang krusial dalam menjembatani interaksi antara robot dan manusia secara efektif. Kemampuan ini memungkinkan robot untuk tidak hanya menerima dan memahami instruksi dari manusia, tetapi juga memberikan respons yang relevan dan akurat, sehingga menciptakan alur kerja yang lebih terstruktur dan efisien. Dengan komunikasi dua arah yang optimal, robot *humanoid service robot* mampu

menyesuaikan perilaku berdasarkan kebutuhan spesifik pengguna, meningkatkan sinergi dalam kolaborasi, serta memperbaiki kualitas *service* secara keseluruhan[8],[9].

Interaksi dengan *service robot* diawali melalui sistem pengenalan suara yang terintegrasi. Proses ini mencakup dua langkah utama: konversi ucapan menjadi teks (*speech-to-text*) dan konversi teks menjadi ucapan (*text-to-speech*). Pada langkah pertama, sistem memanfaatkan algoritma pemrosesan sinyal serta pembelajaran mesin untuk mentransformasi ucapan pengguna menjadi teks yang dapat diolah, sehingga meningkatkan tingkat akurasi dan efisiensi pengenalan suara[10]. Setelah perintah diubah menjadi teks, robot dapat memproses informasi dan memberikan respons yang tepat pada langkah kedua. Pada langkah ini, teknologi *text-to-speech* digunakan untuk mengubah teks menjadi ucapan yang dapat didengar pengguna[11]. Fitur-fitur ini memungkinkan *humanoid service robot* memiliki interaksi yang lebih alami dan responsif, meningkatkan pengalaman pengguna, dan memperkuat hubungan manusia-robot dalam berbagai konteks *service*[12].

Ada beberapa penelitian terdahulu yang telah membahas mengenai algoritma suara pada robot. Qin Tao dkk menerapkan algoritma jaringan syaraf tiruan pengkodean diri (*Self-Coding Neural Network*) dalam sistem pengenalan suara. Penelitian ini menekankan pentingnya peningkatan lebih lanjut mengenai ekstraksi fitur suara, terutama dalam kondisi yang bising. Namun algoritma ini tidak dapat berjalan dengan dataset yang kecil dan memerlukan daya komputasi yang tinggi karena kompleksitasnya, Serta terdapat penurunan akurasi bila terdapat banyak pembicara [13]. Berikutnya Ahmad Reinaldi Akbar melakukan penerapan algoritma *Deep Learning* untuk sistem *speech recognition* bahasa Indonesia pada *humanoid service robot*. Algoritma *Deep Learning* yang dipakai adalah *Long Short-Term Memory* (LSTM) dan menggunakan ekstraksi ciri *Mel Frequency Cepstral Coefficient* (MFCC) [14]. Hasilnya robot sudah mampu mengenali Bahasa pengguna di dalam ruangan dengan tingkat kebisingan 57,5-60,2 . Namun masih terdapat kelemahan dimana lokasi pengambilan datanya pada kondisi lingkungan random yang mengakibatkan menurunnya akurasi dataset. Selanjutnya, Vishnika Veni S dkk melakukan algoritma DNN-HMM dan *Lattice Free Maximum Mutual Information* (LF-MMI) yang memiliki kekurangan dalam menangani kesalahan

pengucapan yang umum, seperti menghilangkan vokal atau konsonan atau membedakan kata-kata yang mirip secara akustik. Untuk DNN, terutama dengan banyak lapisan tersembunyi, risiko *overfitting* menimbulkan tantangan tambahan, dataset yang diuji terbatas pada bahasa inggris dan sumber daya komputasi dalam jumlah besar [15]. Kemudian [16] penelitian ini membahas mengenai *open source toolkit python* VOSK yang diterapkan pada humanoid robot, hasilnya adalah metode ini masih belum mampu memahami kata-kata yang kompleks serta belum cukup adaptif jika terdapat variasi suara. Selain itu bahasa yang dapat diproses sangat terbatas. Maka dari algoritma ini memerlukan penelitian lebih lanjut. Selanjutnya Janardan Banjara dkk menerapkan algoritma *Convolutional Neural Network* (CNN) untuk sistem *speech recognition* khususnya bahasa Nepali India. Terdapat 5 contoh model yang dibandingkan diantaranya *Gated Recurrent Unit* (GRU), *Long Short Term Memory* (LSTM), *Convolutions*, *CTC Loss*, dan *Beam Search Decoding*. Hasilnya adalah mode (CNN-GRU) menjadi model yang paling baik dengan *Character Rate Error* (CER) 23.72%. Namun hasil akurasinya masih rendah, walaupun model sudah hasil dari kombinasi[17].

Lalu untuk penerapan algoritma *Text to Speech* terdapat penelitian yang membahas algortima *Deep Learning* based NLP oleh Eriss dkk [18] dimana dalam penelitian tersebut penggunaan model *deep learning* bisa meningkatkan kualitas sistesis suara. Namun terdapat kekurangan dimana algoritma ini memerlukan data pelatihan yang sangat besar dan berkualitas tinggi untuk mencapai hasil yang optimal. Selanjutnya Bhavyasri Kadali dkk menggunakan library python yaitu *pyttsx*, namun hasil sistesis suara yang dihasilkan masih terbatas pada aksen tertentu dan bahasa [19].

Jika melihat beberapa penelitian mengenai komunikasi 2 arah pada robot diatas masih terdapat keterbatasan pada algoritma yang dipakai, yaitu diperlukannya daya komputasi yang tinggi dan jumlah data yang besar serta berkualitas untuk mencapai akurasi yang tinggi [13][14][17] Serta beberapa penelitian tersebut menggunakan dataset bahasa inggris [15][16], untuk pengaplikasian bahasa indonesia pada penelitian *speech to text* masih sangat terbatas dalam hal dataset. Selain itu penelitian yang menggunakan bahasa daerah pada komunikasi 2 arah pada robot belum dilakukan, terutama pada bahasa daerah

palembang. Sehingga pada penelitian ini dikembangkan sistem *speech to text* secara *real-time* pada *service robot* menggunakan algoritma *Wav2Vec 2.0*. Algoritma *Wav2Vec 2.0* merupakan algoritma yang membutuhkan komputasi yang rendah sehingga dapat menghemat sumber daya serta dapat diaplikasikan pada dataset yang tergolong kecil [13]. Lalu untuk algoritma *text to speech* nya menggunakan algoritma *tacotron 2* menghasilkan sintesis suara yang tidak monoton dan lebih ekspresif [14]. Algoritma ini akan mampu membuat *humanoid service robot* dapat memahami bahasa seperti perintah dan meresponnya dalam bentuk informasi ataupun pekerjaan dengan suara yang lebih natural.

1.2 Rumusan Masalah

Berdasarkan latar belakang, terdapat beberapa permasalahan dalam pengembangan sistem komunikasi dua arah pada *humanoid service robot*, diantaranya kebutuhan daya komputasi tinggi, keterbatasan dataset bahasa Indonesia, serta rendahnya kualitas sintesis suara yang sering monoton. Selain itu penelitian dalam bahasa daerah terutama bahasa Palembang juga belum ada, padahal *service robot* perlu memahami bahasa daerah jika digunakan untuk komunikasi dua arah. Oleh karena itu diperlukan sistem yang lebih efisien dan mampu memproses perintah bahasa Indonesia secara *real-time* untuk meningkatkan interaksi manusia dan robot, serta sistem yang mampu menghasilkan suara yang lebih natural agar pengalaman komunikasi menjadi lebih optimal.

1.3 Tujuan Penulisan

Penelitian ini bertujuan untuk menerapkan sistem komunikasi dua arah antara manusia dan *humanoid service robot* menggunakan bahasa Indonesia dan bahasa daerah Palembang dengan sistem *speech to text* metode *wav2vec 2.0* serta *Tacotron 2* untuk sistem *text to speech*. Selanjutnya *performance* dari *wav2vec 2.0* dalam melakukan sistem *speech to text* secara *real-time* dengan menggunakan bahasa Indonesia dan bahasa Palembang serta Tingkat keberhasilan algoritma *Tacotron 2* dalam menghasilkan suara dan memberikan informasi akan di evaluasi pada penelitian ini.

1.4 Pembatasan Masalah

Pembatasan masalah pada penelitian ini meliputi :

1. Penerapan algoritma *Wav2Vec 2.0* digunakan untuk mengubah suara menjadi teks.
2. Penerapan algoritma *Tacotron 2* digunakan untuk mengubah teks menjadi suara.
3. Bahasa Indonesia yang digunakan adalah bahasa baku.
4. Bahasa Palembang yang digunakan adalah bahasa daerah.
5. Suara respon dari *Tacotron 2* adalah suara perempuan.
6. Pengambilan dataset di Lab Teknik Kendali & Robotika

1.5 Keaslian Penelitian

Penelitian ini merupakan lanjutan dari penelitian-penelitian yang telah dilakukan sebelumnya yang memiliki tema yang sama, namun dengan penerapan metode yang berbeda. Ada beberapa penelitian dengan tema yang sama. Qin Tao dkk [13] melakukan penerapan algoritma *Self-Coding Neural Network* untuk *speech recognition* yang berfokus pada pengidentifikasian identitas pembicara serta pemahaman semantik dari ucapan. Dengan memanfaatkan algoritma *neural network*, sistem ini mampu meningkatkan akurasi pengenalan suara, terutama dalam kondisi bising. Adapun dataset yang digunakan pada penelitian ini terdiri dari 28 kelompok data suara dari 8 pembicara berbeda. Hasilnya algoritma ini terbukti efektif dalam mendenoising suara dan pengenalan suara, namun masih terdapat kelemahan pada memerlukan komputasi yang sangat besar dan akan mengalami penurunan akurasi jika terdapat banyak pembicara.

Penelitian lain dilakukan oleh Ahmad Reinaldi Akbar dengan menggunakan algoritma deep learning yaitu algoritma *Deep Bidirectional Long Short-Term Memory* (LSTM) dan menggunakan ekstraksi ciri *Mel Frequency Cepstral Coefficient* (MFCC) untuk sistem *speech recognition* pada *humanoid service robot*. Hasil yang didapat dari 2 pengujian di dua kondisi ruangan yaitu bising dan tenang, akurasi menunjukkan bahwa nilai *Word Error Rate* (WER) pada ruangan tenang lebih tinggi dibandingkan dengan ruangan yang memiliki noise mesin dan kebisingan. Namun penelitian ini memiliki kelemahan dimana pengambilan data suaranya berada di kondisi lingkungan random yang mengakibatkan penurunan kualitas dataset [14].

Selanjutnya Vishnika Veni S dkk [15] melakukan algoritma DNN-HMM dan *Lattice Free Maximum Mutual Information* (LF-MMI) untuk mengembangkan sistem *speech recognition* untuk individu dengan gangguan bicara akibat penyakit *neurologis*. Dengan pengujian dari setiap pembicara yang mengucapkan 50 kata unik yang berulang sebanyak 5 kali, hal ini menunjukkan peningkatan akurasi pengenalan ucapan dibandingkan dengan model HMM tradisional, dengan peningkatan sebesar 1.01% untuk dataset 20 kata akustik mirip dan 11.2% untuk korpus ucapan terpengaruh 50 kata. Namun terdapat kekurangan dimana algoritma ini dengan banyak lapisan tersembunyi berpotensi akan mengalami *overfitting*, terutama dengan dataset yang kecil serta dataset masih terbatas pada bahasa Inggris.

Berikutnya Nurul Indah Mawaddah menerapkan *open source toolkit python* VOSK untuk sistem *speech recognition* pada robot tari *humanoid* sehingga robot bisa diperintah melalui suara untuk melakukan gerakan. Dari hasil pengujian terdapat 5 perintah yang diucapkan 100 kali dan hanya 8 kali terdapat kesalahan dan WER menunjukkan hasil 0.096 %. Dari hasil pengujian algoritma ini sudah sangat baik dalam melakukan pekerjaan namun terdapat kelemahan dimana dataset yang digunakan sangat terbatas, belum mampu untuk kalimat kompleks dan bahasa yang bisa dipakai masih terbatas[16].

Penelitian lain yang terkait adalah penelitian tentang *speech recognition* untuk Bahasa India khususnya Nepali [17]. Dataset yang dipakai pada penelitian ini berasal dari OpenSLR karena bahasa Nepali sendiri merupakan bahasa yang memiliki *low resource* untuk datasetnya. Dari perbandingan kelima model, model CNN-GRU merupakan yang terbaik dengan *Character Error Rate* (CER) 23.72%. Namun walaupun model sudah hasil dari kombinasi antara CNN-GRU hasil akurasi masih sangat rendah, untuk hasil dari model secara independen menunjukkan hasil yang lebih rendah lagi.

Penelitian selanjutnya mengenai *Text to Speech*, Eriss Eisa dan Babikir Adam menerapkan algoritma deep learning dalam proses sistesis *text to speech* yang berfokus untuk peningkatan pengenalan komunikasi melalui *Natural Language Processing* (NLP). Hasil yang didapatkan dari algoritma ini terbukti mampu mengurangi distorsi aperiodik dalam sintesis suara. Namun dalam dalam

pembuatannya diperlukan dataset yang sangat besar untuk lebih mengoptimalkan sintesis suara yang berkualitas [18].

Penelitian lainnya mengenai text to speech adalah penelitian yang dilakukan Koppala Guravaiah dkk menerapkan algoritma Pyttsx3 pada sistem *text to speech* untuk project "*Third Eye*" mereka. Projek ini diperuntukkan untuk orang dengan gangguan visual untuk bisa mengenali objek di dekatnya. Namun sintesis suara yang dihasilkan masih terbatas dalam aksen dan bahasa, serta suaranya masih monoton [19].

DAFTAR PUSTAKA

- [1] A. C. Yusufadz and A. Rosyidin, "Analisis Penerapan Artificial Intelligence Dan Robotik Pada Industri Manufaktur Indonesia Dalam Menghadapi Era Industri 4.0," *Pros. Semin. Nas. Teknol. ...*, vol. 1, no. 1, pp. 227–232, 2022.
- [2] P. Limbah, M. Jelantah, M. Lilin, A. Di, B. Sampah, and L. Winongo, "Jurnal Pengabdian Kepada Masyarakat: Teknologi dan Aplikasi Journal homepage :," *Pengabdi. Kpd. Masy. Teknol. dan Apl.*, vol. 2, no. 1, pp. 73–82, 2021.
- [3] M. Huda and E. Sri Pudjiarti, "Peran Otomatisasi dan Robotika dalam Era Digital: Trasformasi Bisnis Melalui Otomatisasi dan Robotika dalam Era Digital," *Transform. J. Econ. Bus. Manag.*, vol. 3, no. 1, pp. 254–272, 2024, doi: 10.56444/transformasi.v3i1.1739.
- [4] M. A. Maggioni and D. Rossignoli, "If it Looks Like a Human and Speaks Like a Human ... Communication and Cooperation in Strategic Human-Robot Interactions," *SSRN Electron. J.*, 2022, doi: 10.2139/ssrn.4190882.
- [5] Y. Tirupachuri, G. Nava, L. Rapetti, C. Latella, K. Darvish, and D. Pucci, "Recent Advances in Human-Robot Collaboration Towards Joint Action," *2020 IEEE-RAS 20th Int. Conf. Humanoid Robot.*, 2020, doi: 10.1109/Humanoids47582.2021.9555807.
- [6] H. Jiang, M. Xu, P. Sun, and J. Zhang, "Humanoid service robots versus human employee: how consumers react to functionally and culturally mixed products," *Int. J. Emerg. Mark.*, vol. 17, no. 4, pp. 987–1007, 2022, doi: 10.1108/IJOEM-04-2021-0643.
- [7] J. Miseikis *et al.*, "Lio-A Personal Robot Assistant for Human-Robot Interaction and Care Applications," *IEEE Robot. Autom. Lett.*, vol. 5, no. 4, pp. 5339–5346, 2020, doi: 10.1109/LRA.2020.3007462.
- [8] T. Sievers and R. Möller, "Get It Right: Improving Comprehensibility with Adaptable Speech Expression of a Humanoid Service Robot," *Commun. Comput. Inf. Sci.*, vol. 2047 CCIS, pp. 1–14, 2024, doi:

10.1007/978-3-031-55486-5_1.

- [9] N. Gasteiger, J. Y. Lim, M. Hellou, B. A. MacDonald, and H. S. Ahn, "A Scoping Review of the Literature On Prosodic Elements Related to Emotional Speech in Human-Robot Interaction," *Int. J. Soc. Robot.*, vol. 16, no. 4, pp. 659–670, 2024, doi: 10.1007/s12369-022-00913-x.
- [10] S. Alharbi *et al.*, "Automatic Speech Recognition: Systematic Literature Review," *IEEE Access*, vol. 9, pp. 131858–131876, 2021, doi: 10.1109/ACCESS.2021.3112535.
- [11] F. Khanam, F. A. Munmun, N. A. Ritu, A. K. Saha, and M. F. Mridha, "Text to Speech Synthesis: A Systematic Review, Deep Learning Based Architecture and Future Research Direction," *J. Adv. Inf. Technol.*, vol. 13, no. 5, pp. 398–412, 2022, doi: 10.12720/jait.13.5.398-412.
- [12] J. Berg and S. Lu, "Review of Interfaces for Industrial Human-Robot Interaction," *Curr. Robot. Reports*, vol. 1, no. 2, pp. 27–34, 2020, doi: 10.1007/s43154-020-00005-6.
- [13] Y. changmao and L. yuxuan Q. tao, D. hao, Y. yong, Z. yifei, "Speech recognition based on denoising self coding neural network," *IEEE Access*, no. 710, pp. 327–333, 2020, doi: 10.1109/ICBAIE49996.2020.00075.
- [14] S. Dwijayanti, A. Akbar, and B. Yudho, "Designing Human-Robot Communication in the Indonesian Language Using the Deep Bidirectional Long Short-Term Memory Algorithm," *J. Elektron. dan Telekomun.*, vol. 24, p. 1, Aug. 2024, doi: 10.55981/jet.595.
- [15] V. V. S and S. Chandrakala, "Investigation of DNN-HMM and Lattice Free Maximum Mutual Information Approaches for Impaired Speech Recognition," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3129847.
- [16] N. I. Mawaddah *et al.*, "Automatic Speech Recognition for Human-Robot Interaction on The Humanoid Robot Barelang 7," *EAI*, doi: 10.4108/eai.7-11-2023.2342940.

- [17] S. S. Janardan Banjara, Kaushal Raj Mishra, Jayshree Rathi, Karuna Karki, "Nepali Speech Recognition Using CNN and Sequence Models," *IEEE*, 2020, doi: 10.1109/ICMLANT50963.2020.9355707.
- [18] E. Eisa and B. Adam, "Deep Learning based NLP Techniques In Text to Speech Synthesis for Communication Recognition," *J. Soft Comput. Paradig. ER*, no. December 2020, 2023, doi: 10.36548/jscp.2020.4.002.
- [19] K. Guravaiah and Y. Sai, "ScienceDirect Third Eye : Object Recognition and Speech Generation for Visually Third Eye : Object Recognition and Speech Generation for Visually Impaired Impaired," *Procedia Comput. Sci.*, vol. 218, pp. 1144–1155, 2023, doi: 10.1016/j.procs.2023.01.093.
- [20] K. Darvish, L. Penco, J. Ramos, R. Cisneros, and R. O. Jan, "Teleoperation of Humanoid Robots : A Survey," *IEEE Trans. Robot.*, pp. 1706–1727, 2023, doi: 10.1109/TRO.2023.3236952.
- [21] T. Sievers and M. Ralf, "Connecting AI Technologies as Online Services to a Humanoid Service Robot," *IEEE*, 2023, doi: 10.1109/ICCAE56788.2023.10111181.
- [22] A. Garain, "Normalization of Numeronyms using NLP Techniques," *2020 IEEE Calcutta Conf.*, pp. 20–22, 2020, doi: 10.1109/CALCON49167.2020.9106524.
- [23] N. Sethiya and C. K. Maurya, "End-to-End Speech-to-Text Translation : A Survey," *Publ. Hist.*, vol. 90, 2023, doi: <https://doi.org/10.48550/arXiv.2312.01053>.
- [24] X. Tan, "NaturalSpeech : End-to-End Text to Speech Synthesis with Human-Level Quality," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 6, pp. 1–19, 2024, doi: 10.1109/TPAMI.2024.3356232.
- [25] I. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, "Attention Is All You Need," *Curran Assoc. Inc.*, no. Nips, 2017.
- [26] M. Baevski, A., Zhou, Y., Mohamed, A., & Auli, "Wav2vec 2.0 : A

Framework for Self-Supervised Learning of Speech Representations,” *NIPS*, no. Figure 1, pp. 1–19, 2020, doi: <https://doi.org/10.48550/arXiv.2006.11477>.

- [27] L. Rafael, S. Gris, E. Casanova, F. S. De Oliveira, and A. C. Junior, “Brazilian Portuguese Speech Recognition Using Wav2vec 2.0,” *Adv. Artif. Intell. – IBERAMIA 2022, Lect. Notes Comput. Sci. vol. 13572. Springer, Cham*, pp. 1–11, 2021, doi: 10.1007/978-3-030-98305-5_31.
- [28] A. R. Mandeel, M. Salah, and A. Raheem, “Speaker Adaptation Experiments with Limited Data Speaker Adaptation Experiments with Limited Data Speaker Adaptation Experiments with Limited Data for End-to-End Synthesis using for End-to-End Text-To-Speech Synthesis using for End-to-End Text-To-Speech T,” vol. XIV, no. 3, pp. 55–62, 2022, doi: 10.36244/ICJ.2022.3.7.