EVALUASI KINERJA *SPEECH-TO-TEXT* DENGAN REDUKSI KEBISINGAN *SPECTRAL GATING* DAN *WIENER FILTERING* PADA AUDIO BAHASA LAMPUNG

Tesis

Oleh

RAHMI PERMATA HATI
NPM 2327051006

PROGRAM STUDI S2 ILMU KOMPUTER
JURUSAN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
2025

EVALUASI KINERJA *SPEECH-TO-TEXT* DENGAN REDUKSI KEBISINGAN *SPECTRAL GATING* DAN *WIENER FILTERING* PADA AUDIO BAHASA LAMPUNG

Oleh

RAHMI PERMATA HATI

Tesis

Sebagai Salah Satu Syarat untuk Mendapat Gelar
MAGISTER KOMPUTER

Pada

Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam

PROGRAM STUDI S2 ILMU KOMPUTER
JURUSAN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
2025

ABSTRAK

EVALUASI KINERJA *SPEECH-TO-TEXT* DENGAN REDUKSI KEBISINGAN *SPECTRAL GATING* DAN *WIENER FILTERING* PADA AUDIO BAHASA LAMPUNG

Oleh

RAHMI PERMATA HATI

Pelestarian bahasa daerah seperti Bahasa Lampung menjadi semakin penting di tengah ancaman kepunahan budaya lokal. Penelitian ini bertujuan untuk mengevaluasi kinerja sistem *Speech-to-Text* (STT) terhadap audio berbahasa Lampung dialek Api melalui penerapan dua metode reduksi kebisingan, yaitu *Spectral Gating* dan *Wiener Filtering*. Sistem transkripsi utama dikembangkan menggunakan pendekatan *Hidden Markov Model* (HMM), dan untuk memperkuat evaluasi, penelitian ini turut melibatkan *Whisper*, model STT modern berbasis *deep learning* dari *OpenAI*. Dataset berupa rekaman cerita pendek dari penutur asli diproses melalui tahap reduksi kebisingan sebelum dilakukan transkripsi otomatis. Evaluasi kinerja dilakukan menggunakan tiga metrik utama: *Signal-to-Noise Ratio* (SNR), *Word Error Rate* (WER), dan *Character Error Rate* (CER). Hasil menunjukkan bahwa *Spectral Gating* memberikan peningkatan SNR tertinggi, dengan rata-rata di atas 21 dB, serta secara signifikan menurunkan nilai WER dan CER. Sementara itu, penggunaan *Whisper* pada data uji memperlihatkan peningkatan akurasi transkripsi, terutama pada audio yang telah melalui proses reduksi kebisingan. Penelitian ini menunjukkan bahwa integrasi teknik pengurangan kebisingan dengan sistem STT konvensional dan modern dapat meningkatkan kualitas transkripsi, serta mendukung pelestarian bahasa daerah melalui dokumentasi digital yang lebih akurat.

Kata Kunci : *Speech-to-Text*, Bahasa Lampung, *Hidden Markov Model*, *Whisper*, Reduksi Kebisingan, *Spectral Gating*, *Wiener Filtering*, *Word Error Rate*, *Character Error Rate*, Pelestarian Bahasa.

ABSTRACT

PERFORMANCE EVALUATION OF SPEECH-TO-TEXT WITH SPECTRAL GATING AND WIENER FILTERING NOISE REDUCTION ON LAMPUNG LANGUAGE AUDIO

By

RAHMI PERMATA HATI

The preservation of regional languages such as Lampung is increasingly important amid the threat of cultural extinction. This study aims to evaluate the performance of a Speech-to-Text (STT) system for Lampung language audio (Api dialect) by applying two noise reduction methods: Spectral Gating and Wiener Filtering. The main transcription system is developed using a Hidden Markov Model (HMM) approach, and to enhance the evaluation, the study also incorporates Whisper, a modern deep learning-based STT model from OpenAI. The dataset consists of short narrative recordings collected from native speakers, which were processed through noise reduction prior to transcription. System performance was evaluated using three key metrics: Signal-to-Noise Ratio (SNR), Word Error Rate (WER), and Character Error Rate (CER). Results indicate that Spectral Gating yielded the highest SNR improvement, averaging over 21 dB, and significantly reduced WER and CER. Additionally, the use of Whisper on test data showed improved transcription accuracy, particularly on audio that had undergone noise reduction. This study demonstrates that integrating noise reduction techniques with both conventional and modern STT systems can significantly enhance transcription quality, while also supporting the digital preservation of regional languages through more accurate documentation.

Keyword : Speech-to-Text, Lampung Language, Hidden Markov Model, Whisper, Noise Reduction, Spectral Gating, Wiener Filtering, Word Error Rate, Character Error Rate, Language Preservation.

Judul Tesis	:	**EVALUASI KINERJA *SPEECH-TO-TEXT* DENGAN REDUKSI KEBISINGAN *SPECTRAL GATING* DAN *WIENER FILTERING* PADA AUDIO BAHASA LAMPUNG**
Nama Mahasiswa	:	Rahmi Permata Hati
Nomor Pokok Mahasiswa	:	2327051006
Program Studi	:	S2 Ilmu Komputer
Jurusan	:	Ilmu Komputer
Fakultas	:	Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

Komisi Pembimbing

Dr. rer. nat. Akmal Junaidi, M.Sc.	**Dr. Aristoteles, S.Si., M.Si.**
NIP. 19710129 199702 1 001	NIP. 19810521 200604 1 002

Ketua Jurusan Ilmu Komputer	Ketua Program Studi
Dwi Sakethi, S.Si., M.Kom.	**Favorisen R. Lumbanraja, Ph.D.**
NIP. 19680611 199802 1 001	NIP. 19830110 200812 1 002

MENGESAHKAN

1. Tim Penguji

   Ketua : **Dr. rer. nat. Akmal Junaidi, M.Sc.** .....

   Sekretaris : **Dr. Aristoteles, S.Si., M.Si.** .....

   Penguji I

   Penguji Utama : **Favorisen R. Lumbanraja, Ph.D.** .....

   Penguji II

   Penguji : **Tristiyanto, S.Kom., M.I.S., Ph.D.** .....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Heri Satria, S.Si., M.Si.
NIP 19711001 200501 1 002

3. Direktur Program Pascasarjana

Prof. Dr. Ir. Murhadi, M.Si.
NIP 196403261 99802 1 001

Tanggal Lulus Ujian Tesis: **02 Juni 2025**

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Rahmi Permata Hati
NPM : 2327051006
Jurusan/Prodi : Ilmu Komputer/S2 Ilmu Komputer

Dengan ini menyatakan bahwa tesis saya yang berjudul **"Evaluasi Kinerja *Speech-To-Text* Dengan Reduksi Kebisingan *Spectral Gating* Dan *Wiener Filtering* Pada Audio Bahasa Lampung"** merupakan karya saya sendiri dan bukan karya orang lain. Semua tulisan yang tertuang dalam tesis ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila di kemudian hari terbukti tesis saya merupakan hasil penjiplakan atau dibuat orang lain, maka bersedia menerima sanksi berupa pencabutan gelar yang telah saya terima.

Bandar Lampung 02 Juni 2025

Rahmi Permata Hati
NPM 2327051006

RIWAYAT HIDUP

Penulis dilahirkan di Kotabumi, Lampung Utara, pada tanggal 13 November 1996. Penulis merupakan anak ketiga dari tiga bersaudara, buah hati dari pasangan Bapak Ramadan dan Ibu Fatmi Yulinda.

Penulis menyelesaikan Pendidikan Taman Kanak-Kanak di TK RA Tunas Harapan (DEPAG) Lampung Utara pada tahun 2002, Pendidikan Sekolah Dasar di SD Al-Kautsar Bandar Lampung pada tahun 2008, Pendidikan Sekolah Menengah Pertama di SMPN 23 Bandar Lampung pada tahun 2011, Pendidikan Sekolah Menengah Atas di SMAN 03 Bandar Lampung pada tahun 2014. Penulis memperoleh gelar Ahli Madya di Universitas Lampung pada tahun 2018 dan gelar Sarjana di Universitas Lampung pada tahun 2022. Pada Tahun 2023, penulis terdaftar sebagai mahasiswa Program Pascasarjana Jurusan Ilmu Komputer, Program Studi S2 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung, dan menyelesaikan pendidikan sebagai Magister Ilmu Komputer pada tahun 2025. Selama menempuh pendidikan, penulis berkesempatan menjadi pemakalah (*Oral Presenter*) pada konferensi *International Conference on Applied Science, Mathematics, and Informatics* (ICASMI) pada tahun 2024.

PERSEMBAHAN

Teriring rasa syukur dan cintaku kepada Sang Pencipta yang Maha Kasih,
Allah Subhanahu wa Ta'ala

Alhamdulillah, atas izin Allah Subhanahu wa Ta'ala, tesis ini dapat terselesaikan dengan segala berkah dan pertolongan-Nya. Setiap proses diberi kemudahan dan kekuatan hingga akhirnya sampai pada titik ini

Bismillahirohmannirrohim, Kupersembahkan karyaku ini kepada

Bapak dan Mamaku tercinta
Untuk segala pengorbanannya yang tiada henti memberi
cinta dan kasih sayang yang tak terhingga, kesabaran, perhatian serta iringan doa
yang tak pernah terputus disetiap perjalanan langkah anakmu ini

Uniku Helga Prima Amelia, Abangku Ayattullah Akbar Ramadhani, Abangku Badri Furqon, Tetehku Asmarandani Heryadi Putri, Keponakanku Abdullah Afkar Ramadhani, Adik sepupuku Ratu Inayah Khansa dan semua keluarga besar untuk segala doa, motivasi, dan bantuan yang telah menghantarkanku menyelesaikan pendidikan di jenjang magister

Seluruh keluarga besar Magister Ilmu Komputer 2023 dan orang-orang yang menyayangiku
Terima kasih untuk dukungan dan semangatnya,
Kalian adalah anugerah terindah yang pernah kumiliki

Almamater Tercinta, Jurusan Ilmu Komputer, Universitas Lampung

MOTO

مَافِي قَلْبِي غَيْرُ الله

Ma fi qalbi ghairullah

"Tidak ada di dalam hatiku selain Allah."

حَسْبُنَا اللهُ وَنِعْمَ الْوَكِيلُ

Hasbunallāhu wa ni'ma al-wakīl

"Cukuplah Allah (menjadi penolong) bagi kami, dan Dia adalah sebaik-baik pelindung"

(QS. Ali Imran: 173)

خَيْرُ النَّاسِ أَنْفَعُهُمْ لِلنَّاسِ

"Sebaik-baiknya manusia adalah yang paling bermanfaat bagi manusia lainnya"

(HR. ath-Thabrani)

"Ridha Allah tergantung pada ridha orang tua,

dan murka Allah tergantung pada murka orang tua"

(HR. Tirmidzi)

"Pelajarilah adab sebelum engkau mempelajari ilmu"

(Imam Malik bin Anas)

"Menyerah berarti kita tidak meyakini kekuasaan Allah"

"Allah mempunyai banyak jalan kepada orang yang mempunyai tujuan"

SANWACANA

Alhamdulillah, segala puji syukur penulis panjatkan ke hadirat Allah Subhanahu wa Ta‘ala atas segala limpahan rahmat, berkah, dan hidayah-Nya. Dengan izin dan pertolongan-Nya, serta petunjuk Rasulullah Nabi Muhammad Ṣallallāhu ‘Alaihi Wasallam, penulis dapat menyelesaikan tesis yang berjudul : "*Evaluasi Kinerja Speech-to-Text dengan Reduksi Kebisingan Spectral Gating dan Wiener Filtering pada Audio Bahasa Lampung*". Tesis ini disusun sebagai salah satu syarat untuk memperoleh gelar Magister Ilmu Komputer pada Program Studi S2 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

Penyusunan tesis ini tidak terlepas dari dukungan, doa, arahan, serta bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan rasa terima kasih dan penghargaan setinggi-tingginya kepada :

1. Kedua orang tua dan keluarga tercinta, atas cinta, doa, motivasi, serta dukungan yang tak terhingga dalam setiap langkah dan proses kehidupan penulis. Semoga Allah Subhanahu wa Ta‘ala membalas semua kebaikan dengan keberkahan dunia dan akhirat.
2. Bapak Dr. Eng. Heri Satria, S.Si., M.Si., selaku Dekan FMIPA Universitas Lampung.
3. Bapak Prof. Dr. Ir. Murhadi, M.Si., selaku Direktur Pascasarjana Universitas Lampung.
4. Bapak Dwi Sakethi, S.Si., M.Kom., selaku Ketua Jurusan Ilmu Komputer FMIPA Universitas Lampung.
5. Bapak Dr. rer. nat. Akmal Junaidi, M.Sc., selaku pembimbing utama, yang telah memberikan bimbingan, arahan, masukan, serta dukungan selama proses penyusunan tesis ini.
6. Bapak Dr. Aristoteles, S.Si., M.Si., selaku pembimbing kedua, atas kesediaan dan waktunya memberikan saran, koreksi, serta masukan yang membangun.

7. Bapak Favorisen R. Lumbanraja, Ph.D., selaku penguji utama dan Ketua Program Studi Magister Ilmu Komputer, atas masukan, evaluasi dan kontribusinya dalam penyempurnaan tesis ini.
8. Tristiyanto, S.Kom., M.I.S., Ph.D., selaku penguji kedua, atas saran dan masukan yang sangat bermanfaat dalam penyempurnaan penelitian ini.
9. Ibu Yunda Heningtyas, S.Kom., M.Kom., selaku Sekretaris Jurusan Ilmu Komputer FMIPA Universitas Lampung.
10. Bapak dan Ibu Dosen Jurusan Ilmu Komputer FMIPA Universitas Lampung, atas ilmu, wawasan, dan bimbingannya selama proses perkuliahan.
11. Seluruh staf administrasi dan karyawan FMIPA Jurusan Ilmu Komputer Universitas Lampung, atas bantuan dalam segala urusan administrasi selama proses perkuliahan.
12. Para penutur Bahasa Lampung, khususnya Bapak Maulana Marsad, S.Ag. (Gelar Paksi Tuan), Bapak Zainudin Hasan, S.Ag. (Gelar Sutan Raja Marga), Bapak Irwan, S.E., Ibu Devi Damayanti, S.Si., dan Ibu Dian Novitasari, S.Pd., yang telah berkontribusi langsung dalam penyediaan data penelitian.
13. Keluarga besar Magister Ilmu Komputer angkatan 2023, atas kebersamaan, dukungan, dan semangat yang telah menjadi bagian dari perjalanan akademik penulis.
14. Serta kepada anda yang membaca skripsi ini, semoga tulisan ini dapat berguna dan bermanfaat bagi anda dan yang lainnya.

Penulis menyadari bahwa tesis ini masih jauh dari kata sempurna. Semoga karya ini dapat memberi manfaat dan kontribusi dalam pengembangan teknologi pengolahan suara dan pelestarian Bahasa Lampung.

Bandar Lampung, 02 Juni 2025

Rahmi Permata Hati
NPM 2327051006

DAFTAR ISI

Halaman

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR KODE PROGRAM

I. PENDAHULUAN

1.1 Latar Belakang

Bahasa ibu merupakan salah satu bentuk ekspresi kultural utama suatu etnis atau daerah. Bahasa Lampung, sebagai bagian dari identitas kebudayaan masyarakat Lampung, memiliki peran penting dalam menjaga keberagaman budaya Indonesia. Namun, di tengah globalisasi dan perkembangan teknologi yang pesat, penggunaan bahasa Lampung dalam kehidupan sehari-hari terus mengalami penurunan. Generasi muda cenderung lebih fasih menggunakan bahasa Indonesia atau bahasa asing dibandingkan bahasa ibu mereka. Tanpa upaya pelestarian yang serius, bahasa ini berpotensi punah, seperti yang telah dialami oleh banyak bahasa daerah lainnya di Indonesia. Penurunan penggunaan bahasa daerah juga berdampak pada hilangnya kearifan lokal, tradisi lisan, dan nilai budaya yang melekat dalam bahasa tersebut.

Sebagai bagian dari rumpun bahasa Austronesia, bahasa Lampung memiliki dua dialek utama, yaitu dialek Api dan dialek Nyo. Dialek Api lebih umum digunakan di wilayah selatan Lampung, seperti di Lampung Selatan dan Tanggamus, sedangkan dialek Nyo banyak ditemukan di Lampung bagian utara, seperti di Lampung Utara dan Lampung Tengah (Abidin, 2017). Dialek-dialek ini tidak hanya menjadi sarana komunikasi tetapi juga menjadi cerminan nilai-nilai budaya dan kearifan lokal masyarakat Lampung. Menurut Fakhrurozi et al. (2019), keberadaan dialek ini mencerminkan pola hidup masyarakatnya yang beragam. Oleh karena itu, pelestarian bahasa Lampung berdampak langsung pada adat, budaya, tradisi lisan, hubungan sosial dan lain sebagainya.

Pelestarian bahasa lampung tidak hanya menjadi tanggung jawab pemerintah daerah tetapi juga masyarakat, institusi pendidikan, dan peneliti. Salah satu cara yang menjanjikan untuk mendukung pelestarian bahasa adalah melalui teknologi

speech-to-text. Teknologi ini memungkinkan konversi percakapan verbal menjadi teks, sehingga bahasa Lampung dapat didokumentasikan, dianalisis, dan disebarluaskan. Dalam lingkup pendidikan, teknologi ini dapat digunakan untuk menghasilkan materi pembelajaran bahasa Lampung yang lebih menarik dan interaktif. Di sisi lain, dokumentasi berbasis digital ini juga dapat digunakan untuk melestarikan cerita rakyat, adat istiadat, dan sastra lisan masyarakat Lampung. Dengan demikian, pengembangan teknologi *speech-to-text* tidak hanya mendukung konservasi budaya tetapi juga membuka peluang bagi generasi muda untuk kembali belajar dan menggunakan bahasa ibu mereka.

Salah satu tantangan utama dalam implementasi teknologi *speech-to-text* adalah kualitas input audio yang sering kali terganggu oleh kebisingan lingkungan. Dalam kehidupan sehari-hari, rekaman audio sering kali mengandung suara latar seperti deru kendaraan, angin, suara kerumunan, atau bahkan gangguan teknis dari perangkat perekam itu sendiri. Kebisingan ini tidak hanya mengurangi kualitas rekaman tetapi juga menyebabkan penurunan akurasi sistem *speech-to-text*. Oleh karena itu, reduksi kebisingan menjadi langkah kritis dalam meningkatkan keakuratan transkripsi audio.

Pengolahan sinyal audio telah berkembang dengan berbagai metode untuk mengatasi masalah kebisingan ini. Salah satu metode yang digunakan adalah *Spectral Gating*, yang bekerja dengan menganalisis spektrum frekuensi dalam sinyal audio dan menekan komponen yang dianggap kebisingan di bawah ambang batas tertentu (Shah and Shah, 2023). Metode ini telah terbukti efektif untuk berbagai jenis kebisingan, seperti suara latar kendaraan atau angin, tanpa merusak sinyal utama. Di sisi lain, *Wiener Filtering* adalah metode probabilistik yang memanfaatkan estimasi spektrum sinyal untuk meminimalkan kesalahan kuadrat rata-rata antara sinyal asli dan sinyal hasil pemrosesan (Kumar et al., 2022). Kedua metode ini telah banyak digunakan dalam aplikasi pengolahan sinyal audio, termasuk sistem pengenalan suara, telekomunikasi, dan pengolahan musik (Abdelli and Merazka, 2024).

Namun, hingga saat ini, belum ada penelitian yang secara spesifik mengevaluasi kinerja kedua metode ini dalam konteks *speech-to-text* bahasa Lampung. Padahal, evaluasi semacam ini sangat penting untuk mengetahui metode yang paling efektif dalam mendukung pelestarian bahasa Lampung melalui teknologi digital. Dengan memanfaatkan kedua metode reduksi kebisingan ini, diharapkan sistem *speech-to-text* bahasa Lampung dapat lebih akurat, bahkan dalam kondisi rekaman yang kurang ideal.

Model yang digunakan dalam penelitian ini adalah *Hidden Markov Model* (HMM), yang dikenal sebagai metode yang andal dalam menangani data sekuensial seperti suara. HMM mengandalkan probabilitas transisi antar status tersembunyi untuk memprediksi output berbasis urutan waktu (Graves et al., 2013). Dalam konteks *speech-to-text*, HMM sering digunakan untuk memodelkan hubungan antara fitur akustik dan fonem, memungkinkan sistem untuk mengenali kata yang diucapkan dengan tingkat akurasi yang tinggi (Suryadharma et al., 2014). Penggunaan HMM dalam penelitian ini diharapkan memberikan hasil transkripsi yang optimal, terutama ketika dikombinasikan dengan teknik reduksi kebisingan yang efektif.

Penelitian ini bertujuan untuk mengevaluasi kinerja *speech-to-text* pada audio bahasa Lampung dengan penerapan metode *Spectral Gating* dan *Wiener Filtering* sebagai langkah reduksi kebisingan. Evaluasi ini mencakup pengukuran kualitas rekaman sebelum dan setelah reduksi kebisingan, serta analisis akurasi transkripsi berdasarkan metrik *Word Error Rate* (WER), yang digunakan sebagai standar evaluasi akurasi sistem pengenalan suara otomatis (Neumann et al., 2023). Selain itu, *Character Error Rate* (CER) juga digunakan sebagai metrik yang lebih rinci untuk mengukur kesalahan berbasis karakter dalam sistem pengenalan suara (Karita et al., 2023). Hasil dari penelitian ini diharapkan tidak hanya memberikan kontribusi pada bidang ilmu komputer tetapi juga mendukung pelestarian budaya lokal dan memberikan panduan bagi pengembangan teknologi serupa untuk bahasa daerah lainnya.

Dalam era modern, teknologi telah menjadi alat penting untuk melestarikan budaya yang terancam punah. Dengan memadukan teknologi seperti *speech-to-text* dan pengolahan sinyal audio, penelitian ini tidak hanya menjawab tantangan pelestarian bahasa tetapi juga memberikan kontribusi pada pemanfaatan teknologi untuk tujuan sosial dan budaya. Keberhasilan penelitian ini dapat menjadi model bagi bahasa daerah lainnya di Indonesia yang menghadapi ancaman serupa.

1.2 Rumusan Masalah

Rumusan masalah dari penelitian ini adalah sebagai berikut.

1. Bagaimana penerapan metode *Spectral Gating* dan *Wiener Filtering* memengaruhi kualitas rekaman audio bahasa Lampung berdasarkan pengukuran *Signal-to-Noise Ratio* (SNR)?
2. Seberapa besar tingkat akurasi *speech-to-text* pada audio bahasa Lampung sebelum dan setelah reduksi kebisingan menggunakan kedua metode tersebut?
3. Metode mana yang lebih efektif dalam meningkatkan kinerja *speech-to-text* berbasis *Hidden Markov Model* (HMM) pada audio bahasa Lampung?

1.3 Batasan Masalah

1. Rekaman audio yang digunakan adalah bahasa Lampung dengan durasi pendek.
2. Penelitian difokuskan pada dialek Api, dengan kontribusi dari penutur asli yang memahami, menguasai, dan fasih dalam menggunakan dialek tersebut.

1.4 Tujuan

Tujuan dari penelitian ini adalah sebagai berikut.

1. Menganalisis pengaruh penerapan metode *Spectral Gating* dan *Wiener Filtering* dalam mengurangi kebisingan pada rekaman audio bahasa Lampung berdasarkan pengukuran *Signal-to-Noise Ratio* (SNR).

2. Mengevaluasi tingkat akurasi *speech-to-text* pada audio bahasa Lampung dengan dan tanpa penerapan kedua metode reduksi kebisingan.
3. Menentukan metode reduksi kebisingan yang lebih efektif dalam meningkatkan akurasi transkripsi audio bahasa Lampung dengan *speech-to-text* berbasis *Hidden Markov Model* (HMM).

1.5 Manfaat

Manfaat dari penelitian ini adalah sebagai berikut.

1. Menambah literatur tentang pengolahan sinyal audio untuk bahasa daerah, khususnya dalam konteks reduksi kebisingan dan *speech-to-text*.
2. Memberikan solusi untuk meningkatkan kualitas rekaman audio bahasa Lampung yang dapat digunakan dalam dokumentasi dan pelestarian bahasa.
3. Mengembangkan pendekatan teknologi yang dapat diimplementasikan pada bahasa daerah lain dengan tantangan serupa.

II. TINJAUAN PUSTAKA

2.1 Penelitan Terkait

Penelitian ini memiliki landasan dari berbagai studi sebelumnya yang berkaitan dengan teknologi *speech-to-text*, metode reduksi kebisingan, dan evaluasi kinerja sistem pengenalan suara. Berikut adalah beberapa penelitian yang relevan:

1. Penerapan *Hidden Markov Model* (HMM) untuk Bahasa Daerah Suryadharma et al. (2014) mengeksplorasi penggunaan HMM untuk sistem *speech-to-text* pada bahasa lokal. Penelitian ini menunjukkan bahwa HMM mampu memberikan hasil yang cukup baik meskipun dengan dataset yang terbatas, terutama ketika dikombinasikan dengan optimasi fitur akustik dan model bahasa. Studi ini menjadi dasar penting dalam pengembangan sistem HMM untuk bahasa Lampung, memberikan solusi yang relevan untuk pelestarian bahasa daerah melalui teknologi digital.

2. Reduksi Kebisingan Menggunakan *Wiener Filtering* Penelitian oleh Butarbutar et al. (2023) mengevaluasi keefektifan metode *Wiener Filtering* dalam pengurangan kebisingan audio pada berbagai lingkungan. Dengan memanfaatkan analisis statistik, metode ini terbukti mampu meningkatkan *Signal-to-Noise Ratio* (SNR) secara signifikan. Penelitian ini relevan karena menunjukkan kemampuan *Wiener Filtering* untuk menghasilkan rekaman audio yang lebih bersih dan berkualitas tinggi, yang dapat meningkatkan kinerja sistem *speech-to-text* berbasis HMM. Penelitian ini memberikan wawasan tentang bagaimana kualitas audio dapat ditingkatkan secara substansial melalui pendekatan probabilistik.

3. Peningkatan Suara Menggunakan Spektral Gating Agrawal, Gupta, dan Garg (2023) mengkombinasikan metode spektral gating dengan jaringan saraf dalam untuk meningkatkan kualitas audio. Penelitian ini menunjukkan bahwa spektral gating dapat secara efektif mengurangi kebisingan latar belakang, terutama untuk dataset seperti LibriSpeech dan NOIZEUS. Studi ini memberikan wawasan penting mengenai penggunaan spektral gating untuk pengolahan audio di lingkungan bising, yang relevan untuk meningkatkan akurasi sistem *speech-to-text* berbasis HMM. Penelitian ini juga menunjukkan potensi penggabungan metode ini dengan teknologi lain untuk hasil yang lebih optimal.

4. Evaluasi Kinerja dengan WER dan CER Von Neumann et al. (2023) membahas definisi dan efisiensi *Word Error Rate* (WER) dan *Character Error Rate* (CER) sebagai metrik evaluasi dalam sistem pengenalan suara. Studi ini relevan dengan penelitian ini karena WER dan CER digunakan untuk mengukur akurasi transkripsi *speech-to-text* setelah penerapan metode reduksi kebisingan. Penelitian ini juga menyoroti pentingnya evaluasi yang tepat untuk menilai efektivitas metode pengolahan audio, terutama dalam skenario multi-pembicara atau kondisi rekaman yang kompleks.

5. *Speech-to-Text* dan Dokumentasi Bahasa Lokal Studi oleh Abdelli dan Merazka (2024) menunjukkan bagaimana pengembangan teknologi *speech-to-text* dapat digunakan untuk pelestarian bahasa lokal. Dengan menerapkan metode reduksi kebisingan yang canggih seperti *Wiener Filtering* dan *Spectral Gating*, penelitian ini memberikan bukti empiris bahwa teknologi STT dapat berfungsi sebagai alat dokumentasi dan pelestarian budaya. Studi ini juga menyoroti pentingnya pendekatan multidisiplin untuk mengatasi tantangan bahasa daerah yang memiliki data latih terbatas.

Penelitian-penelitian ini memberikan landasan teoritis dan praktis bagi penelitian ini, baik dalam hal teknologi yang digunakan (HMM), metode pengurangan

kebisingan (*Spectral Gating* dan *Wiener Filtering*), maupun evaluasi kinerja (WER dan CER). Studi-studi ini juga menunjukkan tantangan dan peluang dalam pengembangan teknologi *speech-to-text* untuk bahasa lokal seperti bahasa Lampung.

2.2 Teknologi *Speech-to-Text*

Teknologi s*peech-to-text* (STT) merupakan salah satu terobosan penting dalam bidang pengolahan suara dan kecerdasan buatan. Dengan kemampuan untuk mentranskripsi ucapan manusia menjadi teks tertulis, teknologi ini telah digunakan secara luas di berbagai bidang, mulai dari sistem asisten virtual hingga pelestarian bahasa daerah (Abidin et al., 2020). Kemajuan algoritma pengolahan sinyal audio, seperti penggunaan *Spectrogram* untuk ekstraksi fitur, telah memungkinkan STT untuk mengenali ucapan manusia dengan tingkat akurasi yang semakin tinggi, bahkan dalam kondisi kebisingan yang signifikan (Smith et al., 2019).

Dalam konteks pelestarian bahasa lokal, STT menawarkan peluang untuk mendokumentasikan bahasa yang terancam punah, seperti bahasa Lampung, melalui transkripsi cerita rakyat, percakapan sehari-hari, dan kosakata penting (Widiyanto, 2015). Penjelasan berikut akan mendalami konsep dasar STT, termasuk proses kerja, komponen utama, dan tantangan yang dihadapi, terutama dalam konteks bahasa local (Von Neumann et al., 2023).

2.2.1 *Komponen Utama Speech-to-Text*

Sistem *speech-to-text* terdiri dari tiga komponen utama:

1. Model Akustik (*Acoustic Model*) : Menganalisis fitur akustik dari suara untuk mengenali fonem. Model ini dilatih menggunakan dataset audio yang mencakup berbagai pola ucapan, nada, dan intonasi.
2. Model Bahasa (*Language Model*) : Memperkirakan urutan kata berdasarkan probabilitas linguistik. Model ini memungkinkan sistem memilih kata yang paling sesuai dalam konteks tertentu.

3. Kamus Pelafalan (*Pronunciation Dictionary*) : Menghubungkan fonem dengan kata-kata yang dikenali oleh sistem. Kamus ini penting untuk menyesuaikan pelafalan khusus, terutama untuk bahasa lokal dengan variasi dialek (Graves et al., 2013).

2.2.2 *Konsep Dasar Speech-to-Text*

Speech-to-text (STT) adalah teknologi yang memungkinkan pengubahan ucapan manusia menjadi teks secara otomatis. Teknologi ini menggabungkan algoritma pengolahan sinyal audio dan pembelajaran mesin untuk mengenali, memahami, dan menginterpretasikan suara. Dengan kemampuan ini, STT telah menjadi salah satu teknologi utama dalam pengembangan interaksi manusia dan komputer. Berikut penerapan teknologi STT.

1. Asisten Virtual STT digunakan dalam layanan seperti Siri dan Alexa, yang memproses perintah suara pengguna untuk memberikan respon dalam bentuk teks atau tindakan spesifik. Teknologi ini memungkinkan pengguna untuk melakukan tugas-tugas seperti meminta informasi cuaca atau mengatur pengingat hanya dengan perintah suara (Smith et al., 2019).
2. Sistem Penerjemahan Otomatis dalam aplikasi seperti *Google Translate*, STT digunakan untuk mentranskripsi ucapan menjadi teks. Teks ini kemudian diterjemahkan ke dalam bahasa lain menggunakan model penerjemahan berbasis kecerdasan buatan, menjadikan teknologi ini alat penting dalam komunikasi lintas bahasa (Dewi et al., 2023).
3. Dokumentasi Suara untuk Pelestarian Bahasa Lokal Dalam konteks pelestarian bahasa lokal, STT membantu mendokumentasikan bahasa yang terancam punah, seperti bahasa Lampung. Teknologi ini memungkinkan perekaman cerita rakyat, percakapan sehari-hari, dan kosakata penting untuk pembelajaran dan pelestarian budaya (Gales, 2008).

2.2.3 *Proses Kerja Utama Speech-to-Text*

1. Ekstraksi Fitur Akustik seperti *Spectrogram* diekstraksi dari sinyal suara. Representasi ini menunjukkan intensitas energi dalam domain waktu dan frekuensi, memungkinkan analisis pola suara yang khas. Langkah ini bertujuan untuk menyediakan data yang relevan bagi model pengenalan suara, seperti frekuensi dominan dan intensitas suara. *Spectrogram* telah terbukti menjadi fitur yang sangat efektif dalam sistem STT karena memberikan representasi visual dan numerik dari karakteristik suara (Han et al., 2015).
2. Prediksi Urutan Fonem menggunakan Model Akustik *Hidden Markov Model* (HMM) untuk memprediksi fonem berdasarkan pola akustik yang diekstraksi dari suara. HMM memungkinkan analisis pola sekuensial dengan probabilitas tinggi untuk setiap fonem (Fujimoto and Kawai, 2016).
3. Konversi Fonem menjadi teks ini memperkirakan urutan kata berdasarkan probabilitas linguistik. Model berbasis N-gram atau Transformer sering digunakan untuk memprediksi kata berikutnya dalam konteks tertentu (Ochiai et al., 2020). Kamus pelafalan menghubungkan fonem dengan kata-kata yang dikenali oleh sistem STT, yang sangat penting untuk menangani variasi dialek bahasa lokal (Wang et al., 2023).

2.2.4 *Tantangan Speech-to-Text untuk Bahasa Lokal*

Bahasa lokal seperti bahasa Lampung menghadapi beberapa tantangan unik dalam pengembangan teknologi *speech-to-text*:

1. Bahasa Lampung memiliki keterbatasan data rekaman audio dan teks, sehingga menyulitkan pelatihan model akustik dan bahasa yang berkualitas. Dibutuhkan data tambahan dari berbagai dialek dan konteks untuk meningkatkan akurasi pengenalan suara (Sharma and Raj, 2019).

2. Bahasa Lampung memiliki dua dialek utama, yaitu Api dan Nyo. Perbedaan dalam pelafalan, kosakata, dan struktur kalimat mempersulit pengembangan model universal yang mampu mengenali semua dialek (Abdelli and Merazka, 2024).
3. Sebagian besar sistem STT komersial, seperti *Google Speech* API, tidak mendukung bahasa minoritas karena keterbatasan jumlah penggunanya. Hal ini menjadikan pengembangan berbasis sumber terbuka sebagai solusi utama untuk pelestarian bahasa lokal (Roweis, 2001). Hal ini membuat pengembangan berbasis sumber terbuka menjadi solusi utama.

2.3 *Hidden Markov Model* (HMM)

Hidden Markov Model (HMM) adalah algoritma probabilistik yang digunakan untuk memodelkan data sekuensial, seperti sinyal suara dalam teknologi speech-to-text (STT). HMM memungkinkan sistem untuk menghubungkan fitur akustik dengan fonem melalui transisi antar status tersembunyi berdasarkan probabilitas. Algoritma ini digunakan secara luas dalam pengenalan suara karena kemampuannya memodelkan hubungan temporal antar elemen suara (Fujimoto and Kawai, 2016).

2.3.1 *Rumus Hidden Markov Model (HMM)*

HMM menggunakan pendekatan probabilistik untuk menghitung kemungkinan observasi P(O|λ), yang didefinisikan sebagai:

$$P(O \mid \lambda) = \sum_{Q} P(O \mid Q, \lambda) \cdot P(Q \mid \lambda) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Di mana:

1. $P(O \mid \lambda)$: Probabilitas untuk melihat atau menghasilkan urutan observasi O={o1,o2,...,oT} berdasarkan model HMM λ. Ini adalah nilai yang ingin dihitung oleh sistem, misalnya dalam pengenalan ucapan.

2. Σ_Q : Penjumlahan atas semua kemungkinan urutan state tersembunyi Q={q1,q2,...,qT}. Karena tidak tahu pasti state mana yang sebenarnya terjadi, maka semua kemungkinan harus diperhitungkan.
3. $P(O \mid Q, \lambda)$: Probabilitas menghasilkan observasi O jika diketahui urutan state Q dan model λ Ini biasanya dihitung dari fungsi emisi B pada model.
4. $P(Q \mid \lambda)$: Probabilitas terjadinya urutan state tersembunyi Q dalam model λ dihitung berdasarkan transisi antar state A dan probabilitas awal π.

Rumus ini berasal dari konsep dasar HMM yang banyak digunakan dalam pengenalan suara dan dijelaskan secara rinci dalam penelitian oleh Fujimoto and Kawai (2016) dan Gales (2008).

2.3.2 *Proses Kerja Hidden Markov Model (HMM)*

HMM bekerja dengan memodelkan hubungan antara status tersembunyi dan data observasi melalui probabilitas transisi dan emisi. Proses utamanya melibatkan tiga langkah:

1. Identifikasi Status Tersembunyi (Hidden States): Merepresentasikan elemen fonetik dari sinyal suara, seperti fonem atau unit akustik.
2. Analisis Observasi : Menggunakan fitur akustik, seperti *Spectrogram*, untuk memetakan informasi suara dalam domain waktu dan frekuensi.
3. Probabilitas Transisi Antar Status: Menghitung kemungkinan transisi dari satu status tersembunyi ke status lain, yang direpresentasikan melalui matriks probabilitas transisi.

HMM menggunakan algoritma Viterbi untuk menemukan urutan status tersembunyi dengan probabilitas maksimum, yang kemudian diterjemahkan menjadi teks.

2.3.3 *Keunggulan HMM dalam STT*

1. Kemampuan untuk Menangani Dataset Terbatas:
   HMM bekerja dengan baik pada dataset kecil hingga menengah, menjadikannya ideal untuk pengenalan suara dalam bahasa lokal dengan keterbatasan data (Widiyanto, 2015).
2. Fleksibilitas dalam Memodelkan Variasi Akustik:
   Dengan parameter yang dapat disesuaikan, HMM mampu menangkap variasi suara dari berbagai pengguna, termasuk variasi pelafalan dan intonasi (Fujimoto and Kawai, 2016).
3. Integrasi dengan Fitur Akustik Modern:
   Kombinasi HMM dengan fitur seperti *Spectrogram* atau Mel-Frequency Cepstral Coefficients (MFCC) meningkatkan akurasi pengenalan suara (Smith et al., 2019).

2.3.4 *Keterbatasan HMM*

1. Probabilitas transisi dalam HMM bersifat tetap, sehingga sulit menangkap hubungan temporal yang kompleks. Alternatif seperti Recurrent Neural Networks (RNN) lebih efektif untuk konteks temporal yang panjang (Xu et al., 2015).
2. Dalam kondisi kebisingan tinggi, HMM sering mengalami penurunan akurasi. Oleh karena itu, pengurangan kebisingan seperti *Spectral Gating* atau *Wiener Filtering* diperlukan sebelum proses pengenalan (Smith et al., 2019).

2.4 Reduksi Kebisingan

Reduksi kebisingan adalah langkah kritis dalam pengolahan sinyal audio, khususnya dalam teknologi s*peech-to-text* (STT). Langkah ini bertujuan untuk

meningkatkan akurasi transkripsi dengan memisahkan sinyal suara utama dari kebisingan latar tanpa merusak kualitas sinyal asli. Kebisingan dalam rekaman audio dapat berasal dari berbagai sumber, termasuk lingkungan sekitar (misalnya: suara kendaraan, angin, kerumunan), perangkat perekaman berkualitas rendah, atau gangguan elektromagnetik (Smith et al., 2019).

Proses reduksi kebisingan bertujuan untuk meningkatkan *Signal-to-Noise Ratio* (SNR), yang merupakan parameter penting dalam evaluasi kualitas sinyal audio. SNR yang lebih tinggi mencerminkan sinyal utama yang lebih dominan dibandingkan kebisingan latar, sehingga menghasilkan transkripsi yang lebih akurat (Gannot et al., 2015).

Dalam penelitian ini, dua metode utama digunakan untuk reduksi kebisingan:

1. *Spectral Gating*
2. *Wiener Filtering*

2.4.1 *Spectral Gating*

Spectral Gating adalah teknik reduksi kebisingan yang bekerja dengan menekan atau menghilangkan komponen spektral di bawah ambang batas tertentu. Teknik ini efektif untuk menangani kebisingan latar yang konsisten, seperti suara angin, kipas, atau kerumunan (Shah & Shah, 2023). Analisis Spektrum Frekuensi : Menggunakan Transformasi Fourier Cepat (FFT) untuk mengonversi sinyal audio dari domain waktu ke domain frekuensi. Proses ini membantu mengidentifikasi komponen frekuensi yang dominan (Naik et al., 2021).

Proses Kerja:

1. Analisis Spektrum Frekuensi :
   Menggunakan Transformasi Fourier Cepat (FFT) untuk mengubah sinyal audio dari domain waktu ke domain frekuensi. Proses ini membantu mengidentifikasi komponen frekuensi yang dominan dalam sinyal (Naik et al., 2021).

$$X(f) = \sum_{n=0}^{N-1} x(n) \cdot e^{-j2\pi fn/N} \quad \text{.....(2)}$$

Di mana:

$X(f)$: Spektrum frekuensi.

$x(n)$: Sinyal dalam domain waktu.

N : Panjang sinyal.

2. Penentuan Ambang Batas Kebisingan :

   Ambang batas ditentukan berdasarkan intensitas komponen spektral. Komponen dengan amplitudo di bawah ambang ini dianggap sebagai kebisingan dan dihilangkan.

3. Penerapan Gating :

   Komponen frekuensi di bawah ambang batas dihilangkan, sementara frekuensi di atas ambang tetap dipertahankan. Teknik ini memungkinkan pengurangan kebisingan tanpa merusak sinyal utama (Shah & Shah, 2023).

Kelebihan:

1. Efektif untuk kebisingan yang konsisten.
2. Mudah diterapkan dalam perangkat lunak pengolahan sinyal audio (Agrawal et al., 2023).
3. Spektral gating sangat efektif dalam mengurangi kebisingan latar seperti suara angin, kipas, atau kerumunan. Penelitian oleh Shah dan Shah (2023) menunjukkan bahwa metode ini mampu meningkatkan kejernihan audio secara signifikan tanpa merusak sinyal asli.

2.4.2 *Wiener Filtering*

Wiener Filtering adalah teknik reduksi kebisingan berbasis statistik yang meminimalkan kesalahan kuadrat rata-rata antara sinyal asli dan sinyal yang diestimasi. Teknik ini menggunakan pendekatan probabilistik untuk memisahkan komponen sinyal dari kebisingan.

Proses Kerja:

1. Estimasi Spektrum Sinyal dan Kebisingan :
   Spektrum sinyal dan kebisingan diestimasi menggunakan data statistik dari sinyal yang diinginkan dan kebisingan latar (Butarbutar et al., 2023).
2. Perhitungan Filter *Wiener* :
   Filter *Wiener* dihitung dengan persamaan (3) :

$$H(f) = \frac{S(f)}{S(f)+N(f)} \quad \text{.....(3)}$$

   Di mana:
   $H(f)$: Filter *Wiener*.
   $S(f)$: Spektrum sinyal.
   $S(f) + N(f)$: Spektrum kebisingan.

3. Aplikasi pada Sinyal Campuran:
   Filter diterapkan pada sinyal campuran untuk menghasilkan sinyal bersih dengan kebisingan yang telah diminimalkan.

Kelebihan :

1. Efektif dalam kondisi kebisingan yang terukur dan stabil, seperti kebisingan dari perangkat elektronik atau ruangan tertutup.
2. *Wiener Filtering* sangat efektif dalam kondisi kebisingan yang terukur dan stabil, seperti kebisingan dari perangkat elektronik atau rekaman dalam ruangan. Penelitian oleh Kumar dan Chari (2020) menunjukkan bahwa metode ini mampu meningkatkan SNR secara signifikan, menghasilkan kualitas suara yang lebih bersih.

2.5 Evaluasi Kinerja *Speech-to-Text*

Evaluasi kinerja sistem s*peech-to-text* (STT) sangat penting untuk mengukur efektivitas teknologi dalam mengubah ucapan menjadi teks tertulis. Pendekatan

evaluasi menggunakan tiga metrik utama, yaitu *Word Error Rate* (WER), *Character Error Rate* (CER), dan *Signal-to-Noise Ratio* (SNR). Ketiga metrik ini memberikan analisis yang saling melengkapi untuk menilai akurasi transkripsi dan kualitas sinyal audio.

Word Error Rate (WER) adalah metrik utama untuk mengevaluasi kesalahan transkripsi pada tingkat kata. WER dihitung berdasarkan jumlah substitusi, penghapusan, dan penambahan kata dalam hasil transkripsi dibandingkan dengan teks referensi. Metrik ini memberikan gambaran keseluruhan tentang kemampuan sistem STT dalam mengenali kata-kata yang diucapkan. WER sangat relevan dalam sistem multi-pembicara atau bahasa lokal seperti bahasa Lampung, di mana variasi pelafalan dan struktur kalimat dapat memengaruhi akurasi transkripsi. Penelitian menunjukkan bahwa metode pengurangan kebisingan, seperti *Spectral Gating* dan *Wiener Filtering*, dapat secara signifikan menurunkan nilai WER dengan meningkatkan kejernihan sinyal audio (Von Neumann et al., 2023).

Selain WER, *Character Error Rate* (CER) digunakan sebagai metrik pelengkap untuk mengevaluasi kesalahan transkripsi pada tingkat karakter. CER lebih sensitif terhadap kesalahan kecil seperti penghilangan atau penambahan huruf, yang sering terjadi dalam bahasa dengan struktur fonetik kompleks. Dalam bahasa Lampung, yang memiliki struktur kata pendek, CER memberikan detail tambahan tentang jenis kesalahan yang tidak dapat ditangkap oleh WER. Hal ini menjadikan CER metrik yang ideal untuk mengevaluasi sistem STT dalam konteks bahasa lokal dengan variasi dialek (Karita et al., 2023). CER juga membantu mengidentifikasi kesalahan ejaan atau fonemik yang mungkin memengaruhi keakuratan transkripsi secara keseluruhan.

Sementara itu, *Signal-to-Noise Ratio* (SNR) digunakan untuk mengevaluasi kualitas sinyal audio dengan membandingkan energi sinyal utama terhadap energi kebisingan. SNR yang lebih tinggi mencerminkan kualitas sinyal yang lebih baik, dengan sinyal utama yang lebih dominan dibandingkan kebisingan. Dalam sistem STT, SNR yang rendah sering kali menyebabkan peningkatan nilai WER dan CER,

yang berdampak negatif pada akurasi transkripsi. Oleh karena itu, metode reduksi kebisingan, seperti *Spectral Gating* dan *Wiener Filtering*, sangat penting untuk meningkatkan SNR sebelum sinyal audio diproses lebih lanjut. Dalam konteks lingkungan yang bising, seperti suara kendaraan atau angin, peningkatan SNR dapat secara signifikan meningkatkan akurasi pengenalan suara (Kumar & Chari, 2020). Dalam penerapannya pada bahasa Lampung, peningkatan SNR membantu mengurangi gangguan lingkungan yang sering kali ditemukan dalam rekaman audio bahasa lokal, sehingga memungkinkan pengenalan fonem dan kata dengan lebih baik.

Secara keseluruhan, evaluasi kinerja STT melalui WER, CER, dan SNR memberikan pendekatan holistik untuk menilai efektivitas sistem dalam menangani berbagai kondisi dan skenario. Kombinasi dari ketiga metrik ini memastikan analisis yang komprehensif, terutama untuk sistem STT yang diterapkan pada bahasa lokal seperti bahasa Lampung, di mana tantangan pelafalan dan kondisi lingkungan menjadi faktor yang signifikan dalam menentukan kinerja.

2.5.1 *Word Error Rate (WER)*

Word Error Rate (WER) adalah metrik evaluasi yang digunakan untuk menghitung kesalahan transkripsi berbasis kata dalam sistem s*peech-to-text* (STT). Metrik ini mengukur jumlah kata yang salah dikenali, dihapus, atau ditambahkan oleh sistem dibandingkan dengan teks referensi. Rumus WER adalah sebagai berikut :

$$WER = \frac{S+D+I}{N} \quad \text{.....(4)}$$

Di mana :

S : Jumlah substitusi (kata salah).

D : Jumlah deleksi (kata yang dihapus).

I : Jumlah insepsi (kata tambahan).

N : Jumlah total kata dalam teks referensi.

WER memberikan gambaran tingkat kesalahan transkripsi secara keseluruhan. Sebuah sistem *speech-to-text* dengan WER rendah menunjukkan akurasi yang tinggi dalam mengenali kata-kata yang diucapkan (Graves et al., 2013). Penelitian oleh Von Neumann et al. (2023) menunjukkan pentingnya WER dalam evaluasi sistem multi-pembicara. Dalam konteks bahasa lokal seperti Lampung, WER memberikan ukuran tingkat kesalahan yang mencerminkan kemampuan sistem dalam memahami variasi pelafalan dan dialek.

Salah satu keunggulan utama WER adalah kemampuannya memberikan evaluasi langsung terhadap performa sistem STT dalam skenario nyata. Namun, metrik ini memiliki keterbatasan dalam mendeteksi kesalahan kecil, seperti perbedaan ejaan atau perubahan huruf dalam kata. Untuk mengatasi keterbatasan ini, *Character Error Rate* (CER) digunakan sebagai metrik pelengkap.

2.5.2 *Character Error Rate (CER)*

Character Error Rate (CER) adalah metrik pelengkap yang menghitung kesalahan pada tingkat karakter dalam transkripsi. CER lebih sensitif terhadap kesalahan kecil, seperti penghilangan huruf atau perubahan ejaan, yang mungkin tidak terdeteksi oleh WER. Rumus CER serupa dengan WER, tetapi menggantikan kata dengan karakter, dengan persamaan (5) :

$$CER = \frac{S+D+I}{N} \quad \text{.....(5)}$$

Di mana:

S : Jumlah substitusi (karakter salah).

D : Jumlah deleksi (karakter yang dihapus).

I : Jumlah insepsi (karakter tambahan).

N : Jumlah total karakter dalam teks referensi.

CER lebih sensitif terhadap kesalahan kecil, seperti kesalahan ejaan atau penghilangan huruf. Dalam bahasa Lampung, CER relevan karena struktur kata yang pendek sering kali membuat kesalahan kecil lebih berdampak pada akurasi transkripsi. Penelitian oleh Von Neumann et al. (2023) menunjukkan bahwa CER efektif untuk mengevaluasi akurasi transkripsi dalam konteks bahasa dengan struktur fonetik yang unik, seperti bahasa Lampung.

CER memberikan analisis yang lebih rinci dibandingkan WER, tetapi metrik ini memiliki keterbatasan dalam menangkap konteks linguistik dari kata yang salah. Namun, kombinasi CER dan WER memberikan evaluasi yang komprehensif terhadap sistem STT.

2.5.3 *Signal-to-Noise Ratio (SNR)*

Signal-to-Noise Ratio (SNR) adalah ukuran yang digunakan untuk mengevaluasi kualitas sinyal audio dengan membandingkan kekuatan atau energi dari sinyal yang diinginkan terhadap energi kebisingan di dalam sinyal tersebut. Dalam konteks speech-to-text, SNR merupakan parameter penting yang memengaruhi akurasi sistem pengenalan suara, terutama dalam kondisi lingkungan yang bising (Butarbutar et al., 2023). SNR dinyatakan dalam satuan desibel (dB) dan dihitung menggunakan persamaan (6) :

$$SNR = 10 \cdot \log_{10} \left(\frac{P\ signal}{P\ noise} \right) \quad \text{.....(6)}$$

Di mana :
$P\ signal$: Energi total sinyal suara utama.
$P\ noise$: Energi total kebisingan latar.

Semakin tinggi nilai SNR, semakin dominan sinyal yang diinginkan dibandingkan kebisingan, sehingga kualitas audio lebih baik.

Sebaliknya, SNR yang rendah menunjukkan bahwa kebisingan lebih dominan, yang dapat mengganggu proses transkripsi.
SNR memberikan gambaran tentang dominasi sinyal suara utama dibandingkan kebisingan. Dalam sistem STT, nilai SNR yang lebih tinggi menunjukkan kualitas sinyal yang lebih baik, yang mendukung pengenalan suara dengan akurasi lebih tinggi. Sebaliknya, SNR rendah sering kali menyebabkan kesalahan dalam mengenali fonem dan kata, sehingga meningkatkan nilai WER dan CER (Smith et al., 2019).

SNR memiliki peran yang sangat penting dalam menentukan akurasi sistem *speech-to-text*. Audio dengan SNR yang rendah sering kali menyebabkan kesalahan dalam mengenali fonem dan kata, sehingga meningkatkan *Word Error Rate* (WER) dan *Character Error Rate* (CER). Oleh karena itu, teknik reduksi kebisingan seperti *Spectral Gating* dan *Wiener Filtering* digunakan untuk meningkatkan SNR sebelum proses transkripsi dilakukan (Kumar & Chari, 2020).

III. METODOLOGI PENELITIAN

3.1 Pendahuluan

Metodologi penelitian ini dirancang untuk mengevaluasi kinerja teknologi *Speech-to-Text* (STT) pada bahasa Lampung dialek Api, dengan fokus pada penerapan metode reduksi kebisingan *Spectral Gating* dan *Wiener Filtering*. Tantangan utama dalam sistem STT adalah kualitas input audio yang sering terganggu oleh kebisingan lingkungan, yang dapat menurunkan akurasi transkripsi, diukur menggunakan metrik *Word Error Rate* (WER), *Character Error Rate* (CER), dan *Signal-to-Noise Ratio* (SNR).

Dataset yang digunakan terdiri dari rekaman cerita bahasa Lampung dialek Api, yang merepresentasikan keunikan pelafalan dan struktur bahasa. Rekaman ini akan melalui proses reduksi kebisingan sebelum diolah oleh model STT berbasis *Hidden Markov Model* (HMM).

Tujuan utama penelitian ini adalah menentukan metode reduksi kebisingan terbaik dalam meningkatkan kinerja STT pada bahasa Lampung. Hasil penelitian diharapkan tidak hanya mendukung pelestarian bahasa Lampung melalui dokumentasi digital, tetapi juga memberikan panduan teknis bagi pengembangan sistem serupa untuk bahasa daerah lainnya.

3.2 Tempat dan Waktu Penelitian

3.2.1 Tempat Penelitian

Penelitian ini dilaksanakan di dua lokasi utama. Lokasi pertama adalah Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung, tempat kegiatan analisis, pemrosesan data, dan evaluasi sistem dilakukan. Lokasi kedua adalah Rumah Tokoh Adat Dewan Perwakilan Penyimbang Adat Lampung, Desa Kurungan

Nyawa, Kabupaten Pesawaran, tempat pengumpulan data dilakukan melalui perekaman langsung penutur asli bahasa Lampung dialek Api.

3.2.2 Waktu Penelitian

Penelitian ini dilaksanakan pada bulan Oktober 2024 di semester genap hingga penyelesaian pada bulan Maret 2025. Alur waktu pengerjaan penelitian ini dapat dilihat pada Tabel 1. Pada Tabel 1 menjelaskan tentang alur waktu pengerjaan penelitian yang dibagi menjadi 3 tahap, yaitu :

1. **Tahap Penelitian Awal**
   a. Penentuan Tema, Analisis, dan Pengumpulan Studi Literatur : Tahap ini berfokus pada identifikasi permasalahan, penentuan tema penelitian, dan pengumpulan literatur yang relevan. Literatur yang dikaji mencakup referensi tentang teknologi s*peech-to-text*, metode *Spectral Gating* dan *Wiener Filtering* untuk reduksi kebisingan, serta penerapan *Hidden Markov Model* (HMM) sebagai model transkripsi. Studi ini bertujuan untuk membangun dasar konseptual penelitian.
   b. Pengumpulan Dataset : Dataset berupa rekaman audio bahasa Lampung dialek Api dikumpulkan dengan melibatkan penutur asli. Rekaman dilakukan menggunakan perangkat perekam suara berkualitas tinggi dalam format standar .wav. Dataset mencakup variasi kata, kalimat, dan frasa umum yang digunakan dalam percakapan sehari-hari bahasa Lampung.
   c. Penyusunan Proposal Penelitian : Tahapan ini melibatkan penyusunan proposal penelitian yang mencakup latar belakang, rumusan masalah, tujuan penelitian, metodologi, dan rencana pelaksanaan. Proposal ini menjadi dokumen acuan selama pelaksanaan penelitian.
2. **Tahap Penelitian Lanjut**
   a. Prapemrosesan Data Audio : Rekaman audio yang telah dikumpulkan diproses dengan metode reduksi kebisingan

menggunakan *Spectral Gating* dan *Wiener Filtering*. Setelah itu, data dibagi menjadi potongan kecil (framing) dan dihaluskan menggunakan teknik windowing untuk mempersiapkan tahap analisis lebih lanjut.

b. Ekstraksi Fitur : Ekstraksi fitur dilakukan menggunakan *Spectrogram* sebagai representasi utama sinyal audio. Fitur ini dipilih karena kemampuannya untuk menangkap informasi frekuensi dan intensitas sinyal, yang relevan untuk pengenalan suara.

c. Perancangan Model HMM : Model *Hidden Markov Model* (HMM) dirancang untuk mengenali pola suara berdasarkan fitur yang telah diekstraksi. Model ini digunakan untuk memetakan pola suara ke teks secara otomatis.

d. Pengujian Dataset : Dataset yang telah diproses diuji menggunakan model HMM. Proses ini bertujuan untuk menghasilkan transkripsi teks dari rekaman suara bahasa Lampung dialek Api.

3. Tahap Evaluasi

a. Evaluasi Hasil dilakukan untuk mengukur performa model menggunakan tiga metrik utama:

- *Word Error Rate* (WER): Mengukur kesalahan transkripsi pada tingkat kata.
- *Character Error Rate* (CER): Mengukur kesalahan transkripsi pada tingkat karakter.
- *Signal-to-Noise Ratio* (SNR): Menilai kualitas sinyal audio yang telah diproses.

b. Hasil evaluasi dianalisis untuk menentukan efektivitas metode reduksi kebisingan yang digunakan. Penyusunan hasil penelitian dilakukan sebagai dokumentasi akhir untuk memberikan kontribusi pada pengembangan teknologi s*peech-to-text* bagi bahasa lokal. Tahap ini bertujuan untuk menganalisis performa model dan menyusun hasil penelitian.

Tabel 1. Alur dan Waktu Penelitian

Tahapan	Kegiatan	2024												2025											
		Oktober				November				Desember				Januari				Februari				Maret			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Penelitian Awal	Penentuan Tema, Analisis dan Pengumpulan Studi Literatur																								
	Pengumpulan Dataset																								
	Penyusunan Draft Proposal																								
Penelitian Lanjutan	Prapemrosesan Dataset Audio																								
	Ekstraski Fitur																								
	Perancangan Model																								
Evaluasi	Pengujian Dataset																								
	Penyusunan Draft Hasil dan Seminar Hasil Penelitian																								

3.3 Data dan Alat

3.3.1 Dataset Penelitian

Dataset utama dalam penelitian ini berupa rekaman suara cerita pendek dalam Bahasa Lampung Dialek Api. Cerita ini menggambarkan pengalaman liburan ke pantai dan dirancang untuk mencakup variasi intonasi, kosakata, serta kompleksitas struktur kalimat dalam Bahasa Lampung. Dialek Api, sehingga memberikan keragaman data untuk evaluasi sistem s*peech-to-text* (STT).

Proses pengumpulan data melibatkan empat penutur asli dari Kabupaten Pesawaran yang memiliki keaslian dialek. Masing-masing penutur membacakan cerita ini sebanyak satu kali, menghasilkan total empat rekaman cerita penuh. Rekaman dilakukan menggunakan mikrofon berkualitas tinggi dengan format WAV pada frekuensi sampling 16 kHz untuk memastikan kualitas suara yang optimal.

Setelah proses perekaman selesai, cerita ini akan diproses lebih lanjut dengan cara memotong teks cerita menjadi potongan kalimat. Setiap kalimat yang telah dipotong akan digunakan sebagai unit data untuk mempermudah proses transkripsi dan evaluasi sistem s*peech-to-text* (STT). Pendekatan ini memastikan bahwa evaluasi dilakukan secara granular, memungkinkan analisis kinerja sistem dalam mengenali setiap kalimat secara akurat.

Tabel 2. Potongan Kalimat Cerita Pendek dalam Bahasa Lampung

No	Kalimat dalam Bahasa Lampung
1	Seminggu sai likut, hulun tuhaku wat rencana liburan mit pantai.
2	Jam nunjukko pukul 5 mehayu, lamon keluargaku khadu siap liburan mit Pantai Bajul Mati.
3	Sikam lapah mehayu ulah jarak annyak mahhan mit pantai sekitar 3 jam lapahan.

No	Kalimat dalam Bahasa Lampung
4	Adekku khik nyak mak dijuk ngusung HP ulahni hulun tuhaku ulah sikam haga nikmati nihan lapahan liburan.
5	Jadi, sikam begukhau di mubil pakai permainan tebak hurup pertama.
6	Khua jam khadu lapah. Sikam khadu mula cakak bukit.
7	Adikku khik nyak mulai khadu ngeliak khelaya kejadian liburan.
8	Sikam khua gering nihan ngeliak pemandangan sabah sai helau di bah.
9	Cuma, nyak angkah dapok ngeliak pemandangan sekhebok gawoh, ulah annyak sina nyak khadu ngekhasa pudokh khik sakik ulah khelaya sai biluk-biluk, cakak khik khegoh.
10	Mak ku ngejuk minyak angin. Cawani dapok ngukhangi khasa mual.
11	Akhirni sikam sampai di Pantai Bajul Mati.
12	Pantai hiji wat di Kabupaten Malang, Jawa Timur.
13	Sikam langsung lapah mit gazebo lunik di pantai.
14	Suwa ngekhasako angin, sikam mengan jak usungan sikam.
15	Nyak mak dapok bekhadu ngeliak kehelauan pantai hiji.
16	Pantai Bajul Mati mak pikha terkenal, jadi mak lamon pengunjung sai khatong mit dija, jadi pantaini pagun dawak nihan sai dapok ngejadiko sikam nyaman.

Proses rekaman dilakukan dengan mempertimbangkan faktor usia, jenis kelamin, dan keaslian dialek penutur untuk memastikan variasi data yang representatif. Cerita ini dipilih karena mencerminkan penggunaan bahasa sehari-hari dalam konteks percakapan yang natural. Dataset ini memberikan tantangan transkripsi berupa kompleksitas fonetik dan struktur kalimat yang khas dalam Bahasa Lampung Dialek Api.

Dataset yang dihasilkan akan digunakan dalam proses evaluasi model s*peech-to-text* berbasis *Hidden Markov Model* (HMM) untuk mengukur akurasi transkripsi, yang dievaluasi melalui metrik *Word Error Rate* (WER),

Character Error Rate (CER), dan kualitas sinyal audio berdasarkan *Signal-to-Noise Ratio* (SNR).

3.3.2 Alat dan Perangkat Lunak

Untuk memastikan kelancaran proses pengumpulan data, pemrosesan suara, pemodelan, dan evaluasi sistem, penelitian ini menggunakan alat dan perangkat lunak berikut:

1. **Alat Perekaman Audio**

   Mikrofon SARAMONIC Blink 500 B2 digunakan untuk merekam suara penutur asli Bahasa Lampung Dialek Api. Alat ini memiliki kualitas tinggi dan mendukung frekuensi sampling 16 kHz, yang memastikan suara terekam dengan jelas dan minim gangguan kebisingan. Format rekaman disimpan dalam WAV, yang merupakan standar untuk pemrosesan audio berkualitas tinggi.

2. ***Software***

   Proses reduksi kebisingan, ekstraksi fitur, dan pemodelan dilakukan menggunakan perangkat lunak berikut :

   - Bahasa pemrograman utama yang digunakan dalam penelitian ini yaitu *Python* 3.11.12.
   - *Library* yang digunakan dalam penelitian ini yaitu *librosa*, *soundfile*, *noisereduce*, *numpy*, *pandas*, *hmmlearn*, *jiwer*, dan *whisper*.

3. ***Hardware***

   Proses pemrosesan dan pelatihan model dilakukan pada perangkat komputer dengan spesifikasi berikut :

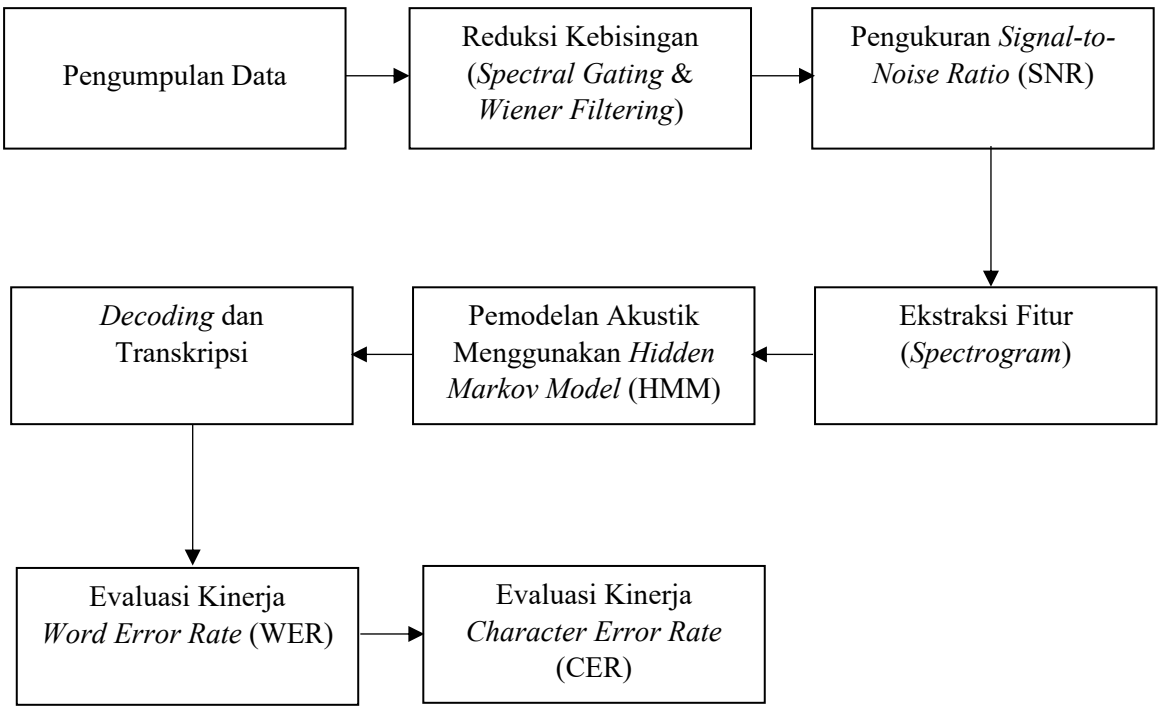
   - Device Name: HP_ENVY
   - Processor: AMD Ryzen 7 4700U with Radeon Graphics, 2.00 GHz
   - RAM: 16.0 GB (15.4 GB usable)
   - Storage: SSD 500 GB

- Operating System: Windows 11 Home Single Language, 64-bit
- OS Version: Version 24H2, Build 26100.2605

3.4 Alur Kerja Penelitian

Eksperimen dilakukan dalam tiga tahap penelitian utama, yang masing-masing memiliki tujuan dan pendekatan berbeda:

1. Penelitian Pertama: Membangun sistem dasar STT berbasis MFCC dan HMM serta menguji efek *noise reduction* terhadap kinerja pengenalan kalimat yang sama (data pelatihan dan pengujian dari sumber yang sama).
2. Penelitian Kedua: Menguji generalisasi sistem terhadap kalimat baru yang memiliki kata-kata yang mirip dengan data pelatihan, serta mencoba pendekatan unit kata untuk pelatihan HMM.
3. Penelitian Ketiga: Mengevaluasi efek nyata dari metode *noise reduction* terhadap kualitas transkripsi dengan menggunakan model *Whisper* (OpenAI) sebagai pendekatan modern.



Gambar 1. Alur Penelitian

Penelitian ini bertujuan untuk mengembangkan model transkripsi suara ke teks dalam Bahasa Lampung dialek Api dengan menerapkan metode *Spectral Gating* dan *Wiener Filtering* untuk reduksi kebisingan, serta menggunakan *Hidden Markov Model* (HMM) sebagai model akustik. Alur kerja penelitian dirancang secara sistematis untuk memastikan setiap tahapan mendukung tujuan penelitian. Masing-masing tahapan akan dijelaskan secara rinci mulai dari proses pembuatan data, pelatihan model, hingga evaluasi dengan metrik *Word Error Rate* (WER) dan *Character Error Rate* (CER).

3.4.1 Pengumpulan Data

Tahap awal penelitian ini adalah pengumpulan data berupa rekaman audio bahasa Lampung dengan dialek Api. Rekaman diambil dari penutur asli yang fasih menggunakan dialek tersebut untuk memastikan keaslian dan keakuratan pelafalan. Data direkam dalam lingkungan dengan tingkat kebisingan yang bervariasi, seperti lingkungan yang tenang hingga cukup bising, guna mencerminkan kondisi nyata di lapangan.

Setiap rekaman berdurasi singkat, sekitar 1–2 menit, untuk mempermudah proses analisis. Data yang dikumpulkan dinormalisasi menggunakan perangkat lunak pengolahan audio untuk memastikan volume dan kualitas rekaman konsisten di seluruh dataset.

3.4.2 Reduksi Kebisingan

Pada tahap ini, rekaman audio yang telah dikumpulkan diproses menggunakan metode *Spectral Gating* dan *Wiener Filtering* untuk mengurangi kebisingan latar. Proses ini bertujuan untuk meningkatkan kualitas sinyal suara sebelum dilakukan transkripsi. Langkah-langkahnya adalah sebagai berikut:

1. *Spectral Gating*:

   Metode ini digunakan untuk menekan komponen frekuensi di bawah ambang batas tertentu, yang biasanya merupakan kebisingan latar seperti suara kipas atau kerumunan. Proses melibatkan analisis spektral menggunakan Transformasi Fourier Cepat (FFT) untuk memisahkan komponen suara dan kebisingan.
2. *Wiener Filtering*:

   Metode ini berbasis statistik dan bertujuan untuk meminimalkan kesalahan kuadrat rata-rata antara sinyal asli dan sinyal hasil reduksi. *Wiener Filtering* bekerja dengan memperkirakan spektrum kebisingan dan spektrum sinyal suara yang diinginkan.

3.4.3 Pengukuran Signal-to-Noise Ratio (SNR)

Kualitas sinyal suara dievaluasi menggunakan metrik *Signal-to-Noise Ratio* (SNR) sebelum dan setelah reduksi kebisingan. Langkah ini memastikan bahwa kedua metode yang diterapkan mampu meningkatkan dominasi sinyal utama dibandingkan kebisingan latar. Perbedaan nilai SNR menunjukkan tingkat efektivitas kedua metode.

3.4.4 Ekstraksi Fitur

Proses ekstraksi fitur dalam penelitian ini menggunakan dataset rekaman suara Bahasa Lampung dialek Api yang telah melalui reduksi kebisingan. Dataset terdiri dari cerita pendek yang diucapkan oleh empat penutur asli. Rekaman ini dipotong menjadi potongan kalimat pendek untuk mempermudah analisis dan pemrosesan.

Fitur yang diambil adalah representasi akustik dari suara menggunakan *Spectrogram*. Rekaman audio diproses menjadi representasi visual yang mencakup intensitas energi pada berbagai frekuensi. Proses ini

bertujuan untuk menyiapkan data agar dapat dianalisis oleh model *Hidden Markov Model* (HMM).

Tahapan ini memastikan bahwa setiap potongan suara memberikan informasi yang cukup untuk mendeteksi pola fonem khas bahasa Lampung dialek Api.

3.4.5 Pemodelan Akustik Menggunakan *Hidden Markov Model* (HMM)

Model akustik pada penelitian ini dibangun menggunakan *Hidden Markov Model* (HMM). Data yang telah diekstraksi fiturnya dibagi menjadi dua bagian:

4. Data Latih (80%): Digunakan untuk melatih HMM agar dapat mengenali pola suara yang khas dalam dataset.
5. Data Uji (20%): Digunakan untuk menguji akurasi model terhadap data baru yang tidak termasuk dalam proses pelatihan.

HMM mempelajari pola suara dari data latih, yang mencakup variasi intonasi dan pelafalan. Pemodelan dilakukan menggunakan perangkat lunak berbasis Python dengan pustaka HTK Toolkit. Tahapan ini memastikan HMM mampu mengenali urutan fonem dari rekaman suara Bahasa Lampung.

3.4.6 *Decoding* dan Transkripsi

Setelah model HMM dilatih, proses *decoding* dilakukan untuk mengubah rekaman suara menjadi teks. Dataset uji yang telah diproses dimasukkan ke dalam model HMM untuk menghasilkan prediksi transkripsi. Proses *decoding* dilakukan dengan menginterpretasikan urutan fonem yang dikenali oleh model dan mencocokkannya dengan kata-kata dalam kamus pelafalan Bahasa Lampung.

Hasil transkripsi disimpan dalam file teks untuk dibandingkan dengan transkripsi referensi (*ground-truth*). Langkah ini memastikan bahwa sistem mampu menghasilkan teks yang mendekati ucapan asli.

3.5 Alur Penelitian Ketiga : Evaluasi *Noise Reduction* pada Model *Whisper*

Penelitian ketiga dilakukan untuk mengevaluasi pengaruh teknik reduksi kebisingan terhadap akurasi sistem transkripsi suara menggunakan model *Whisper*, sebuah sistem modern berbasis *deep learning* dari *OpenAI*. Tidak seperti dua penelitian sebelumnya yang menggunakan pendekatan HMM klasikal, penelitian ini menggunakan model neural besar yang sudah dilatih pada berbagai bahasa dan skenario kebisingan, sehingga memungkinkan analisis yang lebih komprehensif.

Tujuan dari penelitian ini adalah untuk mengetahui apakah metode *Spectral Gating, Wiener Filtering*, dan Kombinasi keduanya dapat meningkatkan kualitas transkripsi pada data berbahasa Lampung, dengan evaluasi berbasis metrik *Word Error Rate* (WER) dan *Character Error Rate* (CER).

3.5.1 Persiapan Data Uji

Tahap awal dari penelitian ketiga dimulai dengan menyusun dataset pengujian yang terdiri dari lima kalimat baru dalam bahasa Lampung dialek Api. Kalimat-kalimat ini dirancang secara manual dan tidak termasuk dalam data pelatihan sebelumnya, sehingga dapat digunakan untuk menguji kemampuan generalisasi sistem transkripsi suara modern dalam mengenali ucapan yang tidak dilatih secara eksplisit.

Setiap kalimat kemudian diproses ke dalam empat versi audio yang berbeda untuk mengamati dampak masing-masing metode noise reduction terhadap kualitas hasil transkripsi. Empat versi tersebut adalah:

1. Versi *Original* (Tanpa Reduksi Kebisingan) : File audio asli sebagaimana direkam, tanpa dilakukan proses pembersihan suara.

2. Versi *Spectral Gating* : File audio diproses menggunakan metode Spectral Gating yang bekerja dengan menekan komponen noise berdasarkan analisis spektrum frekuensi.
3. Versi *Wiener Filtering* : File audio diproses menggunakan filter Wiener, yaitu pendekatan estimasi statistik untuk meredam noise.
4. Versi *Combined* (*Spectral + Wiener*) : Kombinasi dari dua teknik di atas. Pertama dilakukan *Spectral Gating*, kemudian hasilnya diproses kembali menggunakan *Wiener Filtering*.

Dengan demikian, jumlah total file audio uji yang dihasilkan adalah 5 kalimat × 4 versi = 20 file audio. Semua file ini disiapkan dalam format standar (seperti .wav) dengan spesifikasi sampling rate dan bit depth yang seragam, agar dapat diproses langsung oleh sistem *Whisper* pada tahap berikutnya.

3.5.2 Evaluasi dengan *Whisper*

Setelah proses *noise reduction* selesai, langkah berikutnya dalam alur penelitian ketiga adalah mengevaluasi hasil transkripsi audio menggunakan model *Whisper* dari *OpenAI*. *Whisper* merupakan model berbasis neural yang telah dilatih secara multibahasa dan multitugas untuk mengenali dan mentranskripsi suara menjadi teks secara otomatis. Langkah-langkah Evaluasi :

1. Transkripsi Otomatis Menggunakan Model *Whisper* : Semua file audio hasil *noise reduction* (*Original, Spectral Gating, Wiener Filtering*, dan Kombinasi) untuk lima kalimat uji dimasukkan ke dalam sistem *Whisper*. Model ini secara otomatis melakukan proses *decoding* sinyal suara menjadi teks prediksi berdasarkan kemiripan akustik dan struktur linguistik yang telah dipelajarinya.
2. Perbandingan Transkripsi (Prediksi vs *Ground Truth*) : Hasil transkripsi dari model *Whisper* kemudian dibandingkan dengan

transkripsi referensi (*ground truth*) yang telah disiapkan sebelumnya.

3.6 Evaluasi Sistem

Evaluasi dilakukan dengan dua metrik utama yaitu *Word Error Rate* (WER) untuk mengukur persentase kesalahan pada tingkat kata dan *Character Error Rate* (CER) untuk mengukur persentase kesalahan pada tingkat karakter. Evaluasi dilakukan pada semua eksperimen (penelitian 1, 2, dan 3) untuk memperoleh gambaran akurasi dan kemampuan generalisasi sistem transkripsi.

3.6.1 *Word Error Rate* (WER)

Evaluasi kinerja model dilakukan dengan mengukur *Word Error Rate* (WER). Dataset uji yang telah ditranskripsi oleh sistem dibandingkan dengan teks referensi, yang merupakan transkripsi manual dari rekaman. Pengukuran WER digunakan untuk menentukan tingkat akurasi sistem dalam mengenali kata-kata dari dataset. Formula untuk menghitung WER disajikan pada sub bagian 2.5.1. WER pada rekaman sebelum dan setelah reduksi kebisingan dibandingkan untuk mengevaluasi dampak metode *Spectral Gating* dan *Wiener Filtering* terhadap akurasi sistem.

3.6.2 *Character Error Rate* (CER)

Selain WER, penelitian ini juga menggunakan *Character Error Rate* (CER) untuk mengevaluasi kesalahan pada tingkat karakter. CER lebih relevan dalam mendeteksi kesalahan kecil, seperti penghilangan huruf atau perubahan ejaan, yang sering terjadi dalam Bahasa Lampung dengan struktur kata yang pendek. Tingkat CER dihitung berdasarkan formula yang disajikan pada sub bagian 2.5.2. CER dihitung dengan membandingkan hasil transkripsi dengan teks referensi pada tingkat karakter. Data hasil evaluasi CER memberikan analisis tambahan terhadap kinerja sistem, khususnya dalam mengenali detail-detail kecil dalam dataset.

V. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini bertujuan untuk mengevaluasi performa sistem *speech-to-text* (STT) pada audio berbahasa Lampung, dengan fokus utama pada dampak penggunaan metode reduksi kebisingan terhadap hasil transkripsi. Penelitian dilakukan dalam tiga tahapan eksperimen utama, masing-masing dengan pendekatan berbeda: model berbasis HMM-MFCC, evaluasi generalisasi, dan pendekatan modern menggunakan *Whisper*. Berdasarkan hasil yang diperoleh, berikut kesimpulan yang dapat diambil:

1. Pengaruh terhadap Signal-to-Noise Ratio (SNR):
    - Spectral Gating memberikan peningkatan nilai SNR yang paling signifikan dibandingkan dengan Wiener Filtering dan kombinasi keduanya.
    - Metode Wiener Filtering menunjukkan perbaikan SNR namun tidak sebaik Spectral Gating.
    - Sementara itu, penerapan kombinasi dua metode (Spectral Gating + Wiener Filtering) justru menghasilkan penurunan kualitas sinyal jika dilihat dari nilai SNR akhir, yang kemungkinan disebabkan oleh hilangnya informasi penting akibat proses filtrasi ganda.
    - Hal ini menegaskan bahwa dalam konteks rekaman audio bahasa Lampung, penggunaan metode tunggal Spectral Gating lebih optimal untuk menjaga keseimbangan antara pengurangan noise dan informasi sinyal.
2. Penelitian Pertama (HMM-MFCC – Data Kalimat Sederhana):
    - Sistem STT berbasis HMM dan fitur MFCC dapat mengenali kalimat yang identik antara data pelatihan dan pengujian.

- Nilai *Word Error Rate* (WER) dan *Character Error Rate* (CER) yang diperoleh adalah 0.0%, namun ini disebabkan oleh *overfitting* karena data pelatihan dan pengujian berasal dari sumber yang sama.

3. Penelitian Kedua (Evaluasi Generalisasi Model):
   - Ketika diuji dengan kalimat baru (berbasis kata yang sudah pernah dilatih), sistem gagal memberikan hasil transkripsi yang sesuai secara semantik.
   - Model HMM tidak memiliki kemampuan untuk membentuk struktur kalimat dari unit kata, sehingga menghasilkan WER 86,36% dan CER 70,23% pada pendekatan unit kata, menunjukkan ketidakmampuan generalisasi dari model berbasis HMM klasik.
4. Penelitian Ketiga (Evaluasi dengan *Whisper*):
   - *Whisper* sebagai model STT berbasis neural menunjukkan kemampuan yang jauh lebih baik dalam mengenali kalimat baru, meskipun masih ada kesalahan.
   - *Spectral Gating* terbukti menjadi metode reduksi kebisingan yang paling efektif dibandingkan dengan *Wiener* dan kombinasi keduanya.
   - Rata-rata WER dan CER untuk audio hasil *Spectral Gating* adalah yang paling rendah, mengindikasikan bahwa metode ini dapat mempertahankan informasi penting dari sinyal asli sambil mengurangi *noise*.
5. Efektivitas Noise Reduction:
   - Noise *reduction* dengan *Spectral Gating* terbukti meningkatkan akurasi transkripsi pada sistem *Whisper*.
   - Penggunaan dua kali filtering (kombinasi *Spectral* + *Wiener*) menurunkan akurasi karena kemungkinan kehilangan informasi sinyal yang penting.

Tabel 10. Ringkasan Tahapan Penelitian dan Hasil Penelitian

No.	Tahap Penelitian	Tujuan	Metode	Hasil Utama	Kesimpulan
1	**Penelitian Pertama** (HMM-MFCC, Data Identik)	Membangun dan menguji sistem STT berbasis HMM-MFCC dengan data latih dan uji yang identik, untuk menilai baseline kinerja dan mendeteksi tanda overfitting.	**Data Latih & Uji:** 4 file kalimat awal × 4 versi noise (Original, Spectral Gating, Wiener, Combined) = 16 file; data uji sama = 16 file. • Ekstraksi fitur MFCC • Latih 1 HMM per kalimat (n_components=5, covariance_type='diag', n_iter=100) • Decode data uji yang identik memakai HMM masing-masing. • Evaluasi: WER & CER per kondisi.	Semua file uji didekode sebagai teks kalimat pelatihan identik → • **WER = 0 %**, **CER = 0 %** (lolos sempurna secara angka).- Namun karena data uji = data latih, model hanya "menghafal" pola, bukan "mengenali."- Indikasi **overfitting**: akurasi artifisial tinggi tetapi tak merepresentasikan kemampuan generalisasi.	HMM-MFCC dengan data identik menunjukkan overfitting (zer WER/CER palsu). Model "hafal" kalimat latih, sehingga tidak valid untuk menilai kemampuan kenali kalimat baru.
2	**Penelitian Kedua** (Model Unit Kalimat)	Menguji generalisasi HMM berbasis "kalimat utuh" pada 5 kalimat baru (belum pernah dilatih) dengan empat kondisi noise, untuk melihat apakah model kalimat dapat mengenali susunan baru.	**Data Latih:** 4 kalimat awal × 4 versi noise = 16 file. • Ekstraksi MFCC • Latih 1 HMM per kalimat (n_components=5, covariance_type='diag', n_iter=100). **Data Uji:** 5 kalimat baru × 4 versi noise = 20 file. • Setiap file uji diekstrak MFCC → decode oleh ke-4 HMM kalimat. • Evaluasi: WER	Semua 20 file uji "dipetakan" kembali ke kalimat pelatihan ("Seminggu sai likut…") → • **WER ≈ 8,623 %**, **CER ≈ 8,388 %** (rata-rata untuk Original, Spectral, Wiener, Combined).- Angka WER/CER terlihat rendah, tetapi secara semantik model **tidak** mengenali kalimat baru: hasil hanya "mengulang" teks latih.- Menandakan	Model HMM "kalimat utuh" mengalami overfitting: meski WER/CER rendah, model hanya mengulang pola latih. Pendekatan HMM berbasis kalimat tidak mampu mengenali susunan kalimat baru yang tidak identik.

No.	Tahap Penelitian	Tujuan	Metode	Hasil Utama	Kesimpulan
			& CER per kondisi (Original, Spectral, Wiener, Combined).	**overfitting** pada level kalimat: HMM per kalimat gagal mempelajari komponen kata terpisah.	
3	**Penelitian Kedua** (Model Unit Kata)	Mengevaluasi HMM per-kata dalam mengenali susunan kata baru (5 kalimat uji) tanpa language model.	**Unit Kata:** 4 kalimat awal → potong per kata → 191 unit kata unik. • Setiap unit kata diproses ke 4 versi noise (Original, Spectral, Wiener, Combined) → 764 file MFCC. • Latih 1 HMM per unit kata (n_components=5, covariance_type='diag', n_iter=100) = 191 HMM.	**191 HMM berhasil terlatih** tanpa error.- **SNR (per unit kata): • Spectral Gating**: peningkatan SNR tertinggi (~+10–12 dB vs Original). • **Wiener Filtering**: SNR bervariasi (efektif pada noise stasioner, kurang pada noise non-stasioner). • **Combined**: SNR lebih rendah daripada Spectral saja (karena kehilangan frekuensi penting).- Fondasi HMM per-kata siap dipakai untuk Eksperimen 2 (Pengujian Kalimat Baru).	HMM per-kata berhasil dibangun, dengan Spectral Gating sebagai metode noise reduction terbaik untuk unit kata. Menyiapkan model kata-per-kata yang siap dievaluasi pada susunan kalimat.
			Data Uji: 5 kalimat baru × 4 versi noise = 20 file. • Ekstraksi MFCC per segmen pendek. • Setiap segmen diuji ke 191 HMM unit kata → pilih kata dengan log likelihood tertinggi.	**Contoh Test_1_original: • Ground Truth:** "pagi itu hulun tuhaku… rencana liburan" • **Prediksi:** "malang mit khki sai sikam suma lapah liburan sai	HMM per-kata tanpa konteks linguistik gagal menyusun kalimat bermakna. Meskipun setiap kata terlatih, tanpa language

No.	Tahap Penelitian	Tujuan	Metode	Hasil Utama	Kesimpulan
			• Susun urutan kata menjadi kalimat utuh tanpa model bahasa (tanpa N-gram/LSTM). • Evaluasi: WER & CER per file terhadap ground truth.	liburan" • **WER = 86,36 %**, **CER = 70,23 %- Rata-rata 20 file:** WER ≥ 75 %, CER ≥ 60 % → performa sangat buruk.- Prediksi cenderung "acak" dan tidak menyerupai struktur kalimat, sering kali hanya potongan suara tanpa makna.	model urutan kata menjadi acak → WER & CER sangat tinggi. Menunjukkan keterbatasan HMM klasik untuk generalisasi kalimat baru.
4	**Penelitian Ketiga** (Model Whisper)	Mengukur performa Whisper pada 5 kalimat baru dengan variasi noise, serta menentukan metode noise reduction (Spectral, Wiener, Combined) yang paling efektif.	**Data Uji:** 5 kalimat baru × 4 versi noise = 20 file. • Model: Whisper (end-to-end Transformer, pretrained multibahasa). • Input setiap file ke Whisper → hasil transkripsi otomatis. • Evaluasi: Hitung WER & CER untuk kondisi Original, Spectral, Wiener, Combined menggunakan pustaka jiwer.	**Test_4_spectral** (**kasus terbaik**)**: • WER = 36,84 %**, **CER = 13,11 %** (vs Original: WER 47,37 %, CER 16,39 %; Wiener: WER 47,37 %, CER 21,31 %; Combined: WER 52,63 %, CER 18,85 %). • • Rata-rata Whisper (20 file): WER 36,84 %–80,95 %, CER 11,03 %–27,34 %. • • Spectral Gating: paling konsisten menurunkan WER & CER. • Wiener Filtering: hanya efektif pada noise stasioner; pada noise non-stasioner performa stagnan atau memburuk.	Whisper mengenali kalimat baru jauh lebih baik dibanding HMM. **Spectral Gating** adalah metode noise reduction paling efektif bagi Whisper; **Wiener** kurang efektif pada noise non-stasioner; **Combined** menurunkan akurasi. Rekomendasi: gunakan **Whisper + Spectral Gating** untuk STT bahasa Lampung.

No.	Tahap Penelitian	Tujuan	Metode	Hasil Utama	Kesimpulan
				• Combined Filtering: cenderung paling buruk karena hilangnya informasi penting dua kali.	

5.2 Saran

Berdasarkan hasil dan keterbatasan dari penelitian ini, beberapa saran dapat diberikan untuk pengembangan sistem STT berbahasa daerah ke depannya:

1. Penggunaan Model Berbasis Neural :
   Penggunaan model seperti *Whisper* atau wav2vec sangat disarankan untuk pengenalan suara bahasa daerah karena mampu memahami konteks dan struktur kalimat, berbeda dengan HMM yang bersifat statis dan sekuensial.
2. Perluasan Dataset :
   Dataset yang digunakan masih sangat terbatas. Untuk pelatihan sistem STT yang akurat dan general, perlu dikembangkan dataset yang lebih besar, mencakup variasi aksen, kecepatan bicara, dan konteks percakapan.
3. Evaluasi Multi-Level :
   Evaluasi sistem STT umumnya menggunakan dua metrik utama: Word Error Rate (WER) dan Character Error Rate (CER). Meskipun keduanya efektif dalam mengukur kesalahan transkripsi, pendekatan evaluasi ini masih terbatas karena tidak mempertimbangkan aspek fonetik atau semantik dari bahasa. Oleh karena itu, perlu diterapkan metode evaluasi tambahan seperti:
   - Phoneme Error Rate (PER) : Mengukur kesalahan pada tingkat fonem, yang sangat berguna terutama dalam bahasa daerah seperti Lampung yang memiliki fonem khas yang berbeda dengan bahasa Indonesia standar.
   - Semantic Similarity Metrics : Misalnya menggunakan *BERTScore* atau *BLEU* untuk menilai seberapa dekat makna kalimat hasil transkripsi terhadap kalimat referensi. Ini penting karena dalam komunikasi lisan, makna sering kali lebih penting daripada kata-kata yang tepat secara literal.
   - Manual Evaluation : Libatkan penutur asli bahasa daerah untuk memberikan penilaian subjektif terhadap hasil transkripsi dalam hal kelayakan dan keterpahaman.

   Evaluasi multi-level ini akan memberikan gambaran yang lebih komprehensif terhadap kinerja sistem, terutama dalam konteks penggunaan di dunia nyata.

4. Eksplorasi Metode NR Lain:

   Metode *noise reduction* lain seperti RNNoise atau Denoising Autoencoder dapat dieksplorasi untuk menghasilkan sinyal yang lebih bersih tanpa mengorbankan akurasi.

   Dalam penelitian ini digunakan metode *Spectral Gating* dan *Wiener Filtering*. Kedua teknik ini bersifat tradisional dan memiliki keterbatasan dalam menangani noise yang tidak stasioner. Oleh karena itu, saran pengembangan selanjutnya adalah mengeksplorasi pendekatan berbasis deep learning yang lebih adaptif, seperti:

   - RNNoise : Sebuah model ringan berbasis Recurrent Neural Network (RNN) yang dikembangkan oleh Xiph.Org. RNNoise mampu secara efisien membedakan antara sinyal suara dan noise, bahkan dalam kondisi kebisingan yang berat, dan cocok untuk implementasi real-time.
   - Denoising Autoencoder (DAE) : Merupakan jaringan saraf yang dilatih secara khusus untuk merekonstruksi sinyal bersih dari sinyal yang terkontaminasi noise. DAE dapat menangkap pola noise kompleks dan menghasilkan pemulihan sinyal suara yang lebih akurat.
   - Spectral Subtraction berbasis Neural Networks : Kombinasi teknik klasik dengan prediksi berbasis neural untuk memperbaiki kelemahan dalam pengurangan spektrum tradisional.

   Eksplorasi metode ini tidak hanya akan meningkatkan kualitas sinyal audio yang masuk ke sistem STT, tetapi juga dapat memperbaiki performa transkripsi secara keseluruhan.

5. Optimasi Parameter Noise Reduction

   Untuk menjamin bahwa proses noise reduction (NR) tidak menurunkan kualitas sinyal dan meminimalkan artefak, penelitian selanjutnya sebaiknya menerapkan langkah-langkah berikut :

   - Threshold Adaptif : Tentukan ambang deteksi noise secara dinamis berdasarkan kondisi lingkungan rekaman. Dengan cara ini, filter NR dapat menyesuaikan diri agar tidak “over-filtering” (menghapus komponen sinyal berguna) maupun “under-filtering” (meninggalkan sisa noise).

- Pengaturan Intensitas Pengurangan : Sesuaikan tingkat agresivitas pengurangan noise (over-subtraction factor) sehingga tidak menimbulkan artefak “robotik” akibat pengurangan berlebih, atau sebaliknya, tidak menyisakan noise berlebih ketika terlalu konservatif. Penyelarasan parameter ini penting untuk menjaga keseimbangan antara kebersihan sinyal dan naturalitas suara.
- Penerapan Batas Minimum Spektrum : Terapkan spectral floor agar magnitudo frekuensi tidak pernah jatuh di bawah nilai minimum tertentu. Langkah ini mencegah “penggundulan” spektrum yang menyebabkan suara menjadi teredam atau “mengendap,” sehingga karakter asli sinyal tetap terjaga.
- Post-Filtering dan Smoothing : Lakukan perataan cepat (moving average) pada magnitudo spektra dan windowing temporal untuk meredam lonjakan artefak mendadak, seperti klik atau dentuman. Studi lebih lanjut dapat mengeksplorasi filter median 2D atau teknik smoothing lanjutan untuk hasil yang lebih halus.
- Iterasi dan Evaluasi Berkelanjutan : Kombinasikan evaluasi objektif—misalnya WER/CER untuk akurasi transkripsi dan PESQ/POLQA untuk kualitas audio—dengan uji pendengaran subjektif oleh penutur asli. Setiap putaran evaluasi digunakan untuk menyetel ulang parameter NR hingga diperoleh konfigurasi optimal yang meminimalkan artefak tanpa mengorbankan performa STT.

Dengan mengikuti pendekatan terstruktur ini, penelitian dapat menghasilkan metodologi noise reduction yang adaptif, efektif, dan terukur, sekaligus memastikan keseimbangan optimal antara kebersihan sinyal dan kealamian kualitas audio.

DAFTAR PUSTAKA

Abdelli, O., & Merazka, F. (2024). Deep learning for speech denoising with improved Wiener approach. *International Journal of Speech Technology*, vol. 27, no. 4, pp. 997–1012.

Abidin, A. (2017). Dialek Api dan Dialek Nyo pada Bahasa Lampung. *Jurnal Bahasa dan Sastra Daerah*, vol. 3, no. 2, pp. 45–53.

Abidin, A., Dwijayanti, S., & Fadillah, M. F. (2020). Speech-to-text conversion in Indonesian language using a deep bidirectional LSTM. *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 5, pp. 1–7.

Agrawal, J., Gupta, M., & Garg, H. (2023). Performance analysis of speech enhancement using spectral gating with U-Net. *Journal of Electrical Engineering*, vol. 74, no. 5, pp. 365–373.

Al-Omari, A. A., & Al-Ani, M. A. (2020). Arabic speech recognition system based on MFCC and HMMs. *Journal of Computer and Communications*, vol. 8, no. 3, pp. 28–34.

Butarbutar, M., Sachio, K., & Nugroho, M. (2023). Adaptive Wiener filtering method for noise reduction in speech recognition system. *TechRxiv*. https://doi.org/10.36227/techrxiv.23608602.v1

Dewi, I. P. A. K. S., Sari, I. G. A. A. S. W., & Putra, I. M. Y. S. (2023). Implementasi aplikasi penerjemah multi bahasa berbasis Python menggunakan Google Translate API. *Jurnal Ilmu Komputer*, vol. 13, no. 2, pp. 45–52.

Fakhrurozi, F., Putri, S. N., & Nugroho, A. S. (2019). Kearifan lokal dalam Bahasa Lampung: Kajian sosial dan budaya. *Jurnal Ilmu Sosial dan Budaya*, vol. 7, no. 1, pp. 101–115.

Fujimoto, M., & Kawai, H. (2016). Noise robust speech recognition using noise-adaptive hidden Markov models. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 4, pp. 796–806.

Gales, M. J. F., & Young, S. J. (2008). The application of hidden Markov models in speech recognition. *Foundations and Trends in Signal Processing*, vol. 1, no. 3, pp. 195–304.

Gannot, S., Vincent, E., Markovich-Golan, S., & Ozerov, A. (2017). A consolidated perspective on multimicrophone speech enhancement and source separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 4, pp. 692–730.

Graves, A., Mohamed, A.-R., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6645–6649.

Han, K., Wang, Y., Wang, D., Woods, W. S., & Merks, I. (2015). Learning spectral mapping for speech dereverberation and denoising. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 6, pp. 982–992.

Jurafsky, D., & Martin, J. H. (2025). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (3rd ed., draft). Retrieved from https://web.stanford.edu/~jurafsky/slp3/

Karita, S., Sproat, R., & Ishikawa, H. (2023). Lenient evaluation of Japanese speech recognition: Modeling naturally occurring spelling inconsistency. In *Proceedings of the Workshop on Computation and Written Language (CAWL 2023)*, pp. 61–70.

Kumar, M. A., & Chari, K. M. (2020). Noise reduction using modified Wiener filter in digital hearing aid for speech signal enhancement. *Journal of Intelligent Systems*, vol. 29, no. 1, pp. 1360–1378.

Maseri, M., & Mamat, M. (2020). Performance analysis of implemented MFCC and HMM-based speech recognition system. In *Proceedings of the 2020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, pp. 1–6.

Naik, S. S., Bhatikar, G., & Gaude, U. (2021). Analysis of best algorithm for noise reduction in podcasting. *International Journal of Scientific Research in Science and Technology*, vol. 8, no. 3, pp. 246–250.

Neumann, T. von, Boeddeker, C., Kinoshita, K., Delcroix, M., & Haeb-Umbach, R. (2023). On word error rate definitions and their efficient computation for

multi-speaker speech recognition systems. In *Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 16112–16120.

Ramadhina, D. S., Magdalena, R., & Saidah, S. (2020). Individual identification through voice using MFCC and HMM. *Journal of Measurements Electronics Communications and Systems*, vol. 7, no. 1, pp. 26–31.

Roweis, S. T. (2001). One microphone source separation. In *Advances in Neural Information Processing Systems*, vol. 13, pp. 793–799.

Shah, A., & Shah, B. (2023). Speech denoising using spectral gating for noise robust applications. *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 31, no. 3, pp. 11–20.

Sharma, B. K., & Raj, A. (2019). Robust speech recognition using noise adaptive features. In *Proceedings of Interspeech 2019*, pp. 3829–3833.

Smith, R. D., Zhang, L., & Wang, H. (2019). Spectral gating with non-stationary noise modeling for robust ASR. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6440–6444.

Suryadharma, I. K., Budiman, G., & Irawan, B. (2014). Perancangan aplikasi speech to text Bahasa Inggris ke Bahasa Bali menggunakan PocketSphinx berbasis Android. *E-Proceeding of Engineering*, vol. 1, no. 1, pp. 229–237.

Wang, Z., Zhang, Y., & Xie, L. (2023). Text-only domain adaptation with phoneme-guided data splicing for end-to-end ASR. In *Proceedings of Interspeech 2023*, pp. 1234–1238.

Widiyanto, E., Endah, S. N., Adhy, S., & Sutikno. (2014). Aplikasi speech to text berbahasa Indonesia menggunakan Mel-Frequency Cepstral Coefficient dan Hidden Markov Model. In *Prosiding Seminar Nasional Ilmu Komputer*, pp. 39–44. Universitas Diponegoro.

Xu, Y., Du, J., Dai, L. R., & Lee, C. H. (2015). A regression approach to speech enhancement based on deep neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 1, pp. 7–19.