ANALISIS HASIL TRANSKRIPSI SUARA KE TEKS BAHASA LAMPUNG DENGAN EKSTRAKSI FITUR *PERCEPTUAL LINEAR PREDICTION* (PLP) DAN METODE *HIDDEN MARKOV MODEL* (HMM)
(Studi Kasus : Lampung Pepadun Dialek A)

(Tesis)

Oleh

MOHAMAD ZAINUDIN
2327051003

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

ANALISIS HASIL TRANSKRIPSI SUARA KE TEKS BAHASA LAMPUNG DENGAN EKSTRAKSI FITUR *PERCEPTUAL LINEAR PREDICTION* (PLP) DAN METODE *HIDDEN MARKOV MODEL* (HMM)
(Studi Kasus : Lampung Pepadun Dialek A)

(Tesis)

Oleh

MOHAMAD ZAINUDIN
2327051003

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

ANALISIS HASIL TRANSKRIPSI SUARA KE TEKS BAHASA LAMPUNG DENGAN EKSTRAKSI FITUR *PERCEPTUAL LINEAR PREDICTION* (PLP) DAN METODE *HIDDEN MARKOV MODEL* (HMM)
(Studi Kasus : Lampung Pepadun Dialek A)

Oleh

MOHAMAD ZAINUDIN

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

ANALISIS HASIL TRANSKRIPSI SUARA KE TEKS BAHASA LAMPUNG DENGAN EKSTRAKSI FITUR *PERCEPTUAL LINEAR PREDICTION* (PLP) DAN METODE *HIDDEN MARKOV MODEL* (HMM) (Studi Kasus : Lampung Pepadun Dialek A)

Oleh

MOHAMAD ZAINUDIN

Penelitian ini mengkaji kinerja sistem transkripsi suara ke teks pada Bahasa Lampung dialek Pepadun A dengan menerapkan metode ekstraksi fitur *Perceptual Linear Prediction* (PLP) dan pemodelan *Hidden Markov Model* (HMM). Latar belakang penelitian ini adalah melestarikan Bahasa Lampung sebagai warisan budaya lokal melalui teknologi pengenalan suara berbasis bahasa daerah. Sebanyak 800 kalimat dari empat penutur asli diolah melalui tahapan *resampling*, augmentasi, ekstraksi fitur PLP, serta pelatihan model HMM. Evaluasi dilakukan dengan *Word Error Rate* (WER) dan *Character Error Rate* (CER). Hasil menunjukkan bahwa kombinasi PLP dan HMM mampu merepresentasikan ciri akustik Bahasa Lampung secara efektif meskipun dengan keterbatasan data. Penelitian ini memberikan kontribusi terhadap pengembangan sistem pengenalan suara berbasis bahasa daerah dan mendukung upaya pelestarian bahasa melalui pendekatan komputasional.

Kata Kunci: Bahasa Lampung Dialek Pepadun, *Perceptual Linear Prediction* (PLP), *Hidden Markov Model* (HM), *Speech Recognition*.

ABSTRACT

ANALYSIS OF SPEECH-TO-TEXT TRANSCRIPTION IN THE LAMPUNG LANGUAGE USING PERCEPTUAL LINEAR PREDICTION (PLP) FEATURE EXTRACTION AND HIDDEN MARKOV MODEL (HMM) METHOD (Case Study: Lampung Pepadun Dialect A)

Oleh

MOHAMAD ZAINUDIN

This study evaluates the performance of a speech-to-text transcription system for the Lampung language in the Pepadun A dialect by applying *Perceptual Linear Prediction* (PLP) for feature extraction and the *Hidden Markov Model* (HMM) for acoustic modeling. The background of this research lies in the effort to preserve the Lampung language as a cultural heritage through local language-based speech recognition technology. A total of 800 sentences from four native speakers were processed through resampling, data augmentation, PLP feature extraction, and HMM training. System performance was evaluated using *Word Error Rate* (WER) and *Character Error Rate* (CER). The results demonstrate that the combination of PLP and HMM effectively represents the acoustic characteristics of the Lampung language, even with limited data. This research contributes to the development of regional language-based speech recognition systems and supports language preservation through computational approaches.

Keywords: Lampung Language, Pepadun Dialect, Perceptual Linear Prediction (PLP), Hidden Markov Model (HMM), Speech Recognition.

Judul Tesis : **ANALISIS HASIL TRANSKRIPSI SUARA KE TEKS BAHASA LAMPUNG DENGAN EKSTRAKSI FITUR *PERCEPTUAL LINEAR PREDICTION* (PLP) DAN METODE *HIDDEN MARKOV MODEL* (HMM) (Studi Kasus : Lampung Pepadun Dialek A)**

Nama Mahasiswa : Mohamad Zainudin

Nomor Pokok Mahasiswa : 2327051003

Program Studi : S2 Ilmu Komputer

Jurusan : Ilmu Komputer

Fakultas : Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

Komisi Pembimbing

Dr. rer. nat. Akmal Junaidi, M.Sc. NIP 19710129 199702 1 001	**Favorisen R. Lumbanraja, Ph.D.** NIP 19830110 200812 1 002

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

MENGESAHKAN

1. Tim Penguji

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

Sekretaris : **Favorisen R. Lumbanraja, Ph.D.** .....

Penguji I
Penguji Utama : **Dr. Aristoteles, S.Si., M.Si.** .....

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 19640326 198902 1 001

Tanggal Lulus Ujian Tesis : **28 Mei 2025**

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Mohamad Zainudin

NPM : 2327051003

Dengan ini menyatakan bahwa tesis saya yang berjudul **"Analisis Hasil Transkripsi Suara Ke Teks Bahasa Lampung Dengan Ekstraksi Fitur *Perceptual Linear Prediction* (PLP) Dan Metode *Hidden Markov Model* (HMM) (Studi Kasus : Lampung Pepadun Dialek A)"** 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, 28 Mei 2025

METERAI TEMPEL

A0F38AMX352700415

Mohamad Zainudin

NPM. 2327051003

RIWAYAT HIDUP

Penulis dilahirkan di Pringsewu, pada tanggal 05 Agustus 1991, sebagai anak ketiga dari tiga bersaudara.

Penulis menyelesaikan pendidikan formal di SDN 7 Gading Rejo dan selesai pada tahun 2003. Kemudian pendidikan menengah pertama di SMP Muhammadiyah 1 Gading Rejo yang diselesaikan pada tahun 2006, lalu melanjutkan ke pendidikan menengah kejuruan di SMKN 1 Gading Rejo yang diselesaikan pada tahun 2009.

Pada tahun 2009, penulis terdaftar sebagai mahasiswa Jurusan Matematika Program Studi D3 Manajemen Informatika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur ujian tertulis mandiri dan lulus pada Juni 2012. Penulis melanjutkan pendidikan kejenjang selanjutnya yaitu Sarjana Sistem Informasi pada tahun 2020 di Fakultas Komputer Universitas Mitra Indonesia dan lulus pada Agustus 2022. Penulis melanjutkan pendidikan kejenjang selanjutnya yaitu Magister Ilmu Komputer pada tahun 2023 di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung. Selama menjadi mahasiswa, penulis melakukan beberapa kegiatan antara lain.

1. Menjadi *Oral Presenter* pada *Conference on Applied Science, Mathematics, and Informatics* (ICASMI) 2024.
2. Pada tahun 2024, penulis dinyatakan lolos sebagai peserta Uji Kompetensi BNSP untuk skema *Junior Web Developer*.

MOTO

1. “Sesungguhnya bersama kesulitan ada kemudahan.”
(QS. Al-Insyirah: 6)

2. “Sesungguhnya Allah SWT tidak mengubah keadaan suatu kaum sebelum mereka mengubah keadaan diri mereka sendiri. Dan apabila Allah SWT menghendaki keburukan terhadap suatu kaum, maka tak ada yang dapat menolaknya dan tidak ada pelindung bagi mereka selain Dia.”
(QS. Ar-Ra’d: 11)

3. “Strength doesn’t come from what you can do. It comes from overcoming the things you once thought you couldn’t.”
(Rikki Rogers)

4. “Hanya nelayan tangguh yang mampu menaklukkan tingginya ombak.”
(Anonim)

5. “Tak ada pencapaian besar tanpa usaha yang besar.”
(Anonim)

6. “It's not about who is the fastest, but about who never stops that will eventually get there.”
(Anonim)

PERSEMBAHAN

Alhamdulillahirobbilalamin

Segala puji dan syukur saya panjatkan ke hadirat Allah SWT atas rahmat dan karunia-Nya, sehingga tesis ini dapat diselesaikan. Shalawat dan salam semoga tercurah kepada Nabi Muhammad SAW.

Orang Tuaku Tercinta

Dalam setiap langkahku, ada doa kalian yang menguatkan.

Dalam setiap pencapaianku, ada cinta dan pengorbanan yang tak terbalas.

Terima kasih, atas segalanya.

Keluarga Tercinta

Istriku tercinta Suri Gelsi Oktasia, S.Pd., serta anak-anakku Ilona Galeckha Athaleta Zamzain dan Izora Jennamira Bricia Zamzain terima kasih telah menjadi sumber inspirasi, kekuatan, dan kebahagiaan dalam setiap langkah dan perjuanganku. Kehadiran kalian adalah anugerah terbesar dalam hidup ini.

Seluruh Keluarga Besar Magister Ilmu Komputer 2023

Yang selalu memberikan semangat dan dukungan.

Almamater Tercinta, Universitas Lampung dan Jurusan Ilmu Komputer

Tempat bernaung mengemban semua ilmu untuk menjadi bekal hidup.

SANWACANA

Puji syukur kehadirat Allah SWT atas berkah, rahmat dan hidayat-Nya, serta petunjuk dan pedoman dari Rasulullah Nabi Muhammad Sholallahu Alaihi Wasallam penulis dapat menyelesaikan tesis yang berjudul "Analisis Hasil Transkripsi Suara Ke Teks Bahasa Lampung Dengan Ekstraksi Fitur *Perceptual Linear Prediction* (PLP) Dan Metode *Hidden Markov Model* (HMM) (Studi Kasus : Lampung Pepadun Dialek A)" dengan baik dan lancar.

Terima kasih penulis ucapkan kepada semua pihak yang telah membantu dan berperan besar dalam menyusun tesis ini, antara lain.

1. Kedua orang tua tercinta atas segala doa, dukungan, semangat, dan kasih sayang yang tiada henti. Semoga Allah SWT senantiasa melimpahkan kebahagiaan dan keberkahan dalam hidup kalian di 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., sebagai pembimbing utama yang telah memberikan dukungan, arahan, ide, motivasi, kritik dan saran kepada penulis sehingga dapat menyelesaikan tesis ini dengan baik.
6. Bapak Favorisen R. Lumbanraja, Ph.D., sebagai pembimbing kedua yang telah meluangkan waktu untuk membimbing, memberikan ide, arahan, kritik serta saran untuk dapat menyelesaikan tesis ini dengan baik.
7. Bapak Dr. Aristoteles, S.Si., M.Si., sebagai penguji utama yang telah memberikan masukan yang bermanfaat dalam perbaikan tesis ini.

8. Bapak Tristiyanto, S.Kom., M.I.S., Ph.D., sebagai penguji kedua yang telah memberikan masukan yang bermanfaat dalam perbaikan tesis ini.
9. Bapak Dr. Ahmad Faisol, S.Si., M.Sc., sebagai pembimbing akademik yang telah memberikan bimbingan, arahan, dan dukungan selama masa studi hingga penyusunan tesis ini.
10. Ibu Yunda Heningtyas, S.Kom., M.Kom., selaku Sekretaris Jurusan Ilmu Komputer FMIPA Universitas Lampung.
11. Ibu Ade Nora Maela, Mas Samanto, Mas Nofal yang telah membantu segala urusan administrasi penulis di Jurusan Ilmu Komputer.
12. Bapak dan Ibu Dosen Jurusan Ilmu Komputer FMIPA Universitas Lampung yang telah memberikan ilmu dan pengalaman dalam hidup untuk menjadi lebih baik.
13. 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., Ibu Dian Novitasari, S.Pd., atas kesediaannya menjadi penutur Bahasa Lampung yang telah memberikan dukungan sehingga proses penelitian ini berjalan dengan baik.
14. Keluarga Magister Ilmu Komputer 2023 yang tidak bisa penulis sebut satu persatu. Keluarga kedua penulis, rekan kelompok, rekan diskusi, rekan bercanda, dan telah memberi arti dan warna serta pengalaman tak ternilai semasa duduk di bangku kuliah.

Penulis menyadari bahwa tesis ini masih jauh dari kata sempurna, semoga tesis ini membawa manfaat dan keberkahan bagi semua civitas Ilmu Komputer Universitas Lampung aamiin ya rabbal aalamiin.

Bandar Lampung, 28 Mei 2025

Mohamad Zainudin
NPM. 2327051003

DAFTAR ISI

Halaman

DAFTAR TABEL ..... **iii**

DAFTAR GAMBAR ..... **iv**

DAFTAR KODE PROGRAM ..... **v**

I. PENDAHULUAN **1**

1.1. Latar Belakang 1

1.2. Rumusan Masalah 4

1.3. Batasan Masalah 4

1.4. Tujuan Penelitian 4

1.5. Manfaat Penelitian 4

II. TINJAUAN PUSTAKA **6**

2.1. Penelitian Terdahulu 6

2.2. *Resampling* Audio 8

2.3. *Perceptual Linear Prediction* 11

2.4. *Hidden Markov Model* 19

2.5. Sejarah Bahasa Lampung Dialek Pepadun 26

2.6. Karakteristik Fonetik Bahasa Lampung Dialek Pepadun 29

2.7. Implementasi Transkripsi Suara Bahasa Lampung ke Teks 29

2.8. *Word Error Rate* (WER) 30

2.9. *Character Error Rate* (CER) 31

III. METODOLOGI PENELITIAN **33**

3.1. Tempat Penelitian 33

3.2. Waktu Penelitian 33

3.3. Data Suara 38

3.4. Alat dan Perangkat Lunak 40

3.5. Alur Kerja Penelitian 40

IV. HASIL DAN PEMBAHASAN **44**

4.1. Hasil Penelitian 44
4.2. *Splitting* Data 46
4.3. Data Augmentasi 47
4.4. *Perceptual Linear Prediction* (PLP) 55
4.5. *Hidden Markov Model* (HMM) 64
4.6. *Decoding* 68
4.7. *Ground Truth* 69
4.8. Evaluasi Model 71
4.9. Analisis Hasil Evaluasi WER 73
4.10. Analisis Hasil Evaluasi CER 76
4.11. Tantangan Implementasi 79
V. EVALUASI 84
5.1. Simpulan 84
5.2. Saran 85
DAFTAR PUSTAKA 88

DAFTAR TABEL

Tabel Halaman

1. Peneliti Terdahulu 6
2. Probabilitas Emisi. 25
3. Alur dan Waktu Penelitian..... 37
4. Kalimat Bahasa Indonesia dan Bahasa Lampung. 38
5. Pembagian Data. 47
6. Hasil Perhitungan *Pre-emphasis*. 59
7. Hasil Perhitungan *Framming* dan *Windowing*. 60
8. Hasil Perhitngan *Fast Fourier Transform*. 60
9. Hasil Perhitungan *Bark Scale Filterbank*..... 61
10. Perhitungan Hasil *Equal Loudness Pre-emphasis*. 62
11. Perhitungan Hasil *Logarithmic Compression*. 62
12. Perhitungan Hasil *Linear Prediction Coefficients*. 63
13. Pengujian Model. 65
14. Kalimat *Ground Truth*. 70
15. Analisis Prediksi Hasil Evaluasi WER. 74
16. Perhitungan WER..... 75
17. Analisis Hasil Prediksi Kalimat Evaluasi CER. 77
18. Perhitungan CER..... 78

DAFTAR GAMBAR

Gambar Halaman

1. Audio Sebelum *Resampling*. 10
2. Audio Setelah *Resampling*. 10
3. *Model Hidden Model* (Rabiner, 1989). 20
4. Prediksi Cuaca..... 24
5. Peta Provinsi Lampung (Bappeda Lampung, 2018). 27
6. Alur Penelitian..... 41
7. Audio Sebelum Augmentasi. 48
8. Teknik *TimeStretch*..... 50
9. Teknik *PitchShift*. 51
10. Teknik *AddGaussianNoise*. 52
11. Teknik *Shift*..... 53
12. Teknik *Gain*..... 54
13. Audio Original..... 56
14. Audio Ekstraksi PLP..... 57
15. Grafik *Loss Training* HMM. 67
16. Grafik Evaluasi Model HMM. 69
17. Grafik Hasil Perbandingan WER dan CER. 72
18. Bagan Hasil Penelitian. 81

DAFTAR KODE PROGRAM

Pseudocode Halaman

1. *Library* yang digunakan pada HMM..... 44
2. *Splitting* Data..... 46
3. Augmentasi Data. 49
4. *Feature Extraction* PLP..... 55
5. *Hidden Markov Model*. 64
6. *Decoding*. 68
7. Evaluasi Model..... 72

I. PENDAHULUAN

1.1. Latar Belakang

Bahasa Lampung adalah salah satu bahasa daerah yang ada di Indonesia memiliki nilai budaya, sosial dan sejarah yang sangat tinggi (Andina, 2023). Masyarakat Lampung menggunakan bahasa ini di berbagai tempat di Provinsi Lampung dan beberapa daerah di Sumatera Selatan. Bahasa Lampung memiliki struktur, fonologi gramatikal, serta kosakata yang unik dibandingkan dengan bahasa Indonesia dan bahasa daerah lainnya (Andina, 2023). Keberlanjutan Bahasa Lampung menghadapi tantangan yang sangat signifikan dalam beberapa dekade terakhir. Faktor seperti globalisasi, urbanisasi, dominasi bahasa nasional dan internasional yang mempengaruhi berkurangnya masyarakat yang masih menggunakan Bahasa Lampung terutama di kalangan generasi muda (Putri, 2018).

Pelestarian Bahasa Lampung menjadi semakin mendesak mengingat peranannya yang krusial dalam mempertahankan identitas budaya dan warisan leluhur Masyarakat Lampung (Putri, 2018). Salah satu cara yang dapat ditempuh untuk mendukung pelestarian ini adalah melalui penerapan teknologi. Teknologi memiliki potensi besar untuk mendokumentasikan, mempelajari, dan mempromosikan penggunaan Bahasa Lampung dalam kehidupan sehari-hari. Dalam konteks ini, transkripsi suara ke teks berbasis Ekstraksi *Perceptual Linear Prediction* (PLP) dan *Hidden Markov Model* (HMM) muncul sebagai solusi yang tepat untuk melestarikan Bahasa Lampung menggunakan teknologi (Rismawati, 2022).

Teknologi pengenalan suara (*speech recognition*) telah mengalami perkembangan pesat dalam beberapa dekade terakhir (Ali, 2022). Dalam konteks transkripsi suara ke teks, salah satu teknik yang sering digunakan adalah ekstraksi *Perceptual Linear Prediction* (Hermansky, 1990). *Perceptual Linear Prediction* (PLP) mampu mengekstraksi fitur-fitur penting dari sinyal suara yang relevan untuk pengenalan suara, dengan mempertimbangkan karakteristik persepsi pendengaran manusia.

Ekstraksi *Perceptual Linear Prediction* (PLP) adalah teknik yang dikembangkan untuk meniru cara kerja telinga manusia dalam memproses suara (Hermansky, 1990). Teknik ini memungkinkan ekstraksi fitur-fitur penting dari sinyal suara yang relevan untuk keperluan pengenalan suara. *Perceptual Linear Prediction* (PLP) telah digunakan dalam berbagai penelitian terkait pengenalan suara dan menunjukkan peningkatan akurasi dalam berbagai aplikasi (Hermansky, 1990). Keunggulan *Perceptual Linear Prediction* (PLP) terletak pada kemampuannya untuk mempertahankan informasi akustik yang penting dan mengurangi efek kebisingan, sehingga sangat cocok untuk digunakan dalam sistem pengenalan suara (Hermansky, 1990).

Hidden Markov Model (HMM) merupakan salah satu pendekatan statistik yang sangat populer dalam berbagai bidang pengenalan pola, terutama dalam sistem pengenalan suara otomatis. HMM bekerja berdasarkan asumsi bahwa suatu sistem dapat dimodelkan sebagai proses *Markov* tersembunyi, di mana urutan keadaan tidak dapat diamati secara langsung, tetapi dapat disimpulkan melalui observasi yang dihasilkan oleh sistem tersebut. Dalam konteks pengenalan suara, observasi ini dapat berupa fitur akustik yang diekstraksi dari sinyal audio, seperti *Mel-Frequency Cepstral Coefficients* (MFCC) dan *Perceptual Linear Prediction* (PLP) (Hermansky, 1990).

HMM telah banyak digunakan karena kemampuannya dalam menangani data *sekuensial*, dimana urutan dan durasi dari setiap elemen data sangat

penting. Model ini efektif untuk memodelkan transisi antar segmen ucapan dalam sebuah kalimat, seperti peralihan dari fonem ke fonem, sehingga sangat sesuai untuk diaplikasikan dalam sistem pengenalan suara berkelanjutan. HMM telah terbukti secara empiris mampu menghasilkan performa yang baik dalam berbagai penelitian di bidang pengolahan sinyal suara dan pengenalan ucapan (Hermansky, 1990).

Keunggulan utama dari HMM terletak pada kemampuannya dalam memodelkan proses *probabilistik* yang bergantung pada urutan waktu, menjadikannya sangat relevan dalam menangani variasi temporal dari sinyal ucapan manusia. Selain itu, dengan parameter pelatihan yang tepat dan dataset yang *representatif*, HMM dapat menyesuaikan diri terhadap variasi dalam pengucapan, aksen, serta durasi bicara dari berbagai penutur. HMM menjadi penting dalam pengembangan sistem pengenalan suara yang adaptif terhadap data nyata (Rismawati, 2022).

Implementasi teknik ini untuk Bahasa Lampung masih menghadapi berbagai tantangan, mulai dari ketersediaan data suara yang terbatas, hingga adaptasi metode yang sesuai dengan karakteristik unik Bahasa Lampung (Andina, 2023). Oleh karena itu, penelitian ini bertujuan untuk menerapkan ekstraksi *Perceptual Linear Prediction* (PLP) dan *Hidden Markov Model* (HMM) dalam sistem transkripsi suara ke teks Bahasa Lampung, serta mengevaluasi kinerjanya dalam upaya melestarikan dan mempromosikan penggunaan Bahasa Lampung dalam konteks digital.

Berdasarkan permasalahan yang telah diuraikan peneliti melakukan penelitian yang bertujuan untuk menganalisis hasil transkripsi suara ke teks dalam Bahasa Lampung dengan menggunakan ekstraksi fitur *Perceptual Linear Prediction* (PLP) dan metode *Hidden Markov Model* (HMM) pada kasus studi Lampung Pepadun Dialek A.

1.2. Rumusan Masalah

Berdasarkan latar belakang, maka rumusan masalah dalam penelitian ini yaitu sebagai berikut :

1. Bagaimana cara mengimplementasikan teknik ekstraksi *Perceptual Linear Prediction* (PLP) untuk pengolahan suara Bahasa Lampung?
2. Bagaimana kinerja metode *Hidden Markov Model* (HMM) dalam mengukur hasil akurasi transkripsi suara ke teks untuk Bahasa Lampung?

1.3. Batasan Masalah

Berdasarkan pemaparan dari latar belakang, maka batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Penggunaan dataset suara Bahasa Lampung dialek A yang terbatas, mencakup jumlah data suara, variasi penutur, dan kondisi lingkungan perekaman.
2. Penelitian ini hanya mencakup penerapan teknik *Perceptual Linear Prediction* (PLP) untuk ekstraksi fitur akustik dan *Hidden Markov Model* (HMM) untuk pengenalan pola suara.

1.4. Tujuan Penelitian

Berdasarkan pemaparan dari latar belakang, maka tujuan penelitian ini adalah :

1. Mengukur keberhasilan metode *Hidden Markov Model* (HMM) dalam konversi transkripsi suara Bahasa Lampung ke teks.
2. Menganalisis efektivitas *Perceptual Linear Prediction* (PLP) dalam ekstraksi fitur suara Bahasa Lampung.

1.5. Manfaat Penelitian

Hasil penelitian ini dapat mendukung penerapan transkripsi dengan metode *Hidden Markov Model* (HMM) dalam berbagai konteks penggunaan, seperti perjalanan, bisnis, pendidikan, dan interaksi sehari-hari.

1. Memberikan referensi bagi penelitian selanjutnya dalam bidang pengolahan suara dan transkripsi bahasa dengan pendekatan metode yang sama.
2. Memberikan kontribusi akademis dalam bidang teknologi pengenalan suara dan pelestarian bahasa daerah.
3. Melestarikan Bahasa Lampung sebagai salah satu warisan budaya melalui penerapan teknologi modern.
4. Mendorong perkembangan teknologi pengolahan suara dan transkripsi bahasa di Indonesia.

II. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Pada penelitian terdahulu ini bertujuan untuk memberikan gambaran umum mengenai studi sebelumnya yang relevan dengan topik penelitian ini, yaitu implementasi ekstraksi *Perceptual Linear Prediction* (PLP) dan metode *Hidden Markov Model* (HMM) untuk transkripsi suara ke teks Bahasa Lampung. Melalui tinjauan ini, diharapkan peneliti dapat memahami landasan teoritis yang mendasari penelitian mengenai bahasa, serta mengetahui perkembangan terkini dalam bidang pengolahan suara dan transkripsi bahasa. Penelitian yang dibahas dalam bagian ini akan membantu mengidentifikasi masalah penelitian yang masih ada dan bagaimana penelitian ini berkontribusi dalam menyelesaikan masalah tersebut. Pada Tabel 1 di bawah ini, beberapa peneliti dan jurnal yang berhubungan dengan penelitian tersebut .

Tabel 1. Peneliti Terdahulu

No	Peneliti (Tahun)	Judul Penelitian	Metode / Algoritma	Fitur Ekstraksi	Hasil
1	Maseri & Mamat (2020)	*Performance Analysis of Implemented MFCC and HMM-based Speech Recognition System*	*MFCC, HMM, Viterbi & Baum-Welch*	*MFCC (39 vektor)*	Akurasi 95%, menggunakan dataset 2200 ucapan dari 10 digit angka
2	Stadermann & Rigoll (2004)	*A Hybrid SVM/HMM Acoustic Modeling Approach to Automatic*	*SVM + HMM (Hybrid tied-posteriors)*	*RASTA-PLP*	PLP digunakan untuk meningkatkan akurasi pengenalan

No	Peneliti (Tahun)	Judul Penelitian	Metode / Algoritma	Fitur Ekstraksi	Hasil
		Speech Recognition			kata dalam kondisi *noise*
3	Kurzekar & Patil (2014)	*A Comparative Study of Feature Extraction Techniques for Speech Recognition System*	Perbandingan metode *(MFCC, LPC, ZCPA, PLP)*	*MFCC, LPC, PLP, ZCPA*	MFCC dan PLP dianggap sangat efektif untuk pengenalan digit, PLP lebih tahan terhadap *noise*

Beberapa penelitian sebelumnya telah menunjukkan efektivitas penggunaan fitur akustik seperti PLP dan MFCC dalam sistem pengenalan suara berbasis HMM *Hidden Markov Model*. Penelitian Maseri & Mamat (2020) mengimplementasikan sistem pengenalan suara dengan pendekatan MFCC dan HMM menggunakan *toolkit HTK*. Dalam penelitian tersebut, fitur MFCC sebanyak 39 dimensi digunakan untuk merepresentasikan sinyal suara, dan model HMM dilatih dengan *algoritma Baum-Welch* serta diklasifikasi menggunakan *algoritma viterbi*. Hasilnya menunjukkan akurasi pengenalan sebesar 95% untuk pengenalan digit Bahasa Inggris dalam lingkungan bersih. Meskipun penelitian ini belum menggunakan PLP, hasilnya menunjukkan bahwa kombinasi fitur *spektral* dan model HMM memberikan hasil akurasi yang tinggi, yang memperkuat pendekatan penelitian ini dalam menggunakan PLP sebagai fitur alternatif yang lebih dekat dengan persepsi pendengaran manusia.

Penelitian Stadermann & Rigoll (2004) meneliti sistem *hybrid* SVM/HMM dengan menggunakan fitur *Rasta*-PLP, yang merupakan turunan dari PLP. Penelitian ini bertujuan untuk meningkatkan robustitas terhadap *noise* dengan mengubah representasi akustik menjadi *probabilitas posterior* yang dikombinasikan dalam *framework* HMM. Hasil eksperimen pada dataset *Aurora2* menunjukkan bahwa pendekatan *hybrid* dengan *Rasta*-PLP memberikan penurunan tingkat kesalahan yang signifikan dibandingkan baseline HMM-GMM konvensional. Penelitian ini sangat relevan dengan

fokus penelitian saat ini karena menunjukkan bahwa PLP (atau turunannya) sangat efektif dalam menangani variabilitas dan *noise* dalam pengenalan ucapan berbasis HMM.

Sementara itu, Kurzekar & Patil (2014) melakukan kajian komparatif terhadap beberapa teknik ekstraksi fitur seperti MFCC, LPC, ZCPA, dan PLP dalam sistem pengenalan suara. Penelitian ini menekankan bahwa PLP merupakan salah satu fitur yang paling mendekati persepsi manusia dan lebih tahan terhadap *noise* dibandingkan MFCC. Hasil studi menunjukkan bahwa PLP sangat efektif dalam pengenalan digit dan aplikasi *real-time*. Meskipun penelitian ini tidak secara eksplisit menggunakan HMM sebagai model klasifikasi, PLP yang digunakan sangat sesuai untuk diterapkan dalam *framework* HMM, yang menjadi inti dari penelitian ini.

2.2.*Resampling* Audio

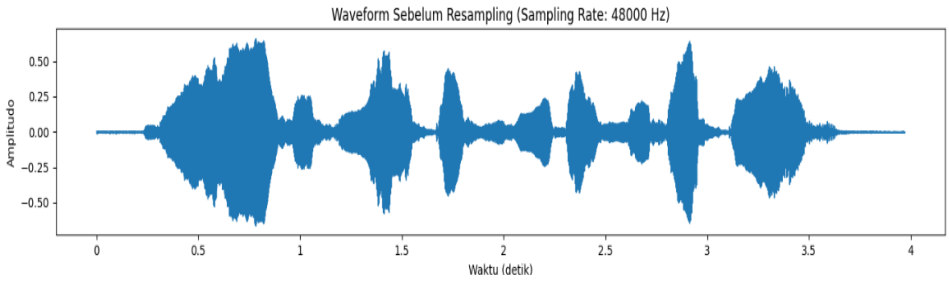
Proses pengolahan sinyal suara dalam sistem pengenalan ucapan (*speech recognition*) melibatkan beberapa tahapan penting dalam *preprocessing*, salah satunya adalah proses *resampling*. *Resampling* merupakan teknik untuk mengubah laju pengambilan sampel (*sampling rate*) dari sinyal audio agar selaras dengan kebutuhan sistem. Langkah ini dilakukan karena data audio yang dikumpulkan dari berbagai sumber sering kali memiliki *sampling rate* yang berbeda-beda, seperti *44,1 kHz*, *22,05 kHz*, atau *48 kHz*. Jika tidak diseragamkan, perbedaan ini dapat memengaruhi proses ekstraksi fitur dan performa model secara keseluruhan (Davis & Mermelstein, 1980).

Seluruh data suara diubah ke laju *sampling* sebesar *16 kHz* (16.000 sampel per detik). Pemilihan laju sampling *16 kHz* karena frekuensi ini telah menjadi standar umum dalam banyak sistem pengenalan suara modern, termasuk sistem berbasis *Hidden Markov Model* (HMM) maupun *Deep Learning*. Frekuensi *16 kHz* mampu untuk merepresentasikan informasi penting dalam sinyal ucapan manusia, yang sebagian besar berada pada rentang frekuensi antara *0 kHz* hingga *8 kHz*. Berdasarkan prinsip *Nyquist*,

laju sampling minimal yang diperlukan untuk merepresentasikan sinyal dengan frekuensi maksimum f adalah $2f$ sehingga laju *16 kHz* memadai untuk menangkap komponen penting dari spektrum suara manusia (Rabiner & Schafer, 2011).

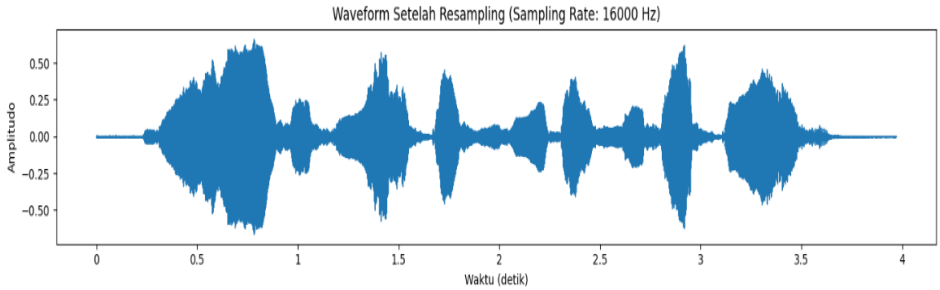
Selain itu, pemilihan laju sampling ini juga didasari oleh pertimbangan efisiensi komputasi. Penggunaan *sampling rate* yang lebih tinggi akan menghasilkan ukuran file yang lebih besar, yang secara langsung akan meningkatkan kebutuhan penyimpanan dan waktu komputasi, terutama pada proses ekstraksi fitur dan pelatihan model. Sebaliknya, dengan menggunakan laju sampling *16 kHz*, ukuran data menjadi lebih kecil dan proses pemrosesan menjadi lebih cepat tanpa kehilangan informasi akustik yang signifikan. Oleh karena itu, *resampling* ke *16 kHz* dapat dianggap lebih optimal antara kualitas representasi sinyal dan efisiensi pemrosesan.

Berdasarkan Gambar 1, ditampilkan bentuk gelombang (*waveform*) dari sinyal audio file kalimat *devi/devi_5/kalimat7.wav* sebelum dilakukan proses *resampling*, dengan frekuensi pengambilan sampel (*sampling rate*) sebesar *48.000 Hz*. Pada kondisi ini, setiap detik sinyal direpresentasikan oleh 48.000 titik data (sampel), sehingga memberikan resolusi temporal yang sangat tinggi. Hal ini memungkinkan detail halus dari sinyal suara, termasuk komponen frekuensi tinggi. Namun, tingginya laju sampling ini tidak selalu efisien untuk sistem pengenalan suara, karena informasi fonetik penting dalam ucapan manusia umumnya terkonsentrasi pada rentang *0–8.000 Hz.* Selain itu, ukuran data yang besar juga akan meningkatkan beban komputasi dan waktu pemrosesan dalam proses pelatihan dan inferensi model, terutama dalam sistem berskala besar.



Gambar 1. Audio Sebelum *Resampling*.

Pada Gambar 2, dari sinyal audio file kalimat *devi/devi_5/kalimat7.wav* menunjukkan bentuk *waveform* dari sinyal setelah dilakukan proses resampling ke 16.000 Hz, yaitu menurunkan laju sampling dari sinyal asli. Proses ini dilakukan dengan metode interpolasi dan *filtering* agar tetap mempertahankan informasi penting dari sinyal suara. Meski terdapat sedikit perbedaan detail pada bentuk gelombangnya, struktur utama dari pola ucapan tetap sama, sehingga informasi fonetik relevan untuk sistem pengenalan ucapan.



Gambar 2. Audio Setelah *Resampling*.

Dengan demikian, proses *resampling* dari *48 kHz* ke *16 kHz* dilakukan untuk menyesuaikan dengan standar umum sistem pengenalan suara, frekuensi *16 kHz* sudah cukup untuk menangkap informasi penting dari ucapan manusia yang umumnya berada pada rentang frekuensi *0–8 kHz*. *Resampling* ini memberikan hasil terbaik antara kualitas informasi suara, efisien, pemrosesan dan memastikan bahwa seluruh *spectrum* penting dari sinyal ucapan dapat ditangkap secara akurat oleh sistem pengenalan suara.

2.3.*Perceptual Linear Prediction*

Perceptual Linear Prediction (PLP) adalah teknik yang digunakan dalam analisis sinyal suara untuk mengekstraksi fitur penting yang relevan dengan cara telinga manusia mendengar. *Perceptual Linear Prediction* (PLP) dirancang untuk meniru cara kerja sistem pendengaran manusia, sehingga memberikan peningkatan akurasi dalam pengenalan suara, terutama dalam kondisi akustik yang beragam (Hermansky, 1990). *Perceptual Linear Prediction* (PLP) didasarkan pada konsep *perceptual processing*, yaitu bagaimana telinga manusia dan sistem pendengaran memproses dan menafsirkan sinyal suara. Proses ini melibatkan beberapa langkah utama :

1. *Preprocessing* Sinyal Suara

Pre-emphasis bertujuan untuk memperkuat komponen frekuensi tinggi dalam sinyal suara. Frekuensi tinggi cenderung melemah atau mengalami attenuation lebih besar dibandingkan frekuensi rendah selama proses perekaman suara. Dengan memperkuat frekuensi tinggi, sinyal menjadi lebih seimbang dan siap untuk langkah pemrosesan berikutnya. Persamaan (1) adalah formula untuk memperkuat frekuensi tinggi.

$$y(n) = x(n) - \alpha . x(n-1) \quad \text{.....(1)}$$

$y(n)$: Sinyal output setelah *pre-emphasis*

$x(n)$: Sinyal input asli

α : Faktor *pre-emphasis* (biasanya bernilai 0,95)

Dengan menerapkan *pre-emphasis*, perbedaan antara frekuensi tinggi dan rendah menjadi lebih seimbang. Hal ini membantu model STT lebih mudah mengenali pola suara dan meningkatkan akurasi transkripsi. Teknik ini sangat penting dalam penelitian *Speech-to-Text* untuk Bahasa Lampung, karena bahasa ini memiliki fonem unik yang memerlukan kejelasan tinggi dalam proses ekstraksi fitur. Namun, meskipun *pre-emphasis* meningkatkan kejelasan suara, teknik ini juga dapat menyebabkan bias pada sinyal suara.

Bias ini muncul karena penguatan frekuensi tinggi dapat mengubah karakteristik alami suara, terutama jika sinyal mengalami distorsi akibat *noise* atau rekaman berkualitas rendah. Jika tidak diterapkan dengan parameter yang tepat, *pre-emphasis* bisa terlalu menekan frekuensi rendah, yang dapat mengurangi informasi penting dalam suara dan menyebabkan kesalahan dalam transkripsi.

Sehingga, *pre-emphasis* menjadi tahap awal yang membantu meningkatkan kualitas suara sebelum masuk ke proses ekstraksi fitur PLP, yang nantinya digunakan dalam sistem HMM untuk pengenalan suara. Oleh karena itu, pemilihan faktor *pre-emphasis* (α) yang optimal sangat penting untuk menghindari bias yang dapat mengganggu hasil transkripsi.

2. *Framing* dan *Windowing*

Setelah *pre-emphasis*, sinyal suara dipecah menjadi frame kecil dengan durasi sekitar 20-30 ms, karena suara manusia bersifat berubah dan tidak tentu dalam rentang waktu ini. Membagi sinyal suara menjadi *frame* kecil sangat penting untuk menganalisis sinyal secara lokal. *Hamming Window* digunakan untuk menghaluskan sinyal di ujung *frame*, dengan Persamaan (2) dan (3).

***Hamming Window*:**

$$w(n) = 0{,}54 - 0{,}46 \cos\left(\frac{2\pi n}{N-1}\right) \quad (2)$$

Sinyal yang di-*windowing*:

$$x_\omega(n) = x(n) \,.\, \omega\,(n) \quad (3)$$

N : Jumlah sampel dalam satu *frame*

$\omega\,(n)$: Nilai *window* untuk sample ke- n

$x_\omega(n)$: Sinyal setelah *windowing*

Dalam menerapkan teknik *framing* dan *windowing*, sinyal suara yang panjang dipotong menjadi bagian-bagian kecil yang disebut *frame*. Hal ini dilakukan karena suara manusia selalu berubah seiring waktu, sehingga lebih mudah dianalisis jika dibagi menjadi segmen pendek, biasanya sekitar 20-30 milidetik. Jika sinyal suara langsung dipotong tanpa *windowing*, akan ada perubahan di ujung potongan, yang bisa menyebabkan gangguan atau distorsi saat suara dikonversi ke bentuk digital. Untuk mengatasi masalah ini, digunakan *hamming window*, yaitu teknik yang membuat tepi potongan suara lebih halus.

Dengan cara ini, setiap potongan suara tetap memiliki transisi yang mulus, sehingga saat digabungkan kembali, hasilnya tidak terdengar patah atau aneh. Teknik ini sangat penting dalam sistem *Speech-to Text* (STT), terutama untuk bahasa seperti Bahasa Lampung, yang memiliki bunyi khas yang harus diproses dengan jelas. Selain itu, teknik *Framing* dan *Windowing* juga membantu mengurangi kesalahan dalam mengenali suara. Dengan *windowing*, suara yang sudah dipotong tidak kehilangan informasi penting dan lebih mudah dipahami oleh komputer. Jika suara diproses dengan lebih rapi, model kecerdasan buatan seperti *Hidden Markov Model* (HMM) akan lebih akurat dalam mengubah suara menjadi teks. Dengan kata lain, *Framing* dan *Windowing* membantu komputer mengenali suara dengan lebih baik, mengurangi kesalahan transkripsi, dan membuat hasil *Speech-to-Text* lebih akurat.

3. *Fast Fourier Transform* (FFT)

FFT digunakan untuk mengubah sinyal suara dari domain waktu ke domain frekuensi. Proses ini penting karena komponen frekuensi dalam sinyal suara memberikan informasi tentang pola spektrum suara. Hasil FFT berupa spektrum energi frekuensi dari setiap frame sinyal suara. FFT bertujuan untuk mengidentifikasi distribusi energi pada berbagai komponen frekuensi, menghasilkan representasi spektrum yang akan diproses lebih lanjut oleh

Mel Filter Bank dan meningkatkan efisiensi dalam analisis sinyal suara, dengan Persamaan (4).

$$X(k) = \sum_{n=0}^{N-1} x_{\omega}(n).\ e^{-j2\pi kn/N} \quad \text{.....} \quad (4)$$

$X(k)$: Amplitudo spektrum pada frekuensi ke k
$x_{\omega}(n)$: Sinyal setelah *windowing*
N : Ukuran FFT

FFT merupakan bagian dari proses mengubah suara menjadi bentuk yang bisa dipahami oleh komputer. Saat kita berbicara, suara kita berupa gelombang yang berubah seiring waktu. Namun, agar komputer bisa mengenali suara dengan lebih baik, kita perlu melihat frekuensi yang ada di dalam suara tersebut. *Transformasi Fourier* digunakan untuk mengubah suara dari bentuk waktu ke frekuensi, sehingga kita bisa tahu bunyi apa saja yang terkandung dalam suara itu. Setelah suara dipotong menjadi bagian kecil (*framing*) dan dibuat lebih halus (*windowing*), langkah selanjutnya adalah menganalisis frekuensi yang ada di dalamnya. $X(k)$ dalam rumus ini menunjukkan seberapa kuat bunyi pada frekuensi tertentu. Ukuran FFT (N) menentukan seberapa rinci kita bisa melihat frekuensi tersebut. Semakin besar nilai (N), semakin detail informasi yang bisa diperoleh.

Dalam sistem *Speech-to-Text* untuk Bahasa Lampung, proses ini sangat penting karena membantu komputer mengenali bunyi khas dari bahasa tersebut. Dengan mengetahui pola frekuensi dalam suara, sistem bisa lebih akurat dalam menerjemahkan suara menjadi teks. Teknik ini juga digunakan dalam model *Hidden Markov Model* (HMM) agar sistem bisa lebih dapat memahami dan mengonversi suara menjadi kata yang benar.

4. *Bark Scale Filterbank*

Bark scale filterbank merupakan proses untuk membagi suara ke dalam beberapa bagian frekuensi berdasarkan cara telinga manusia mendengar. *Skala bark* lebih mendekati persepsi manusia dibandingkan pembagian

frekuensi biasa, karena mengikuti sensitifitas telinga terhadap perubahan nada, *filterbank* digunakan untuk menangkap informasi penting dari suara di setiap bagian frekuensi. Setiap bagian disebut pita *bark*, dan sistem akan menghitung berapa besar energi suara di tiap pita tersebut. Pada tahapan ini, suara diubah menjadi bentuk yang lebih dimengerti oleh mesin, karena fokusnya mengikuti cara kerja pendengaran manusia.

Hasil dari tahap ini akan digunakan untuk membuat fitur suara yang lebih akurat saat dilatih ke model, dengan Persamaan (5).

$$E_{bark}(m) = \sum_{k=0}^{n-1} |X(k)|^2 \,.\, H_m(\mathrm{k}) \quad \text{.....} \quad (5)$$

$E_{bark}(m)$: Energi pada *filter bark* ke-m

$X(k)|^2$: *Spektrum* sinyal suara

H_m(k) : Respons dari filter Bark ke m pada frekuensi ke k

Persamaan *bark scale filterbank* bertujuan untuk menghitung total energi sinyal suara dalam setiap pita frekuensi *bark*. Dalam proses ini, sinyal suara yang semula berada di domain waktu diubah terlebih dahulu ke domain frekuensi menggunakan *Fast Fourier Transform* (FFT), menghasilkan spektrum frekuensi yang dilambangkan sebagai $X(k)$. Untuk mendapatkan spektrum daya atau *power spectrum*, setiap komponen frekuensi diambil kuadrat dari magnitudonya, yaitu $|X(k)|^2$. Selanjutnya, *power spectrum* ini dikalikan dengan respons filter *bark* H_m (k), yaitu fungsi yang merepresentasikan sensitivitas filter ke (m) terhadap setiap frekuensi. Proses ini memungkinkan hanya frekuensi yang berada di dalam rentang tertentu, yakni rentang yang ditentukan oleh *filter bark* ke (m) yang berkontribusi terhadap energi total filter tersebut. Perkalian ini dilakukan untuk seluruh frekuensi, lalu dijumlahkan, sehingga menghasilkan nilai energi $E_{bark}(m)$, yaitu seberapa besar energi sinyal suara dalam *filter bark* ke-m. Secara konseptual, pendekatan ini mensimulasikan bagaimana telinga manusia memproses suara berdasarkan persepsi frekuensi, bukan distribusi

linier, sehingga fitur yang diekstrak menjadi lebih relevan untuk mengenali fonem dalam bahasa, termasuk dalam konteks bahasa daerah.

5. *Equal-Loudness Pre-emphasis*

Pada tahap ini, energi spektral hasil *filter bank* diperbaiki sesuai dengan *equal-loudness contour*. Hal ini mensimulasikan *sensitivitas* telinga manusia terhadap frekuensi yang berbeda Intensitas frekuensi disesuai dengan sensitivitas pendengaran manusia, memperkuat frekuensi yang lebih relevan secara perseptual, mengurangi energi yang kurang penting, dapat dilihat dengan Persamaan (6).

$$H(f) = \left(\frac{f^2+56.8.10^6 f^4}{(f^2+6.3.10^6)(f^2+3.8.10^8)}\right) \cdot \frac{1}{f^2+9.58.10^3} \quad \text{.....} \quad (6)$$

$H(f)$: Respons frekuensi *equal-loudness*

f : Frekuensi dalam Hz

Dalam merepresentasikan respons frekuensi *equal-loudness*, yang digunakan dalam proses ekstraksi fitur suara untuk menyesuaikan cara komputer memahami suara dengan cara manusia mendengarnya. Dalam sistem *Speech-to-Text* (STT), pendengaran manusia tidak merespons semua frekuensi dengan tingkat sensitivitas yang sama. Frekuensi tertentu, terutama di rentang 1 *kHz* hingga 5 *kHz*, lebih mudah didengar dibandingkan frekuensi yang sangat rendah atau sangat tinggi. Oleh karena itu, *filter equal-loudness* diterapkan untuk memperkuat atau melemahkan bagian tertentu dari spektrum suara agar lebih sesuai dengan cara manusia mendengar.

Persamaan $H(f)$ menunjukkan bagaimana energi pada berbagai frekuensi disesuaikan dengan persepsi manusia. Bagian pertama dari persamaan menyesuaikan tingkat kepekaan telinga manusia terhadap frekuensi tertentu, sedangkan, bagian kedua berfungsi sebagai faktor normalisasi untuk memastikan distribusi energi tetap seimbang. Dengan menerapkan

fungsi ini, suara dengan frekuensi tinggi yang cenderung lebih lemah dalam rekaman dapat diperkuat, sementara suara dengan frekuensi rendah yang lebih dominan dapat sedikit dikurangi untuk menciptakan keseimbangan yang lebih baik dalam analisis fitur suara, penggunaan *equal-loudness filtering* sangat penting karena membantu meningkatkan akurasi pengenalan suara dengan mempertajam karakteristik yang relevan. Hal ini memungkinkan sistem untuk lebih mudah membedakan bunyi-bunyi yang mirip dan mengurangi gangguan dari komponen suara yang kurang signifikan, teknik ini tidak hanya meningkatkan kualitas fitur suara yang diekstraksi, tetapi juga membantu model *Hidden Markov Model* (HMM) atau jaringan saraf dalam memahami suara dengan lebih baik, sehingga meningkatkan akurasi transkripsi suara ke teks.

6. *Logarithmic Compression*

Logarithmic Compression atau kompresi logaritmik adalah proses yang bertujuan untuk menyesuaikan sinyal suara digital dengan karakteristik persepsi pendengaran manusia. Dalam sistem pendengaran manusia, perubahan energi suara tidak dipersepsikan secara *linear*. Artinya, telinga manusia lebih peka terhadap perubahan kecil pada energi rendah, dan kurang peka terhadap perubahan besar pada energi tinggi. Untuk meniru perilaku ini dalam pengolahan sinyal digital, kita menerapkan fungsi logaritma pada energi spektral hasil filter Bark. Secara matematis, transformasi ini mengubah energi *linier* menjadi skala logaritmik, seperti dalam Persamaan (7).

$$\mathrm{C}(m) = \log(E_{bark}(m)) \quad \text{.....} \quad (7)$$

$E_{bark}(m)$: Energi pada *filter Bark* ke m (hasil dari *filterbank bark*)

$C(m)$: Nilai setelah kompresi logaritmik untuk *filter bark* ke m

Sebuah filter pada *filterbank bark* dilambangkan dengan $E_{bark}(m)$, yang merepresentasikan total energi sinyal dalam pita frekuensi *bark* ke- m. Nilai

ini kemudian dikompresi menggunakan fungsi logaritma sehingga menghasilkan $C(m)$. Hasil dari kompresi logaritmik ini akan digunakan sebagai bagian dari vektor fitur PLP. Tujuannya adalah untuk membuat fitur lebih stabil dan efisien, terutama ketika menghadapi variasi dalam intensitas suara (misalnya suara keras atau pelan). Dalam konteks sistem *Speech-to-Text* (STT), hal ini memungkinkan model mengenali pola suara dengan lebih akurat, karena input yang diterima telah disesuaikan dengan sensitivitas pendengaran manusia.

7. *Linear Prediction Coefficients* (LPC)

Linear Prediction Coefficients (LPC) dihitung untuk menyederhanakan fitur *cepstral*. LPC merepresentasikan sinyal suara dalam bentuk prediksi *linear*, yang sangat efisien dalam menyimpan informasi akustik. Tujuan dari LPC adalah merepresentasikan sinyal suara dalam bentuk prediksi *linear*, menangkap pola akustik yang relevan dari sinyal suara, menghasilkan fitur akhir PLP yang siap digunakan untuk analisis atau klasifikasi, dengan Persamaan (8).

$$x[n] = -\sum_{i=1}^{p} ai.x[n-i] \quad \text{.....} \quad (8)$$

$x[n]$: Sinyal prediksi

ai : Koefisien prediksi *linear*

p : Order prediksi

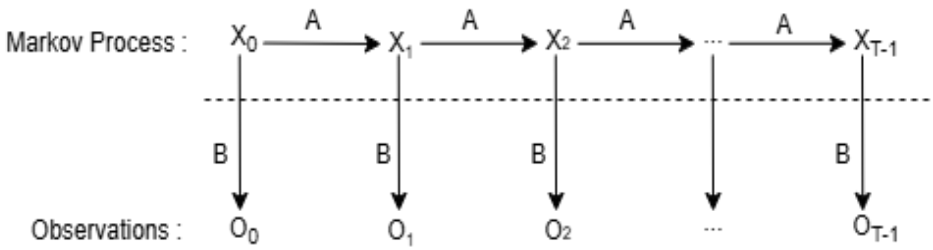
Model prediksi *linear* yang digunakan dalam proses ekstraksi fitur suara, khususnya dalam metode *Linear Predictive Coding* (LPC) yang juga digunakan dalam *Perceptual Linear Prediction* (PLP). Model ini berfungsi untuk memprediksi nilai sinyal suara saat ini $x[n]$ berdasarkan nilai-nilai sebelumnya. Pada sistem *Speech-to-Text* (STT), pendekatan ini membantu dalam merepresentasikan pola suara dengan lebih efisien, karena suara manusia memiliki sifat yang dapat diprediksi dalam jangka pendek, persamaan ini, $x[n]$ adalah sinyal suara yang diprediksi, sedangkan ai

adalah koefisien prediksi yang menentukan seberapa besar pengaruh dari nilai-nilai sebelumnya terhadap nilai saat ini. Parameter p disebut sebagai order prediksi, yang menentukan jumlah sampel sebelumnya yang digunakan untuk memperkirakan sinyal saat ini. Semakin besar nilai p, semakin kompleks model prediksinya, yang berarti bisa menangkap lebih banyak informasi dalam suara, tetapi juga bisa menyebabkan *overfitting* jika tidak diatur dengan baik.

Setelah tahapan *Perceptual Linear Prediction* (PLP) dilakukan, hasil yang diperoleh adalah koefisien PLP, yang merupakan sekumpulan angka numerik yang merepresentasikan ciri akustik dari sinyal suara. Hasil ini berupa vektor fitur yang mencerminkan karakteristik *spektrum* suara pada setiap *frame* sinyal. Metode ini memungkinkan sistem untuk menangkap pola suara manusia dengan lebih baik, sehingga suara dapat diubah menjadi fitur numerik yang lebih mudah dipahami oleh model HMM. Kemudian *Vektor* fitur yang diperoleh melalui PLP menjadi input utama dalam pemodelan *Hidden Markov Model* (HMM), yang digunakan untuk mengenali pola suara dan melakukan transkripsi ke teks Bahasa Lampung Pepadun A.

2.4.*Hidden Markov Model*

Hidden Markov Model (HMM) adalah model *probabilistik* yang digunakan untuk memodelkan sistem dengan dua komponen utama, yaitu *state* tersembunyi dan observasi. Dalam HMM, state tersembunyi (x_t) membentuk proses *Markov*, di mana setiap *state* hanya bergantung pada *state* sebelumnya. Sementara itu, observasi (O_t) yang dapat diamati dihasilkan dari state tersembunyi melalui probabilitas emisi. (Rabiner, 1989).



Gambar 3. *Model Hidden Model* (Rabiner, 1989).

1. Probalitas Awal

Pada tahap ini mendefinisikan distribusi probabilitas *state* tersembunyi pada $t = 0$. Parameter ini menentukan kemungkinan *state* awal yang akan dipilih dalam proses HMM. Probabilitas awal didefinisikan melalui Persamaan (9).

$$\pi_i = \mathrm{P}(X_0 = i), 1 \leq i \leq N \quad \text{.....} \quad (9)$$

π_i : *probabilitas state* tersembunyi X_0 berada pada *state* i

N : jumlah total *state* tersembunyi

Probabilitas awal π_i menunjukkan kemungkinan bahwa pada waktu awal $t = 0$, sistem berada dalam *state* tersembunyi ke i dari total N state tersembunyi. Nilai ini sangat penting karena akan menentukan bagaimana sistem memulai analisisnya terhadap sinyal suara. Misalnya, dalam pengenalan ucapan, sistem harus memilih keadaan awal yang paling mungkin, seperti apakah suara yang masuk lebih cenderung ke huruf vokal atau konsonan, sebelum melanjutkan ke tahap selanjutnya dalam proses *decoding* suara menjadi teks.

2. Matriks Transisi

Matriks transisi (A) menggambarkan probabilitas perpindahan dari satu *state* tersembunyi ke *state* tersembunyi lainnya seiring waktu. Pada HMM, *state* saat ini (X_t) hanya bergantung pada state sebelumnya (X_{t-1}), sesuai dengan prinsip rantai *markov*. Secara matematis, probabilitas transisi didefinisikan dapat dilihat pada Persamaan (10).

α_{ij}=P$(X_t = j|X_{t-1} = i), 1 \leq i, j \leq N$ .....(10)

α_{ij} : *probabilitas* berpindah dari *state* i ke *state* j

N : jumlah *state* tersembunyi

Matriks transisi menggambarkan kemungkinan perpindahan dari satu *state* tersembunyi ke state tersembunyi lainnya seiring waktu. Pada *Speech-to-Text* (STT), sistem harus memahami bagaimana suara berubah dari satu keadaan ke keadaan berikutnya. Matriks transisi ini membantu sistem dalam memprediksi urutan fonem atau kata berdasarkan hubungan antar suara dalam suatu kalimat atau ucapan.

Pada persamaan tersebut, α_{ij} menunjukkan probabilitas bahwa sistem berpindah dari state tersembunyi ke-i ke *state* tersembunyi ke-j pada waktu berikutnya. Dalam HMM *state* saat ini (X_t) hanya bergantung pada state sebelumnya (X_{t-1}), sesuai dengan prinsip rantai *Markov*. Ini berarti bahwa keadaan suara saat ini hanya dipengaruhi oleh keadaan sebelumnya, bukan oleh keseluruhan sejarah suara sebelumnya. Misalnya, jika sistem mengenali sebuah fonem, maka probabilitas transisi akan membantu memperkirakan fonem berikutnya berdasarkan pola bahasa yang umum digunakan. Jika probabilitas perpindahan antara dua *state* tinggi, maka sistem akan lebih cenderung memilih transisi tersebut dalam proses *decoding* suara ke teks.

3. Matriks Emisi

Matriks emisi (B) menggambarkan hubungan antara *state* tersembunyi dengan observasi yang dapat diukur. Probabilitas emisi menunjukkan kemungkinan keluaran observasi (O_t) diberikan *state* tersembunyi tertentu (X_t). Secara matematis, probabilitas *emisi* didefinisikan pada Persamaan (11).

$b_j(O_t)$=P$(O_t|X_t = j)$, $1 \leq j \leq N, 1 \leq O_t \leq M$ ..... (11)

$b_j(O_t)$: *probabilitas* observasi (O_t) muncul ketika *state* tersembunyi berada pada *state j*

N : jumlah *state* tersembunyi

M : total *observasi*

Matriks emisi menentukan kemungkinan suatu observasi muncul ketika sistem berada dalam *state* tertentu. Dalam persamaan ini, $b_j(O_t)$ menunjukkan probabilitas bahwa observasi tertentu (O_t) terjadi jika sistem berada dalam state tersembunyi ke- j. Maka matriks emisi membantu sistem memetakan karakteristik suara ke state tersembunyi yang sesuai, sehingga model dapat mengenali pola suara dengan lebih baik. Misalnya, jika seseorang mengucapkan fonem tertentu, matriks emisi menentukan seberapa besar kemungkinan sistem mengenali fonem tersebut berdasarkan fitur suara yang diperoleh. Dengan probabilitas emisi yang terlatih dengan baik, model HMM dapat lebih akurat dalam mengidentifikasi suara dan mengonversinya ke teks, meskipun terdapat variasi dalam intonasi, aksen, atau kualitas rekaman suara.

Hidden Markov Model (HMM) bekerja dengan asumsi bahwa setiap keadaan tersembunyi bergantung hanya pada keadaan sebelumnya, bukan pada semua keadaan sebelumnya. Ini dikenal sebagai asumsi *Markov*, yang menyederhanakan perhitungan dalam model (Rismawati, 2022). Dalam konteks *speech-to-text*, *Hidden Markov Model* (HMM) memodelkan aliran sinyal suara untuk mengenali kata-kata dalam sebuah ucapan (Rabiner, 1989). Model ini bekerja dengan menguraikan sinyal suara menjadi fitur-fitur tertentu dan menggunakan probabilitas transisi serta distribusi emisi untuk mengidentifikasi rangkaian kata *Hidden Markov Model* (HMM) dapat memperkirakan kata atau fonem mana yang paling mungkin sesuai dengan urutan sinyal suara yang diterima.

Proses kerja *Hidden Markov Model* (HMM) mencakup tiga tahapan utama, yaitu:

1. *Inisialisasi*: Sistem mulai pada salah satu status tersembunyi (θ_1) berdasarkan *probabilitas* awal (π_i).
2. *Transisi* Status: *State* tersembunyi berpindah dari X_{t-1} ke X_t berdasarkan matriks transisi (A).
3. *Emisi Observasi*: Pada setiap state X_t , *observasi* O_t dihasilkan berdasarkan *matriks emisi* (B).
4. *Probabilitas* Total: total untuk suatu urutan *observasi* $O(O_1, O_{2,} O_3 \ldots O_{T-1})$ adalah kombinasi dari semua kemungkinan jalur status tersembunyi, dapat dilihat dalam Persamaan (12):

$$P(O|\lambda) = \sum_X P(X_0) \prod_{t=1}^{T} P(X_t|X_t) \prod_{t=0}^{T} P(O_t|X_t) \quad \text{.....(12)}$$

dimana λ = (A, B, π) adalah parameter model HMM.

$P(O|\lambda)$: Menunjukkan probabilitas bahwa observasi suara tertentu terjadi berdasarkan model HMM dengan parameter λ.

$\sum_X P(X_0)$: Menjumlahkan semua kemungkinan *state* awal dalam model.

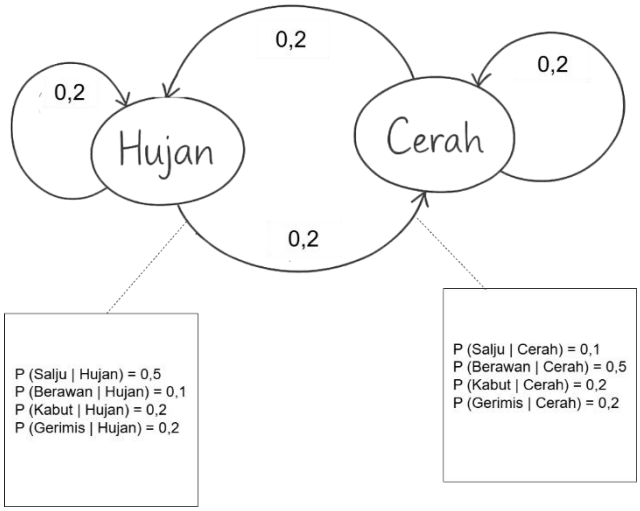
$P(X_t|X_{t-1})$: Probabilitas transisi dari *state* sebelumnya (X_{t-1}) ke state saat ini (X_t) sesuai dengan matriks transisi A.

$P(O_t|X_t)$: Probabilitas bahwa observasi tertentu (O_t) muncul saat berada dalam *state* (X_t), yang didefinisikan dalam matriks emisi B.

Persamaan ini menggambarkan probabilitas suatu urutan observasi (O) pada *Hidden Markov Model* (HMM), yang sangat penting dalam sistem *Speech-to-Text* (STT). Tujuan utama dari persamaan ini adalah menghitung kemungkinan bahwa suatu urutan observasi suara dihasilkan oleh model HMM tertentu yang memiliki parameter λ = (A, B, π).

4. Cara Kerja *Hidden Markov Model*

Pada Gambar 4, sebuah stasiun cuaca mencatat kondisi cuaca setiap jam. Cuaca sebenarnya (state tersembunyi) terdiri dari dua jenis: Hujan dan Cerah. Namun, yang diamati oleh stasiun cuaca adalah observasi seperti Salju, Berawan, Kabut, atau Gerimis.



Gambar 4. Prediksi Cuaca.

Model Hidden Markov (HMM) digunakan untuk:

a. Memodelkan transisi antar state cuaca.
b. Menghitung peluang observasi berdasarkan state.
c. Menyimpulkan state tersembunyi berdasarkan observasi menggunakan *Teorema Bayes*.

Informasi Model :

1) State Cuaca : Hujan dan Cerah
2) Transisi Antar State :

   Hujan ke Hujan : 0,2

   Hujan ke Cerah : 0,2

   Cerah ke Hujan : 0,2

   Cerah ke Cerah : 0,2

Probabilitas Emisi (Observasi | State)

Pada Tabel 2, menunjukkan probabilitas emisi, yaitu peluang munculnya observasi tertentu berdasarkan keadaan cuaca yang sebenarnya (state tersembunyi). Dalam konteks model *Hidden Markov*, probabilitas emisi menggambarkan kemungkinan stasiun cuaca mengamati kondisi tertentu seperti salju, berawan, kabut, atau gerimis, berdasarkan apakah cuaca sebenarnya sedang hujan atau cerah.

Tabel 2. Probabilitas Emisi.

Observasi	P (Observasi | Hujan)	P (Observasi | Cerah)
Salju	0,5	0,1
Berawan	0,1	0,5
Kabut	0,2	0,2
Gerimis	0,2	0,2

Probabilitas emisi ini menjadi komponen penting dalam proses inferensi *Hidden Markov Model* (HMM), khususnya saat menghitung kemungkinan state tersembunyi berdasarkan observasi yang diperoleh. Dengan demikian, nilai emisi ini akan digunakan pada perhitungan selanjutnya untuk menentukan distribusi probabilitas state pada waktu tertentu serta untuk melakukan prediksi dan klasifikasi terhadap pola cuaca yang diamati.

Perhitungan pertama berapa P ($q_2 = Hujan$)?

P ($q_1 = Hujan$) = 0,5 dan P ($q_1 = Cerah$) = 0,5

Asumsi distribusi awal (*initial state*), maka :

P (q_2=Hujan) = P (q_1=Hujan). P (Hujan - Hujan) + P (q_1=Cerah). P (Cerah - Cerah)

= 0,5 x 0,2 + 0,5 x 0,2 = 0,1 + 0,1 = 0,2

Perhitungan kedua berapa P ($O_2 = Berawan$)?

Dihitung dari total kemungkinan observasi berawan, maka :

P (O_2=Berawan) = P (q_2=Hujan). P (Berawan | Hujan) + P (q_2=Cerah). P (Berawan | Cerah)

Nilai P (q_2=Cerah) = 1 – P (q_2=Hujan) = 1 – 0,2 = 0,8

Jadi nilai P (O_2=Berawan) adalah = 0,2 x 0,1 + 0,8 x 0,5 = 0,02 + 0,4 = 0,42

Perhitungan ketiga berapa P ($q_2 = Hujan \mid O_2 =$ Berawan)?

Dihitung menggunakan *Teorema Bayes*, maka :

$$\text{P}\,(q_2 = Hujan \mid O_2 = \text{Berawan}) = \frac{\text{P}\,(q_2{=}Hujan).\ \text{P}\,(Berawan \mid Hujan)}{\text{P}\,(O_2{=}\ \text{Berawan})}$$

$$= \frac{0{,}2\ x\ 0{,}1}{0{,}42} = \frac{0{,}02}{0{,}42} = 0{,}0476 \text{ atau } 4{,}76\%$$

Berdasarkan hasil perhitungan dengan metode *Hidden Markov Model* (HMM), diperoleh bahwa probabilitas cuaca pada jam ke-2 adalah Hujan sebesar 0,2, sedangkan peluang munculnya observasi Berawan pada jam yang sama adalah 0,42. Selanjutnya, dengan menggunakan Teorema Bayes, ditemukan bahwa kemungkinan cuaca sebenarnya adalah Hujan jika diamati Berawan hanya sebesar 0,0476 atau sekitar 4,76%. Hal ini menunjukkan bahwa observasi Berawan lebih kuat diasosiasikan dengan kondisi Cerah daripada Hujan.

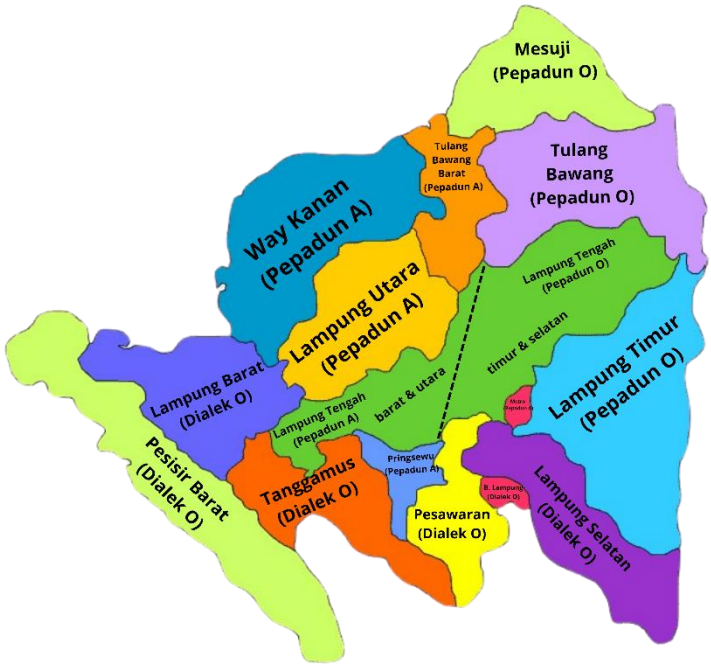
Dengan demikian, HMM mampu mengidentifikasi kemungkinan state tersembunyi (cuaca sebenarnya) berdasarkan observasi yang tampak, sekaligus menunjukkan bagaimana pola probabilistik dapat digunakan untuk menganalisis dan memprediksi kondisi cuaca secara kuantitatif.

2.5.Sejarah Bahasa Lampung Dialek Pepadun

Bahasa Lampung merupakan salah satu bahasa daerah yang termasuk dalam rumpun bahasa Melayu-Polinesia dalam keluarga Austronesia. Bahasa ini dituturkan oleh masyarakat di Provinsi Lampung dan sebagian wilayah Sumatera Selatan. Dialek dalam Bahasa Lampung terbagi menjadi dua, yaitu dialek Api (A) dan dialek Nyo (O), dengan dialek Nyo lebih dikenal sebagai dialek Pepadun (Roveneldo, 2022).

Dialek Pepadun, yang digunakan oleh komunitas adat Pepadun, memiliki karakteristik unik dibandingkan dialek Api. Dialek ini berkembang di wilayah pedalaman Lampung, di mana komunitas Pepadun memiliki sistem sosial yang lebih egaliter dibandingkan komunitas Saibatin di wilayah pesisir (Udin et al., 1998). Dalam sejarahnya, dialek Pepadun berkembang seiring dengan migrasi penduduk dan interaksi budaya di Sumatera bagian selatan. Tradisi lisan, seperti pantun, cerita rakyat, dan ritual adat, menjadi media penting untuk melestarikan dialek ini (Udin et al., 1998).

Bahasa Lampung dialek Pepadun menunjukkan kekayaan budaya lokal yang tercermin dalam variasi kosakata dan struktur kalimatnya. Interaksi dengan bahasa dan budaya lain turut memengaruhi perkembangan dialek ini, menciptakan kekayaan linguistik yang unik. Namun, modernisasi dan pengaruh globalisasi menjadi tantangan besar dalam pelestarian bahasa ini, terutama di kalangan generasi muda (Udin et al., 1998). Dialek api dituturkan oleh masyarakat yang tinggal di daerah Pesisir Provinsi Lampung.



Gambar 5. Peta Provinsi Lampung (Bappeda Lampung, 2018).

Berdasarkan Gambar 5, yang menunjukkan peta administratif Provinsi Lampung, dapat diamati dengan jelas sebaran dialek Bahasa Lampung. Dialek terbagi ke dalam dua subvarian utama, yaitu Pepadun A dan Pepadun O, yang masing-masing memiliki persebaran wilayah dan ciri fonetik yang berbeda. Dialek Pepadun A, yang dikenal lebih konservatif dan mempertahankan struktur fonetik tradisional, banyak ditemukan di wilayah bagian utara dan barat Provinsi Lampung. Wilayah utama penutur dialek ini mencakup Kabupaten Way Kanan, Lampung Utara, dan Tulang Bawang Barat. Selain itu, bagian barat dan utara Kabupaten Lampung Tengah juga menunjukkan karakteristik fonetik yang sejalan dengan Pepadun A. Bahkan, sebagian wilayah utara Kabupaten Pringsewu juga terpengaruh oleh dialek ini, meskipun berada di zona transisi antara Pepadun dan Saibatin.

Sementara itu, dialek Pepadun O, yang lebih dinamis dan terbuka terhadap pengaruh fonetik luar, tersebar di wilayah yang lebih ke tengah dan timur provinsi. Dialek ini digunakan secara luas oleh masyarakat di Kabupaten Tulang Bawang, Mesuji, dan Lampung Timur, serta bagian timur dan selatan dari Kabupaten Lampung Tengah. Sebagian penutur di Kota Metro, khususnya yang berasal dari keluarga Pepadun, juga menggunakan varian ini dalam komunikasi sehari-hari. Karakteristik fonetik dari Pepadun O mencakup tempo bicara yang lebih cepat, intonasi yang lebih bervariasi, serta kecenderungan perubahan bunyi vokal tunggal menjadi rangkaian dua vokal dalam satu suku kata.

Perbedaan wilayah sebaran ini penting untuk diperhatikan dalam penyusunan sistem pengenalan suara berbasis Bahasa Lampung, karena variasi antar dialek dapat berpengaruh langsung terhadap akurasi hasil transkripsi suara ke teks. Sistem harus dilatih dengan data yang representatif dari masing-masing subdialek agar mampu mengenali pola bunyi yang khas dari tiap wilayah.

2.6.Karakteristik Fonetik Bahasa Lampung Dialek Pepadun

Bahasa Lampung dialek Pepadun merupakan salah satu varian bahasa daerah yang memiliki struktur bunyi yang khas dan cukup rumit. Beberapa bunyi khas dalam dialek ini adalah gabungan bunyi hidung dan bunyi letupan, seperti bunyi "mb", "nd", dan "ngg", yang sering muncul di awal atau tengah kata. Selain itu, juga terdapat bunyi hentakan pendek di tenggorokan yang muncul tiba-tiba, terutama di akhir atau tengah kata, yang dapat membedakan arti suatu kata dengan kata lain (Roveneldo, 2015). Ciri lain yang menonjol adalah variasi vokal, termasuk kecenderungan perubahan vokal menjadi dua bunyi sekaligus terutama di akhir kata.

Pola tekanan suara atau intonasi juga menjadi bagian penting dalam pengucapan Bahasa Lampung dialek Pepadun. Tekanan pada suku kata tertentu bisa menandai penekanan makna atau bahkan membedakan antara kata perintah dan pernyataan biasa. Selain itu, perubahan satu bunyi kecil saja dapat menyebabkan perubahan makna yang cukup besar dalam percakapan dan komunikasi sehari-hari, cara pengucapan yang cepat atau lambat juga dapat mempengaruhi arti, terutama dalam kalimat yang mengandung unsur adat atau formalitas.

Keunikan lain dari dialek Pepadun adalah adanya pengaruh lingkungan sosial dan geografis yang menyebabkan sedikit perbedaan dalam pelafalan antar wilayah. Misalnya, pelafalan oleh penutur dari wilayah pesisir cenderung lebih cepat dengan intonasi menurun, sementara penutur dari daerah pedalaman memiliki gaya bicara yang lebih tenang dan stabil. Variasi ini menyebabkan sistem pengenalan suara harus cukup fleksibel untuk dapat mengenali gaya bicara yang berbeda dari berbagai penutur (Hanawalt, 2007).

2.7.Implementasi Transkripsi Suara Bahasa Lampung ke Teks

Bahasa Lampung adalah salah satu dari ratusan bahasa lokal di Indonesia. Penelitian tentang Bahasa Lampung masih terbatas, terutama dalam konteks

pengolahan bahasa alami. Sebuah studi tentang struktur Bahasa Lampung dan perbandingannya dengan bahasa Indonesia dapat memberikan wawasan yang berharga untuk pengembangan algoritma transkripsi bahasa. Penggunaan *Perceptual Linear Predictive* (PLP) dan *Hidden Markov Model* (HMM) dalam konteks ini melibatkan beberapa langkah:

1. Pengumpulan Data: Mengumpulkan data bicara dalam Bahasa Lampung, yang kemudian digunakan untuk pelatihan model *Hidden Markov Model* (HMM).
2. Pra-pengolahan: Menggunakan *Perceptual Linear Predictive* (PLP) untuk menganalisis sinyal bicara dan ekstraksi fitur yang relevan yang mencerminkan karakteristik unik dari Bahasa Lampung.
3. Modeling dengan *Hidden Markov Model* (HMM): Membangun model *Hidden Markov Model (HMM)* yang menggunakan fitur-fitur dari *Perceptual Linear Predictive* (PLP) sebagai pengamatan. Model ini dilatih untuk mengenali pola spesifik dalam data bicara Bahasa Lampung.
4. Pengenalan dan Transkripsi: Menggunakan model *Hidden Markov Model* (HMM) yang telah dilatih untuk mengenali dan menerjemahkan bicara Bahasa Lampung ke dalam teks secara otomatis.

2.8. *Word Error Rate* (WER)

WER digunakan untuk mengukur tingkat kesalahan pengenalan kata dalam hasil transkripsi. Parameter ini menghitung jumlah kesalahan yang terjadi pada level kata, yang meliputi: kata yang salah dikenali (*substitusi)*, kata yang berlebih (*insersi)*, dan kata yang hilang (hapus) dibandingkan dengan jumlah total kata dalam teks referens (Goldwater & Manning, 2010). Nilai WER dapat dihitung menggunakan parameter dapat dilihat pada Persamaan (13).

$$WER = \frac{S+I+D}{N} \quad (13)$$

S : jumlah kata yang salah dikenali (*substitusi*)

I : jumlah kata yang berlebih (*insersi*)

D : jumlah kata yang hilang (hapus)

N : jumlah total kata dalam teks referensi

WER digunakan untuk mengevaluasi apakah model STT dapat mengenali fonem dan kata dalam kalimat dengan baik. Jika WER tinggi, ini berarti model perlu diperbaiki, misalnya dengan melatih model dengan lebih banyak data suara, meningkatkan teknik ekstraksi fitur atau menggunakan model akustik yang lebih baik.

2.9.*Character Error Rate* (CER)

CER digunakan untuk mengukur tingkat kesalahan pada level karakter dalam hasil transkripsi. Parameter ini menghitung jumlah kesalahan yang terjadi pada *level* karakter, meliputi: karakter yang salah dikenali (*substitusi*), karakter yang berlebih (*insersi*), dan karakter yang hilang (hapus) dibandingkan dengan jumlah total karakter dalam teks referensi (Nguyen et al., 2020). Dapat dilihat pada parameter dapat dilihat pada Persamaan (14).

$$CER = \frac{S+I+D}{C} \quad (14)$$

S : jumlah kata yang salah dikenali (*substitusi*)

I : jumlah kata yang berlebih (*insersi*)

D : jumlah kata yang hilang (hapus)

C : jumlah total kata dalam teks referensi

CER sangat berguna untuk mengevaluasi sejauh mana sistem dapat mengenali fonem yang mirip, terutama jika suatu bahasa memiliki huruf atau aksen khusus yang membedakannya dari bahasa lain. Jika nilai CER

tinggi, ini menunjukkan bahwa model STT memerlukan perbaikan, misalnya dengan meningkatkan teknik ekstraksi fitur, menambah jumlah data latih, atau memperbaiki model akustik.

III. METODOLOGI PENELITIAN

3.1. Tempat Penelitian

Penelitian dilaksanakan di Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung dan Rumah Tokoh Adat Dewan Perwakilan Penyimbang Adat Lampung Desa Kurungan Nyawa Kabupaten Pesawaran.

3.2. Waktu Penelitian

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

1. Tahap Penelitian Awal

a. Penentuan Tema, Analisis, dan Pengumpulan Studi Literatur

Tahapan ini bertujuan untuk membangun dasar konseptual dan persiapan awal penelitian. Fokus kegiatan pada tahap ini adalah identifikasi permasalahan penelitian, penentuan tema, serta pengumpulan studi literatur yang relevan. Studi literatur mencakup referensi terkait teknologi *Speech-to-Text*, metode *Perceptual Linear Prediction* (PLP) untuk ekstraksi fitur suara, dan penerapan *Hidden Markov Model* (HMM) dalam proses transkripsi.

b. Pengumpulan Dataset

Kegiatan ini berfokus pada pengumpulan dataset berupa rekaman audio Bahasa Lampung Dialek Pepadun A. Proses ini melibatkan rekaman suara penutur asli menggunakan perangkat perekam suara berkualitas tinggi dengan format standar seperti *.wav. Dataset* yang dikumpulkan mencakup variasi kata, kalimat, dan frasa umum yang digunakan dalam Bahasa Lampung.

c. Penyusunan Draft Proposal

Tahapan ini mencakup penyusunan proposal penelitian yang terdiri dari latar belakang, rumusan masalah, tujuan penelitian, metodologi, serta jadwal pelaksanaan kegiatan. Draft proposal ini menjadi acuan dalam melaksanakan seluruh rangkaian penelitian.

2. Tahap Penelitian Lanjut

Tahap penelitian lanjutan merupakan bagian penting dalam pelaksanaan penelitian karena berfokus pada implementasi teknis, pemrosesan data, dan perancangan model sistem pengenalan suara. Tahap ini bertujuan untuk mengubah data audio menjadi representasi tekstual menggunakan pendekatan berbasis model statistik *Hidden Markov Model* (HMM). Adapun tahapan yang dilakukan meliputi prapemrosesan data audio, ekstraksi fitur, perancangan dan pelatihan model HMM, serta pengujian kinerja sistem menggunakan dataset Bahasa Lampung.

Pada tahap prapemrosesan data, rekaman suara terlebih dahulu dibersihkan dari gangguan suara latar (*noise*) melalui proses *noise reduction*. Hal ini penting dilakukan agar informasi *linguistik* dari sinyal suara dapat diidentifikasi dengan lebih akurat oleh sistem. Setelah itu, sinyal audio dibagi menjadi potongan kecil menggunakan teknik *framing* agar dapat dianalisis dalam skala waktu yang lebih pendek. Tiap frame tersebut kemudian dilunakkan dengan menerapkan fungsi jendela (*windowing*), seperti *hamming window*, untuk meminimalkan efek *distorsi* pada batas antar *frame* yang dapat memengaruhi hasil ekstraksi fitur.

Tahap berikutnya adalah ekstraksi fitur menggunakan metode *Perceptual Linear Prediction* (PLP). Metode ini dipilih karena kemampuannya dalam meniru proses pendengaran manusia dengan lebih baik. PLP bekerja dengan mengestimasi karakteristik spektrum dari sinyal suara dan menghasilkan koefisien numerik yang merepresentasikan informasi fonetik yang terkandung di dalamnya. Koefisien PLP ini kemudian digunakan sebagai representasi fitur utama yang akan digunakan sebagai input pada model HMM.

Setelah fitur diperoleh, langkah selanjutnya adalah merancang model *Hidden Markov Model* (HMM). HMM merupakan model statistik yang banyak digunakan dalam pengenalan pola sekuensial, termasuk dalam sistem pengenalan suara. Dalam konteks ini, HMM digunakan untuk memodelkan pola sekuensial dari fitur PLP yang merepresentasikan ucapan manusia. Setiap model HMM dilatih untuk mengenali pola dari satu kalimat tertentu dalam Bahasa Lampung, sehingga pada tahap pelatihan, dilakukan proses pemetaan antara urutan fitur PLP dengan label kalimat yang sesuai. Proses pelatihan ini dilakukan secara iteratif untuk menemukan parameter optimal dari model.

Selanjutnya adalah pengujian dataset, yaitu menguji kinerja model HMM yang telah dilatih dalam mengenali data suara. Dataset pengujian terdiri dari rekaman suara Bahasa Lampung yang berbeda dari data pelatihan. Pada tahap ini, sistem akan menerima input berupa sinyal suara dan menghasilkan output berupa teks dalam Bahasa Lampung. Hasil pengenalan dibandingkan dengan transkripsi manual (*ground truth*) untuk menghitung metrik evaluasi seperti *Word Error Rate* (WER), *Character Error Rate* (CER), dan akurasi prediksi.

3. Tahap Evaluasi

Tahap ini bertujuan untuk menganalisis performa model dan menyusun hasil penelitian.

a. Evaluasi Hasil

Mengukur akurasi model menggunakan:

- *Word Error Rate* (WER): Menghitung kesalahan pada level kata.
- *Character Error Rate* (CER): Menghitung kesalahan pada level karakter.

Tahap evaluasi ini bertujuan untuk mengukur sejauh mana model yang telah dibangun mampu mengenali pola suara dengan baik dan akurat dalam kondisi nyata.

Tabel 3. 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 dan Pengujian Dataset																								
Evaluasi	Evaluasi Hasil																								
	Penyusunan Draft Hasil dan Seminar Hasil Penelitian																								

3.3.Data Suara

Data utama dalam penelitian ini berupa rekaman suara dalam Bahasa Lampung Dialek Pepadun A yang terdiri dari kalimat yang diucapkan oleh penutur asli. Proses pengumpulan data melibatkan partisipasi dari 4 penutur asli yang tinggal di wilayah Pepadun meliputi Kabupaten Pesawaran. Masing-masing penutur diminta untuk membaca kalimat-kalimat yang telah disiapkan. Setiap rekaman suara dilakukan dengan *mikrofon* kualitas tinggi dan disimpan dalam format *wav.* Suara direkam menggunakan frekuensi *sampling 16 kHz*. Kalimat penutur dapat dilihat pada Tabel 4.

Tabel 4. Kalimat Bahasa Indonesia dan Bahasa Lampung.

No	Kalimat dalam Bahasa Indonesia	Kalimat dalam Bahasa Lampung
1	Ana membeli kopi di warung	Ana ngebeli kupi di warung
2	Kakak bermain sepak bola di lapangan	Daing main bal di lapangan
3	Bibi mencuci pakaian di sungai	Lita mupoh pakaian di duwai
4	Bapak mencari ikan di sungai dengan jaring	Buyah ngunut iwa di duwai makai jaring
5	Kami membaca buku di ruang tamu	Sikam ngebaca buku di ruang tamu
6	Dede menanam sayur di kebun	Dede nanom sayuran di huma
7	Dia memasak nasi goreng di dapur	Dia ngeguai kan guring di dapor
8	Adik memelihara kucing di rumah	Adek ngisik kucing di mahan
9	Ibu mengajar anak membaca di kamar.	Ibuk ngelajar sanak ngebaca di kamar
10	Saya menyiram bunga di halaman	Sikam nyiram bunga di pengadapan
11	Paman menanam padi di sawah	Ama nanom pari di sabah
12	Kakek menggendong cucu di halaman	Atu ngebabay umpu di pengadapan

No	Kalimat dalam Bahasa Indonesia	Kalimat dalam Bahasa Lampung
13	Kucing tidur di atas kursi	Kucing pedom di lammung kerasi
14	Nenek membawa keranjang di pasar	Ateh ngusung keranjang di pasar
15	Bibi membersihkan rumah di pagi hari	Mak wan ngeberasihko mahan jeno pagi
16	Nenek menganyam tikar di bawah pohon	Ateh ngayam sulan di debah batang
17	Ayah mengendarai mobil baru	Ayah ngusung mubil baru
18	Ibu mencuci pakaian kotor	Umik mupoh pakaian kamak
19	Adik memakan buah mangga	Adek nganik buah manga
20	Kakak bermain sepeda di lapangan	Daing main sepida di lapangan

Kalimat yang digunakan memiliki struktur yang sederhana dan umum, yaitu mengikuti pola subjek predikat objek keterangan, seperti **ana membeli kopi di warung** dan **paman menanam padi di sawah**. Struktur ini dipilih untuk memudahkan penyesuaian dalam proses pelabelan dan ekstraksi fitur suara, serta meminimalkan potensi kesalahan interpretasi dalam pengucapan. Selain itu, konten kalimat didesain untuk mencerminkan aktivitas sehari-hari yang familiar bagi penutur, sehingga menghasilkan pelafalan yang konsisten.

Setiap penutur mengucapkan kalimat sebanyak 10 kali, total kalimat yang terekam dari seluruh responden berjumlah 800 kalimat. Pemilihan penutur dilakukan dengan memperhatikan faktor usia, jenis kelamin, dan keaslian dialek untuk memastikan variasi data yang *representatif* dalam konteks pengenalan suara Bahasa Lampung. Penutur terdiri dari 2 laki-laki berusia 46 dan 49 tahun, serta 2 perempuan berusia 35 dan 45 tahun, yang semuanya merupakan penutur asli Bahasa Lampung dialek A.

3.4.Alat dan Perangkat Lunak

Pada proses pengumpulan data, ekstraksi fitur, pemodelan, dan evaluasi dalam penelitian ini berjalan lancar dan menghasilkan data yang valid, diperlukan penggunaan alat dan perangkat lunak yang sesuai. Penelitian ini menggunakan alat dan perangkat lunak sebagai berikut:

1. **Alat Perekaman Audio**

Mikrofon berkualitas tinggi untuk merekam suara dengan jelas. *SARAMONIC Blink* 500 B2 dengan frekuensi sampling *16 kHz*.

2. ***Software***

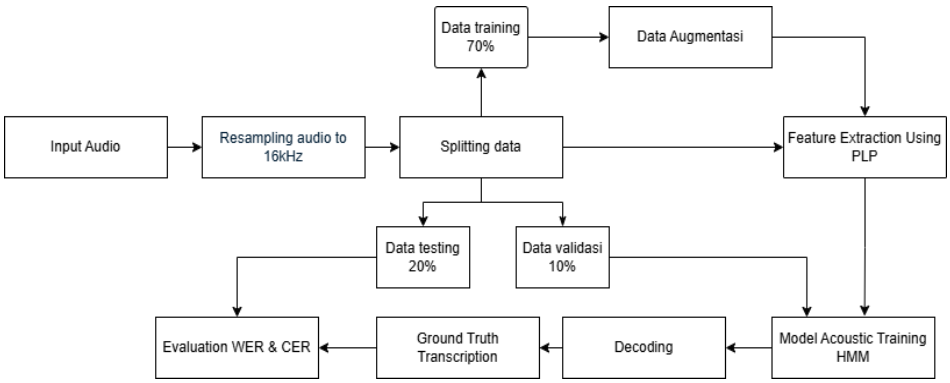
- Sistem Operasi menggunakan *Windows 11 Home Single Language*.
- Penelitian ini menggunakan bahasa pemrograman *Python 3.11.12*.
- Library yang digunakan *librosa soundfile*, *openpyxl*, *audiomentations*, *python_speech_features*, *librosa scipy*, *hmmlearn* dan *jiwer*.

3. ***Hadware***

- Prosesor : *Intel Core i7 1th Gen Intel(R) Core (TM) i7-11800H @ 2.30GHz*
- Memori: *16 GB RAM*
- Penyimpanan: *SSD 500 GB*
- Kartu Grafis: *NVIDIA GPU*

3.5.Alur Kerja Penelitian

Model transkripsi suara ke teks ini dikembangkan dengan menerapkan metode *Perceptual Linear Prediction* (PLP) sebagai teknik ekstraksi fitur dan *Hidden Markov Model* (HMM) sebagai model akustik, Alur penelitian dapat dilihat pada Gambar 6.



Gambar 6. Alur Penelitian.

1. Input Audio

Tahap awal dalam penelitian ini adalah pengumpulan data berupa sinyal suara berformat *.wav*. Data ini digunakan sebagai input utama dalam proses pengenalan ucapan. File audio tersebut merepresentasikan berbagai kalimat berbahasa daerah yang akan diolah lebih lanjut.

2. *Resampling* Audio

Tahap selanjutnya adalah proses resampling audio menjadi 16kHz**.** Tujuan dari tahap ini adalah untuk menyeragamkan laju sampel sinyal suara agar sesuai dengan standar pengolahan sinyal dan ekstraksi fitur PLP.

3. *Splitting* Data

Dataset kemudian dibagi menjadi 3 yaitu data latih (*training*), validasi, dan pengujian (*testing*) dengan proporsi tertentu, 70% data *training*, 10% data validasi dan 20% data *testing*. Pembagian ini penting agar model dapat dilatih secara efektif, divalidasi selama proses pelatihan, dan dievaluasi performanya secara objektif di akhir.

4. Data *Augmentasi*

Untuk meningkatkan keragaman dan jumlah data pelatihan, dilakukan augmentasi audio pada data train sebanyak dua kali (*repeat=2*) menggunakan teknik *TimeStretch*, *PitchShift*, *AddGaussianNoise*, *Shift*, dan *Gain*, sehingga jumlah data bertambah dari 560 menjadi 1.680, dengan

tujuan agar model mampu mengenali variasi suara yang lebih luas dan mempertahankan performa dalam kondisi yang berbeda-beda. Pada penelitian ini data yang di augmentasi adalah data *train* (latih).

5. *Feature Extraction* PLP

Setelah data audio disiapkan, dilakukan ekstraksi fitur menggunakan metode *Perceptual Linear Prediction* (PLP). PLP digunakan karena mampu meniru persepsi pendengaran manusia, sehingga diharapkan dapat meningkatkan akurasi dalam pengenalan ucapan. Hasil ekstraksi fitur ini berupa representasi numerik dari sinyal audio yang menjadi input bagi model akustik.

6. Model *Acoustic Training* HMM

Model akustik dibangun menggunakan algoritma *Hidden Markov Model* (HMM). Model ini dilatih menggunakan fitur hasil ekstraksi dari data pelatihan. HMM dipilih karena kemampuannya dalam memodelkan urutan waktu (*temporal sequence*) seperti dalam ucapan. Setiap kalimat diwakili oleh satu HMM.

7. *Decoding*

Setelah model dilatih, proses *decoding* dilakukan terhadap data uji untuk menghasilkan prediksi kalimat. Prediksi dilakukan dengan memilih model HMM yang memberikan skor probabilitas tertinggi terhadap fitur input. Ini merupakan bentuk sederhana dari *decoding* berbasis *maximum likelihood*.

8. *Ground Truth Transcription*

Ground truth adalah transkripsi asli dari kalimat dalam data uji. *Ground truth* ini menjadi acuan untuk mengevaluasi sejauh mana hasil prediksi (*decoding*) model sesuai dengan kenyataan.

9. *Evaluation* WER dan CER

Evaluasi dilakukan dengan menghitung *Word Error Rate* (WER) dan *Character Error Rate* (CER) antara hasil *decoding* dan *ground truth*. WER

digunakan untuk mengukur kesalahan pada tingkat kata, sedangkan CER pada tingkat karakter. Kedua metrik ini umum digunakan dalam evaluasi sistem pengenalan suara.

V. EVALUASI

5.1.Simpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai transkripsi suara ke teks Bahasa Lampung dengan metode *Perceptual Linear Prediction* (PLP) dan *Hidden Markov Model* (HMM), maka dapat disimpulkan beberapa hal sebagai berikut:

1. Dataset penelitian terdiri dari 800 file audio yang merepresentasikan 20 kalimat unik dalam Bahasa Lampung Dialek Pepadun A. Setiap kalimat diucapkan oleh beberapa penutur berbeda dan disimpan dalam subfolder tersendiri. Seluruh data direkam dalam format *.wav* dan dilakukan proses *resampling* menjadi *16 kHz* untuk memenuhi kebutuhan ekstraksi fitur akustik standar pada sistem pengenalan suara.
2. Proses *pra-pemrosesan* dan augmentasi data berhasil meningkatkan keragaman sinyal suara. Teknik *TimeStretch*, *PitchShift*, *AddGaussianNoise*, *Shift*, dan *Gain*. Penerapan augmentasi ini bertujuan untuk menghasilkan variasi ucapan sehingga model dapat dilatih untuk mengenali beragam kondisi pelafalan dan meningkatkan kemampuan generalisasi model. Pada penelitian ini data yang di augmentasi adalah data *train* (latih), augmentasi data train dilakukan 2 kali repeat augmentasi dari 560 data train original menjadi 1680 data audio kalimat.
3. Ekstraksi fitur PLP berhasil menghasilkan representasi akustik suara yang lebih stabil dan menyerupai persepsi pendengaran manusia.

4. Proses PLP yang mencakup transformasi ke *skala Bark*, *equal-loudness pre-emphasis*, *logarithmic compression*, serta *linear prediction* telah diimplementasikan secara utuh dan digunakan untuk membangun input pelatihan model HMM.
5. Model akustik HMM dilatih untuk setiap kalimat secara terpisah menggunakan fitur PLP. *Decoding* dilakukan dengan memilih kalimat dengan skor probabilitas tertinggi berdasarkan prinsip *maximum likelihood*. Penelitian ini menggunakan pendekatan *fixed sentence recognition* (pengenalan kalimat tetap) dan belum melakukan *decoding* kata per kata atau fonem per fonem.
6. Evaluasi performa model menunjukkan hasil yang cukup baik dan stabil. Nilai *Word Error Rate* (WER) rata-rata yang diperoleh adalah 17,79%, sedangkan *Character Error Rate* (CER) rata-rata adalah 14,60%. Selain itu, nilai terbaik (WER terbaik dan CER terbaik) sebesar 0,00% berhasil dicapai pada beberapa kalimat tertentu, menunjukkan bahwa sistem mampu mengenali beberapa kalimat dengan sempurna.
7. Kesalahan prediksi umumnya ditemukan pada kalimat yang memiliki kemiripan fonetik atau pada kualitas audio yang kurang optimal. Hal ini menunjukkan bahwa model masih dapat ditingkatkan, terutama dengan memperbanyak variasi data latih dari segi penutur maupun intonasi pelafalan.
8. Secara keseluruhan, model yang dikembangkan berbasis PLP dan HMM telah mampu mengenali kalimat Bahasa Lampung dialek Pepadun A dengan akurasi yang cukup tinggi. Hasil penelitian ini membuktikan bahwa pendekatan ini efektif untuk pengembangan sistem pengenalan suara berbasis bahasa daerah.

5.2.Saran

Berdasarkan hasil penelitian dan analisis yang telah dilakukan, terdapat beberapa saran yang dapat diberikan untuk pengembangan dan penelitian lanjutan di masa depan, sebagai berikut:

1. Untuk meningkatkan kemampuan generalisasi sistem, disarankan agar jumlah kalimat diperbanyak, kalimat tidak ulang, mencakup variasi struktur kalimat dan kosakata. Melibatkan lebih banyak penutur dengan latar belakang usia, jenis kelamin, dan logat berbeda dapat membantu model mengenali karakteristik fonetik Bahasa Lampung secara lebih komprehensif.
2. Perekaman data dilakukan dengan kualitas audio yang baik di lingkungan yang minim gangguan. Jika terdapat pengulangan kalimat, maka kalimat tersebut sebaiknya diucapkan oleh penutur yang berbeda untuk memperkaya variasi fonetik dan meningkatkan kemampuan generalisasi model terhadap karakter suara yang beragam.
3. Penelitian ini menggunakan pendekatan *fixed sentence classification* dengan satu model HMM per kalimat. Untuk meningkatkan fleksibilitas sistem, disarankan menggunakan pendekatan berbasis fonem atau kata agar sistem dapat mengenali kalimat baru yang belum pernah dilatih secara spesifik.
4. PLP terbukti bekerja cukup baik, namun kesalahan tetap muncul pada kalimat-kalimat dengan fonetik mirip. Oleh karena itu, direkomendasikan untuk menggabungkan PLP dengan fitur lain seperti MFCC atau *pitch* untuk menghasilkan representasi suara yang lebih kaya dan akurat.
5. Evaluasi saat ini menggunakan pembagian data secara *stratified sampling* (*train, validation, test*). Penggunaan *stratified sampling* dapat membantu memperoleh gambaran performa model yang lebih stabil dan akurat terhadap seluruh distribusi data.
6. *Decoding* berbasis *maximum likelihood* sudah cukup efektif, integrasi metode *decoding* yang lebih canggih seperti *Viterbi decoding* atau penggabungan model bahasa berbasis *n-gram* atau *Connectionist Temporal Classification* (CTC) dapat lebih meningkatkan akurasi sistem.

7. Penelitian ini berfokus pada dialek Pepadun A. Penelitian lanjutan diharapkan memperluas cakupan ke dialek lain dalam Bahasa Lampung untuk memperkaya kemampuan model dalam memahami variasi bahasa daerah.

DAFTAR PUSTAKA

Ali N.F, R. H. (2022). A Review on Automatic Speech Recognition: Indonesian Speech Recognition. *RC EEE 2021 Bangkok Thailand*.

Andina, E. (2023). Implementasi dan Tantangan Revitalisasi Bahasa Daerah di Provinsi Lampung. *Jurnal Masalah-Masalah Sosial, 14 No 1, 15-30*.

Bappeda Provinsi (2008). Peta Kabupaten Daerah/Kota Provinsi Lampung, *https://bappeda.lampungprov.go.id/berkas/uploads/I9Zwcw3s12SsOkctE6cyUQTaQKYMUpOZ5S2D2QdK.pdf*. Diakses tanggal 30 April 2024.

Cazau D, N. G. (2017). Investigation on the Use of Hidden-Markov Models in Automatic Transcription of Music. *Journal arXiv, 1704.03711v1 Page 1-27*.

Dahl, G. E. (2012). Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition. *IEEE Transactions on Audio, Speech, and Language Processing, 20 No 1, 30-39.*

Dave, N. (2013). Feature extraction methods LPC, PLP and MFCC in speech recognition. *International journal for advance research in engineering and technology Volume 1, Issue VI, July 2013, 1-4.*

Davis, S. B., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 28(4), 357–366.*

Goldwater, S., Jurafsky, D., & Manning, C. D. (2010). Which words are hard to recognize? Prosodic, lexical, and disfluency factors that increase speech recognition error rates. *Speech Communication, 52(3), 181–200*

Hanawalt, C. (2007). Bitter or Sweet? The Vital Role of Sociolinguistic Survey in Lampungic Dialectology. *Studies in Philippine Languages and Cultures*, 16, 31–50

Hermansky, H. (1990). Perceptual Linear Predictive (PLP) Analysis of Speech. *The Journal of the Acoustical Society of America, 87 No4, 1738-1752*.

Huang, X., Acero, A., & Hon, H.-W. (2001). Spoken Language Processing: A Guide to Theory, Algorithm, and System Development. *Prentice Hall.*

Kumar, A., Dua, M., & Choudhary, T. (2014). Continuous hindi speech recognition using monophone based acoustic modeling. International *Journal of Computer Applications (ICACEA), 1, 15-19.*

Kurzekar, N. Gaikwad, & M. Patil. (2014). A Comparative Study of Feature Extraction Techniques for Speech Recognition System. *International Journal of Computer Applications, vol. 89, no. 17, pp. 12–16.*

Liu, C. R., Qu, ., & Yang, X. K. (2019). Long short term memory networks weighted prediction error for far-field speech recognition. Proceedings of 2019. *IEEE 8th Joint International Information Technology and Artificial Intelligence Conference, ITAIC.*

Maseri & Mamat (2020). Performance Analysis of Implemented MFCC and HMM-based Speech Recognition System, *International Journal of Computer Applications, vol. 177, no. 25, pp. 1–6*.

Mon, S. M., & Tun, H. M. (2015). Speech-To-Text conversion (STT) system using hidden markov model (HMM). *International Journal of Scientific &technology Research, 4(6), 349–351.*

Nguyen, M. H., Luong, M. T., & Quoc, V. L. (2020). Improving Character Error Rate in Speech Recognition via Hybrid Training. *IEEE Access, 8, 215476–215486.*

Octodhiyanto, I. (2023). Suku Jawa Jadi Mayoritas di Lampung. Diakses dari *http://www.harianhaluan.com/news/109588208/ternyata-ini-alasan-suku-jawa- jadi-mayoritas-di-lampung*

Paul, A. K., Das, D., & Kamal, M. M. (2009). Bangla speech recognition system using LPC and ANN. Proceedings of the Seventh International Conference on Advances in Pattern Recognition, *IEEE, (ICAPR 2009), 171–174*.

Park, D. S., Chan, W., Zhang, Y., Chiu, C. C., Zoph, B., Cubuk, E. D., & Le, Q. V. (2019). SpecAugment: A simple data augmentation method for automatic speech recognition. *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH), 2613–2617*.

Peddinti, V., Povey, D., & Khudanpur, S. (2015). Audio augmentation for speech recognition. Proceedings of Interspeech 2015, 3586–3589. *International Speech Communication Association*.

Putri, N. W. (2018). Pergeseran Bahasa Daerah Lampung pada Masyarakat Kota Bandar Lampung. *Jurnal Penelitian Linguistik, 12 No 1, 15-25*.

Rabiner, L. R., & Juang, B.H. (1993). Fundamentals of speech recognition. Englewood Cliffs, *NJ: Prentice Hall*.

Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE, 77 No 2, 257–286.*

Rabiner, L., & Schafer, R. (2011). Theory and applications of digital speech processing (1st ed.). *Pearson*.

Rismawati, K. (2022). Analisis Part of Speech Tagging dalam Bahasa Lampung Dialek A dengan Metode HMM (Hidden Markov Model). *Jurnal Teknologi Pintar*, 2 No 10, 1-9.

Roveneldo, R. (2015). Korespondensi Bunyi Bahasa Lampung Dialek Tulangbawang. *Jurnal Madah*, 6(1), 91–100

Roveneldo. (2022). Perbedaan kosakata rumah panggung antara bahasa Lampung dialek O dan dialek A dalam kajian dialektologi. *International Journal of Scientific & Technology Research, 11(6), 3586–3589.*

Salamon, J., Bello, J. P. (2017). Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification. *IEEE Signal Processing Letters, 24(3), 279–283.*

Sechidis, K., Tsoumakas, G., & Vlahavas, I. (2011). On the stratification of multi-label data. Machine Learning and Knowledge Discovery in Databases, *ECML PKDD, Vol. 6912, pp. 145–158. Springer.*

Suyanto, S., & Amin, M. F. (2017). Pemakaian Bahasa Jawa di Provinsi Lampung Bedasar Data Sensus Penduduk 2010. Nusa: *Jurnal Ilmu Bahasa dan Sastra, 12(3), 81–92*

Stadermann & Rigoll, "A Hybrid SVM/HMM Acoustic Modeling Approach to Automatic Speech Recognition," *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2004, pp. 577–580

Udin, H. N., Akhyar, W., Wetty, N. N., Rusminto, N. E., & Sanusi, A. E. (1998). Sastra lisan Lampung dialek Pubiyan.*Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan*. Jakarta.

Walker, D. F. (1976). A grammar of the Lampung language: The Pesisir dialect of Way Lima. Canberra: Pacific Linguistics, *The Australian National University.*

Zainal, A. G., & Kagungan, D. (2023). Analisis proses akomodasi komunikasi dan adaptasi masyarakat pendatang dan pribumi di Kabupaten Lampung Barat (Skripsi). *Universitas Lampung*.

Zainal, Z. (2023). Struktur fonologis dialek Pepadun dalam perspektif linguistik. *Jurnal Kajian Bahasa dan Budaya, 8(2), 22–34.*

Zalmansyah, A. (2019). Bahasa Lampung di Kalangan Anak Muda Lampung. *In Kelasa, Vol. 14, No. 2, Desember 2019: 145—158. scholar.archive.org*

Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2017). Understanding deep learning requires rethinking generalization. *In Proceedings of the 5th International Conference on Learning Representations* (ICLR 2017)