



**STUDI EKSPERIMENTAL LIP-READING  
VOKAL BAHASA INDONESIA MENGGUNAKAN PENDEKATAN  
VISUAL-ONLY DAN MULTIMODAL AUDIO-VISUAL**

**TUGAS AKHIR  
SKRIPSI**

**UNIVERSITAS  
MERCU BUANA**  
**FIRMAN AL RAHMAT**  
**41523110025**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2026**



**STUDI EKSPERIMENTAL LIP-READING  
VOKAL BAHASA INDONESIA MENGGUNAKAN PENDEKATAN  
VISUAL-ONLY DAN MULTIMODAL AUDIO-VISUAL**

**TUGAS AKHIR  
SKRIPSI**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana**

**FIRMAN AL RAHMAT**  
**41523110025**  
**MERCU BUANA**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2026**

## HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : FIRMAN AL RAHMAT

NIM : 41523110025

Program Studi : Fakultas Ilmu Komputer / Teknik Informatika

Menyatakan dengan sebenar-benarnya bahwa Tugas Akhir Berjudul:

“Studi Eksperimental Lip-Reading Vokal Bahasa Indonesia Menggunakan Pendekatan Visual-Only dan Multimodal Audio-Visual” adalah hasil karya saya sendiri, tidak mengandung unsur plagiarisme, pelanggaran hak cipta atau konten ilegal dalam bentuk apapun dan tidak melanggar hukum atau pihak manapun.

Apabila di kemudian hari ditemukan pelanggaran terhadap pernyataan ini, saya bersedia menanggung seluruh konsekuensi hukum dan membebaskan Universitas Mercu Buana dari segala bentuk tuntutan hukum dan saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Demikian pernyataan ini saya buat dengan sebenar-benarnya untuk digunakan sebagaimana mestinya.

UNIVERSITAS  
MERCU BUANA

Jakarta, 22 Januari 2026

  
METERAI  
TEMPAN  
107FANX278567419  
Firman Al Rahmat.

**PERNYATAAN SIMILARITY CHECK**  
**/SIMILARITY CHECK STATEMENT**

Saya yang bertanda tangan dibawah ini menyatakan, bahwa karya ilmiah yang ditulis oleh  
*/The undersigned, hereby declare that the scientific work written by*

Nama /Name : FIRMAN AL RAHMAT  
NIM /Student ID Number : 41523110025  
Program Studi /Study Program : Teknik Informatika

Dengan Judul Tugas Akhir

*/The title:*

**“Studi Eksperimental Lip-Reading Vokal Bahasa Indonesia Menggunakan Pendekatan Visual-Only dan Multimodal Audio-Visual”**

telah dilakukan pengecekan *similarity* dengan sistem Turnitin pada tanggal:

*/Similarity checks have been carried out with the Turnitin system on the date:*

3 Januari 2026

dengan nilai persentase sebesar :

*/with a percentage value of:*

15%

dinyatakan memenuhi standar sesuai dengan ketentuan berlaku di **Fakultas Ilmu Komputer** Universitas Mercu Buana. */declared to meet standards in accordance with applicable regulations at the Faculty of Computer Science, Universitas Mercu Buana.*

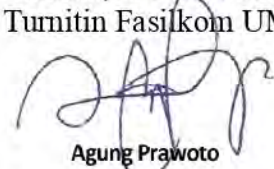
File hasil cek similarity turnitin:

*/Turnitin similarity report file*

[https://drive.google.com/file/d/1su8FukWI7AtuwrMYTb53fza-12jcY4zC/view?usp=drive\\_link](https://drive.google.com/file/d/1su8FukWI7AtuwrMYTb53fza-12jcY4zC/view?usp=drive_link)



Jakarta, 3 Januari 2026  
Admin Turnitin Fasilkom UMB



Agung Prawoto

**Agung Prawoto, S.Kom., B.Sc**

NIK : 322970503

## HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : FIRMAN AL RAHMAT  
NIM : 41523110025  
Fakultas/Program Studi : Fakultas Ilmu Komputer / Teknik Informatika  
Judul Tugas Akhir : Studi Eksperimental Lip-Reading Vokal Bahasa Indonesia Menggunakan Pendekatan Visual-Only dan Multimodal Audio-Visual

Telah berhasil dipertahankan pada sidang tanggal 22 Januari 2026 dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing



Muhaimin Hasanudin, S.T., M.Kom.

NIDN/NUPTK: 0420027508

UNIVERSITAS  
MERCU BUANA

Jakarta, 22 Januari 2026

Mengetahui,

Dekan Fakultas Ilmu Komputer



Dr. Bambang Jokonowo, S.Si., MTI  
NIDN/NUPTK: 0320037002

Ketua Program Studi  
Teknik Informatika



Dr. Hadi Santoso, S.Kom., M.Kom  
NIDN/NUPTK: 0225067701

## KATA PENGANTAR

Puji syukur kehadiran Tuhan yang Maha Esa, atas segala rahmat dan ridha-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang merupakan salah satu persyaratan kelulusan Program Studi Strata Satu (S1) pada jurusan Teknik Informatika, Universitas Mercu Buana.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari sempurna, karena kesempurnaan sejatinya hanya milik Tuhan yang Maha Esa. Oleh karena itu, saran dan masukan yang membangun senantiasa penulis terima dengan senang hati. Serta berkat dukungan, motivasi, bantuan, bimbingan, dan doa dari banyak pihak, penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Andi Adriansyah, M.Eng. selaku Rektor Universitas Mercu Buana.
2. Bapak Dr. Bambang Jokonowo, S.Si., MTI selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Dr. Hadi Santoso, S.Kom., M.Kom. selaku Ketua Program Studi Teknik Informatika Universitas Mercubwana.
4. Bapak Dosen Pembimbing Muhaimin Hasanudin, S.T., M.Kom. Yang telah memberikan pengarahan, motivasi, menyediakan waktu, tenaga, dan pikiran sehingga selama pembuatan Tugas Akhir ini terjadwal dengan baik.
5. Ibunda saya yang selalu mensupport dan mendukung saya selama menjalani masa studi sebagai mahasiswa Universitas Mercubwana. Dan doa tertinggi Untuk Alm Ayah saya semoga diberikan sisi terbaik di sisi-Nya.
6. Rekan kuliah, Rekan Kerja dan para kontributor dataset. Yang membantu dalam penelitian ini.

Akhir kata, penulis berharap semoga Tuhan yang Maha Esa membalas kebaikan dan selalu mencurahkan rahmat, hidayah, serta panjang umur kepada kita semua, aamiin. Terima Kasih.

Jakarta, 22 Januari 2026

Firman Al Rahmat

## HALAMAN PERSETUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

|                   |   |                                                                                                                             |
|-------------------|---|-----------------------------------------------------------------------------------------------------------------------------|
| Nama              | = | FIRMAN AL RAHMAT                                                                                                            |
| NIM               | = | 41523110025                                                                                                                 |
| Program Studi     | = | Teknik Informatika                                                                                                          |
| Judul Tugas Akhir | = | Studi Eksperimental Lip-Reading Vokal<br>Bahasa Indonesia Menggunakan Pendekatan<br>Visual-Only dan Multimodal Audio-Visual |

Demi Pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana Hak Bebas Royalty Non-eksklusif (Non-Exclusive Royalty-Free Right) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalty Non-eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

# MERCU BUANA

Jakarta, 22 Januari 2026

Yang Menyatakan



Firman Al Rahmat

**STUDI EKSPERIMENTAL LIP-READING  
VOKAL BAHASA INDONESIA MENGGUNAKAN PENDEKATAN  
VISUAL-ONLY DAN MULTIMODAL AUDIO-VISUAL**

**FIRMAN AL RAHMAT**

**ABSTRAK**

Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi sistem pengenalan vokal Bahasa Indonesia berbasis lip-reading menggunakan pendekatan visual-only dan multimodal audio-visual. Permasalahan utama dalam lip-reading adalah keterbatasan informasi visual yang menyebabkan terjadinya ambiguitas antar vokal, terutama pada vokal yang memiliki konfigurasi bibir yang serupa. Oleh karena itu, penelitian ini mengintegrasikan fitur visual geometris bibir dan fitur audio untuk meningkatkan akurasi klasifikasi. Metode penelitian yang digunakan adalah kuantitatif eksperimental dengan desain eksperimen komparatif. Dataset diperoleh melalui perekaman pengucapan vokal /a/, /i/, /u/, /e/, dan /o/ menggunakan webcam. Data video diproses dengan MediaPipe Face Mesh untuk mengekstraksi landmark bibir dan menghitung fitur geometris, sedangkan data audio dianalisis menggunakan fitur statistik dan spektral. Fitur visual dan audio kemudian digabungkan menggunakan pendekatan feature-level fusion. Proses klasifikasi dilakukan menggunakan algoritma Logistic Regression, Support Vector Machine, dan Random Forest dengan pembagian data latih dan data uji sebesar 80:20. Hasil penelitian menunjukkan bahwa pendekatan visual-only menghasilkan akurasi maksimum sebesar 48,48% pada model Random Forest. Sementara itu, pendekatan multimodal mampu meningkatkan akurasi secara signifikan hingga mencapai 98,67% pada model SVM dan 100% pada model Random Forest. Selain itu, nilai F1-score juga mengalami peningkatan yang konsisten pada seluruh model. Kesimpulan dari penelitian ini menunjukkan bahwa integrasi fitur audio dan visual secara efektif meningkatkan performa sistem lip-reading dalam mengklasifikasikan vokal Bahasa Indonesia. Pendekatan multimodal terbukti lebih unggul dibandingkan visual-only dan berpotensi dikembangkan lebih lanjut dalam sistem pengenalan ujaran berbasis audio-visual.

**Kata Kunci:** lip-reading, pengenalan vokal, multimodal audio-visual, machine learning, fitur geometris

# EXPERIMENTAL STUDY OF INDONESIAN VOWEL LIP-READING USING VISUAL-ONLY AND MULTIMODAL AUDIO-VISUAL APPROACHES

FIRMAN AL RAHMAT

## ABSTRACT

*This study aims to develop and evaluate an Indonesian vowel recognition system based on lip-reading using visual-only and multimodal audio-visual approaches. The main challenge in lip-reading lies in the limitation of visual information, which causes ambiguity among vowels, especially those with similar lip configurations. Therefore, this research integrates geometric visual features of the lips and audio features to improve classification accuracy. The research employs a quantitative experimental method with a comparative experimental design. The dataset was obtained through webcam recordings of vowel pronunciations /a/, /i/, /u/, /e/, and /o/. Video data were processed using MediaPipe Face Mesh to extract lip landmarks and compute geometric features, while audio data were analyzed using statistical and spectral features. Visual and audio features were combined using a feature-level fusion approach. Classification was performed using Logistic Regression, Support Vector Machine, and Random Forest algorithms with an 80:20 training-testing data split. The results show that the visual-only approach achieved a maximum accuracy of 48.48% using the Random Forest model. In contrast, the multimodal approach significantly improved performance, reaching 98.67% accuracy with SVM and 100% with Random Forest. The F1-score also consistently increased across all models. In conclusion, the integration of audio and visual features effectively enhances the performance of lip-reading systems for Indonesian vowel classification. The multimodal approach outperforms the visual-only method and demonstrates strong potential for further development in audio-visual speech recognition systems.*

**Keywords:** lip-reading, vowel recognition, multimodal audio-visual, machine learning, geometric features

## DAFTAR ISI

|                                                                  |      |
|------------------------------------------------------------------|------|
| HALAMAN SAMPUL .....                                             | 0    |
| HALAMAN JUDUL.....                                               | i    |
| HALAMAN PERNYATAAN KARYA SENDIRI .....                           | ii   |
| HALAMAN SURAT KETERANGAN HASIL UJI TURNITIN .....                | iii  |
| HALAMAN PENGESAHAN.....                                          | iv   |
| KATA PENGANTAR .....                                             | v    |
| HALAMAN PERSETUJUAN PUBLIKASI TUGAS AKHIR DI REPOSITORI UMB..... | vi   |
| ABSTRAK .....                                                    | vii  |
| ABSTRACT.....                                                    | viii |
| DAFTAR ISI.....                                                  | ix   |
| DAFTAR TABEL .....                                               | xi   |
| DAFTAR GAMBAR .....                                              | xii  |
| DAFTAR LAMPIRAN.....                                             | xiii |
| BAB I PENDAHULUAN .....                                          | 1    |
| 1.1 Latar Belakang .....                                         | 1    |
| 1.2 Perumusan Masalah .....                                      | 2    |
| 1.3 Tujuan Penelitian .....                                      | 3    |
| 1.4 Manfaat Penelitian .....                                     | 4    |
| 1.5 Batasan Penelitian .....                                     | 5    |
| 1.6 Sistematika Penulisan .....                                  | 6    |
| BAB II TINJAUAN PUSTAKA.....                                     | 10   |
| 2.1 Teori Utama .....                                            | 10   |
| 2.2 Teori Pendukung .....                                        | 19   |
| 2.3 Penelitian Terdahulu .....                                   | 38   |
| 2.4 Gap Penelitian .....                                         | 44   |
| BAB III METODE PENELITIAN.....                                   | 48   |
| 3.1 Pendekatan Penelitian .....                                  | 48   |
| 3.2 Desain Penelitian.....                                       | 49   |
| 3.3 Subjek Penelitian.....                                       | 51   |
| 3.4 Instrumen Penelitian.....                                    | 52   |
| 3.5 Teknik Pengumpulan Data.....                                 | 54   |
| 3.6 Teknik Analisis Data.....                                    | 57   |
| 3.7 Prosedur Penelitian.....                                     | 62   |

|                                   |                                              |     |
|-----------------------------------|----------------------------------------------|-----|
| 3.8                               | Evaluasi Hasil Penelitian.....               | 63  |
| 3.9                               | Timeline Penelitian .....                    | 64  |
| BAB IV HASIL DAN PEMBAHASAN ..... |                                              | 65  |
| 4.1                               | Dataset dan Preprocessing.....               | 65  |
| 4.2                               | Ekstraksi Fitur .....                        | 69  |
| 4.3                               | Exploratory Data Analysis (EDA) .....        | 75  |
| 4.4                               | Pelatihan Model Visual-Only.....             | 80  |
| 4.5                               | Pelatihan Model Multimodal Audio-Visual..... | 86  |
| 4.6                               | Perbandingan Visual dan Multimodal.....      | 89  |
| 4.7                               | Ringkasan Hasil Penelitian .....             | 93  |
| BAB V KESIMPULAN DAN SARAN.....   |                                              | 97  |
| 5.1                               | Kesimpulan .....                             | 97  |
| 5.2                               | Saran.....                                   | 98  |
| DAFTAR PUSTAKA .....              |                                              | 99  |
| LAMPIRAN.....                     |                                              | 103 |



## DAFTAR TABEL

|                                                                                                   |    |
|---------------------------------------------------------------------------------------------------|----|
| Tabel 2. 1 Vokal Monoftong dalam Bahasa Indonesia .....                                           | 11 |
| Tabel 2. 2 Karakteristik Akustik Vokal Bahasa Indonesia .....                                     | 14 |
| Tabel 2. 3 Perbedaan MediaPipe dan Dlib .....                                                     | 27 |
| Tabel 3. 1 Tabel Perangkat Lunak yang Dipakai .....                                               | 53 |
| Tabel 3. 2 Tabel Hyperparameter Tuning yang Digunakan.....                                        | 61 |
| Tabel 3. 3 Timeline Penelitian .....                                                              | 64 |
| Tabel 3. 1 Tabel Perangkat Lunak yang Dipakai .....                                               | 53 |
| Tabel 3. 2 Tabel Hyperparameter Tuning yang Digunakan.....                                        | 61 |
| Tabel 3. 3 Timeline Penelitian .....                                                              | 64 |
| Tabel 4. 1 Tabel Landmark Bibir MediaPipe Face Mesh .....                                         | 68 |
| Tabel 4. 2 Tabel Feature Level Fusion Multimodal.....                                             | 75 |
| Tabel 4. 3 Tabel Parameter Hasil Hyperparameter Tuning .....                                      | 83 |
| Tabel 4. 4 Perbandingan Kinerja Model Visual-Only dan Multimodal.....                             | 90 |
| Tabel 4. 5 Tabel Perbandingan Karakteristik Vokal: Teori Zenten (1989) dan Hasil Penelitian ..... | 95 |



## DAFTAR GAMBAR

|                                                                            |    |
|----------------------------------------------------------------------------|----|
| Gambar 2. 1 Matriks Vokal IPA Untuk Bahasa Indonesia .....                 | 11 |
| Gambar 2. 2 Workflow Visual Speech Recognition (VSR) .....                 | 17 |
| Gambar 2. 3 Citra Grey Scale .....                                         | 20 |
| Gambar 2. 4 Citra RGB.....                                                 | 21 |
| Gambar 2. 5 Blok Diagram Facial Landmark Detection (Pambudi, 2024) .....   | 22 |
| Gambar 2. 6 Ekstraksi Fitur Area Bibir dari Penelitian Aminah (2024) ..... | 23 |
| Gambar 2. 7 Facial Landmark Mediapipe 468 Titik.....                       | 24 |
| Gambar 2. 8 Fitur Geometrik Bibir di Penelitian Aminah (2024) .....        | 26 |
| Gambar 2. 9 Proses Feature Fusion di Penelitian Song (2022).....           | 36 |
| Gambar 3. 1 Arsitektur Penelitian.....                                     | 50 |
| Gambar 3. 2 Contoh Subjek yang Memenuhi Kriteria .....                     | 52 |
| Gambar 3. 3 Flowchart Pengumpulan dan pengolahan data.....                 | 55 |
| Gambar 3. 4 Struktur Folder Dataset .....                                  | 56 |
| Gambar 3. 5 Alur Ekstraksi Fitur Visual .....                              | 57 |
| Gambar 3. 6 Flowchart Analisis Multimodal.....                             | 58 |
| Gambar 3. 1 Arsitektur Penelitian.....                                     | 50 |
| Gambar 3. 2 Contoh Subjek yang Memenuhi Kriteria .....                     | 52 |
| Gambar 3. 3 Flowchart Pengumpulan dan pengolahan data.....                 | 55 |
| Gambar 3. 4 Struktur Folder Dataset .....                                  | 56 |
| Gambar 3. 5 Alur Ekstraksi Fitur Visual .....                              | 57 |
| Gambar 3. 6 Flowchart Analisis Multimodal.....                             | 58 |
| Gambar 4. 1 Distribusi Video Dataset .....                                 | 67 |
| Gambar 4. 2 Visualisasi Landmark Mulut .....                               | 69 |
| Gambar 4. 3 Hasil Extraksi Visual.....                                     | 71 |
| Gambar 4. 4 Distribusi Tiap Vokal.....                                     | 72 |
| Gambar 4. 5 Distribusi Fitur Vokal A.....                                  | 73 |
| Gambar 4. 6 Distribusi Fitur Vokal E .....                                 | 73 |
| Gambar 4. 7 Distribusi Fitur Vokal I .....                                 | 73 |
| Gambar 4. 8 Distribusi Fitur Vokal O.....                                  | 73 |
| Gambar 4. 9 Distribusi Fitur Vokal U.....                                  | 73 |
| Gambar 4. 10 Dataset Multimodal .....                                      | 75 |
| Gambar 4. 11 Distribusi Data per Kelas Vokal .....                         | 76 |
| Gambar 4. 12 Heatmap Kolerasi Fitur Multimodal .....                       | 76 |
| Gambar 4. 13 Rata-rata Nilai Fitur Per Kelas Vokal .....                   | 78 |
| Gambar 4. 14 Hasil Pengechekan Missing Value .....                         | 80 |
| Gambar 4. 15 Flowchart Pelatihan Model Visual-Only .....                   | 85 |
| Gambar 4. 16 Hasil Evaluasi Model Visual-Only .....                        | 85 |
| Gambar 4. 17 Flowchart Konfigurasi Model Multimodal .....                  | 88 |
| Gambar 4. 18 Hasil Evaluasi Model Multimodal Audio-Visual.....             | 88 |

## LAMPIRAN

|                                                      |     |
|------------------------------------------------------|-----|
| Lampiran 1. Kartu Asistensi Tugas Akhir .....        | 103 |
| Lampiran 2. Curriculum Vitae .....                   | 104 |
| Lampiran 3. Surat Pernyataan & Pengalihan HAKI ..... | 105 |
| Lampiran 4. Bukti bayar BNSP .....                   | 107 |
| Lampiran 5. Form Revisi Dosen Penguji.....           | 108 |

