TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2393-2403
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9869

Evaluasi Montreal Forced Aligner dan Goodness of Pronunciation untuk Penilaian Pelafalan Bahasa Sunda

Abdul Fatahillah[*], Sigit Puspito Wigati Jarot

Program Studi Teknik Informatika, Sekolah Tinggi Teknologi Terpadu Nurul Fikri, Depok, Indonesia
Email: [1,*]abdu22215ti@student.nurulfikri.ac.id, [2]sigit.jarot@nurulfikri.ac.id
Email Penulis Korespondensi: abdu22215ti@student.nurulfikri.ac.id

Abstrak–Bahasa Sunda merupakan bahasa daerah dengan penutur terbanyak kedua di Indonesia, namun ketersediaan sistem penilaian pelafalan otomatis untuk bahasa ini masih sangat terbatas. Penelitian ini melakukan evaluasi sistematis terhadap pipeline *Montreal Forced Aligner* (MFA) dan *Goodness of Pronunciation* (GOP) untuk penilaian pelafalan Bahasa Sunda dalam konteks purwarupa aplikasi pembelajaran berbasis suara. Dataset terdiri dari 2.500 sampel ujaran valid yang diperoleh dari 50 penutur asli Bahasa Sunda, mencakup 10 kosakata *basa loma* dengan 20 fonem unik. Evaluasi MFA menunjukkan kegagalan *alignment* yang bersifat total dan sistemis: seluruh 2.500 file (100%) teridentifikasi bermasalah, dengan 17 dari 20 fonem memiliki durasi tepat 10 milidetik pada seluruh segmennya. Tiga konfigurasi parameter yang berbeda menghasilkan tingkat kegagalan identik (100%), mengkonfirmasi bahwa kegagalan bukan akibat parameter melainkan keterbatasan intrinsik MFA pada data single-word berdurasi sangat pendek (rata-rata 0,69 detik) untuk bahasa *low-resource*. Evaluasi GOP menunjukkan *top-1 accuracy* global hanya 26,1%, dengan dominasi anomali fonem /l/ sebagai *top-1* untuk 14 dari 20 fonem. Pengujian fungsional membuktikan sistem tidak mampu membedakan ucapan benar dari ucapan salah. Dari sisi teknis, purwarupa aplikasi berbasis React Native dan FastAPI berhasil diimplementasikan dengan 6 dari 8 skenario pengujian *black-box* berhasil dijalankan. Penelitian ini memberikan tiga kontribusi utama: (1) kontribusi empiris berupa bukti kuantitatif pertama bahwa *pipeline* MFA-GOP standar tidak dapat diterapkan secara langsung untuk Bahasa Sunda sebagai bahasa *low-resource* dengan audio *single-word* berdurasi pendek; (2) kontribusi metodologis berupa *baseline* empiris dan kerangka evaluasi yang dapat direplikasi untuk bahasa daerah lain di Indonesia; serta (3) kontribusi praktis berupa purwarupa aplikasi *client-server* React Native–FastAPI yang menjadi titik tolak pengembangan sistem penilaian pelafalan Bahasa Sunda berbasis pendekatan alternatif.

Kata Kunci: Bahasa Sunda; *Forced Alignment*; *Goodness of Pronunciation*; *Low-Resource Language*; Penilaian Pelafalan Otomatis; *Singlew-Word Audio*

Abstract–Sundanese is the second most widely spoken regional language in Indonesia, yet automated pronunciation assessment systems for this language remain extremely scarce. This study presents a systematic evaluation of the *Montreal Forced Aligner* (MFA) and *Goodness of Pronunciation* (GOP) pipeline for Sundanese pronunciation assessment within a prototype voice-based learning application. The dataset comprises 2,500 valid utterance samples collected from 50 native Sundanese speakers, covering 10 *basa loma* vocabulary items spanning 20 unique phonemes. MFA evaluation revealed total and systemic alignment failure: all 2,500 files (100%) were identified as problematic, with 17 of 20 phonemes consistently assigned exactly 10-millisecond durations. Three distinct parameter configurations produced identical failure rates (100%), confirming that the failures are intrinsic to MFA's limitations with very short-duration single-word audio (mean 0.69 seconds) for low-resource languages. GOP evaluation yielded a global top-1 accuracy of only 26.1%, characterized by anomalous dominance of the /l/ phoneme as top-1 for 14 of 20 phonemes. Functional testing demonstrated the system's inability to discriminate correct from incorrect utterances. On the technical side, the React Native and FastAPI prototype application was successfully implemented, with 6 of 8 black-box test scenarios passing. This research provides three principal contributions: (1) empirical contribution in the form of the first quantitative evidence that the standard MFA-GOP pipeline cannot be directly applied to Sundanese as a low-resource language with short-duration single-word audio; (2) methodological contribution in the form of an empirical baseline and replicable evaluation framework applicable to other regional languages of Indonesia; and (3) practical contribution in the form of a React Native–FastAPI client-server prototype that serves as a starting point for further development of Sundanese pronunciation assessment systems using alternative approaches.

Keywords: Sundanese Language; Forced Alignment; Goodness of Pronunciation; Low-Resource Language; Automatic Pronunciation Assessment; Single-Word Audio

1. PENDAHULUAN

Bahasa daerah merupakan elemen penting identitas budaya bangsa yang perlu dijaga dan dilestarikan (Kamaly et al., 2025). Indonesia dikenal sebagai negara dengan keragaman bahasa luar biasa, dengan lebih dari 700 bahasa daerah yang tercatat (Ashar & Handayani, 2025), menempatkannya sebagai negara dengan keragaman bahasa terbanyak kedua setelah Papua Nugini (Rusyana & Rohmah, 2024). Namun, seiring globalisasi, bahasa daerah menghadapi ancaman serius akibat menurunnya pewarisan antar generasi, UNESCO mencatat setidaknya ada 143 bahasa daerah yang mulai terancam punah (Mandolang et al., 2024). Kaharuddin et al. (2024) menunjukkan bahwa penetrasi bahasa dominan di era komunikasi digital menjadi faktor utama yang mempercepat ancaman kepunahan bahasa daerah di Indonesia. Selain itu, Bahasa Sunda menduduki peringkat kedua sebagai bahasa daerah paling banyak dituturkan di Indonesia, dengan jumlah penutur mencapai 34 juta jiwa berdasarkan sensus penduduk 2020 (Rusyana & Rohmah, 2024), namun vitalitasnya dinilai mengalami penurunan penutur aktif di kalangan generasi muda (Sulastri Ai et al., 2023). Bahasa Sunda memiliki sistem fonologi yang khas, seperti vokal /*é*/, /*eu*/, dan konsonan retrofleks, yang menantang bagi pelajar *non-native* (Rusyana & Rohmah, 2024). Yanti Rut Susanti (2022) juga mencatat bahwa generasi muda di wilayah Jawa Barat semakin jarang menggunakan Bahasa Sunda dalam komunikasi sehari-hari. Untuk menjaga keberlangsungannya, diperlukan upaya yang adaptif dengan perkembangan teknologi (Nurjanah et al., 2025).

TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2393-2403
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9869

Pengembangan sistem penilaian pelafalan otomatis atau *Automatic Pronunciation Verification* (APV) masih didominasi untuk bahasa-bahasa mayor dengan ketersediaan data besar (Kheir et al., 2023; Liu et al., 2025). Beberapa penelitian terkait telah dilakukan: Rahmah dan Juhriah (2021) mengembangkan aplikasi pembelajaran Bahasa Sunda berbasis Android, namun tanpa komponen penilaian pelafalan otomatis; Tri Pujiani dan Miftahuddin (2022) membangun sistem ASR Sunda menggunakan PCA dan VQ yang baru mencapai deteksi kemiripan kata, belum mengintegrasikan *forced alignment* atau GOP; Pratama dan Amrullah (2024) mengevaluasi Whisper untuk bahasa rendah sumber daya dan menemukan keterbatasan signifikan; Arisaputra et al. (2024) mengevaluasi model XLS-R pada ASR Bahasa Sunda dan menemukan bahwa model *multilingual* masih menghadapi tantangan besar; serta Cryssiover dan Zahra (2024) membangun ASR Sunda menggunakan Wav2Vec 2.0 yang mencapai WER 23,5% pada dataset 53 jam. Sementara untuk APV khusus Sunda, eksplorasi yang efektif masih sangat jarang (Getman et al., 2023), menciptakan *research gap* yang signifikan.

Tinjauan terhadap sebelas penelitian kunci mengidentifikasi tiga celah evaluatif yang spesifik: (1) belum ada penelitian yang mengukur akurasi *phone alignment* MFA secara kuantitatif pada data ujaran Bahasa Sunda; (2) belum ada studi yang menguji dan melaporkan korelasi antara skor GOP dengan penilaian subjektif dari penutur asli Sunda; dan (3) belum ada eksplorasi terhadap kelayakan implementasi *pipeline* MFA-GOP dalam purwarupa aplikasi pembelajaran. Status Sunda sebagai bahasa *low-resource* membuat pendekatan *end-to-end* berbasis *self-supervised learning* yang membutuhkan data sangat besar kurang layak (Kim et al., 2022), sehingga evaluasi terhadap pendekatan *pipeline* yang lebih efisien seperti MFA-GOP menjadi pilihan yang rasional.

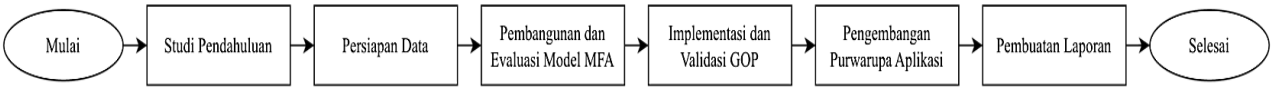
Untuk menjembatani kesenjangan tersebut, penelitian ini bertujuan mengevaluasi penerapan *Montreal Forced Aligner* (MFA) dan *Goodness of Pronunciation* (GOP) dalam mendeteksi dan menilai kualitas pelafalan Bahasa Sunda, mengukur akurasi sistem dalam memberikan umpan balik pelafalan, serta menilai kelayakan model akustik tersebut untuk diimplementasikan pada aplikasi pembelajaran berbasis suara. Evaluasi ini diwujudkan melalui sebuah purwarupa aplikasi yang memanfaatkan MFA untuk penyelarasan fonetik antara suara pengguna dengan data referensi penutur asli (Mcauliffe et al., 2017), sedangkan GO P berfungsi memberikan skor penilaian kualitas pengucapan secara objektif (Witt & Young, 2000). Melalui pendekatan ini, pengguna dapat memperoleh umpan balik langsung berupa nilai dan visualisasi kesesuaian suara (Zhang et al., 2021), sehingga proses belajar menjadi lebih interaktif dan terukur.

Berdasarkan latar belakang dan tujuan tersebut, penelitian ini memberikan tiga kontribusi utama. Pertama, kontribusi empiris, yaitu menyajikan bukti kuantitatif pertama mengenai kinerja *pipeline* MFA-GOP standar pada Bahasa Sunda sebagai bahasa *low-resource* dengan karakteristik audio *single-word* berdurasi pendek, melalui audit *alignment* menyeluruh terhadap 2.500 sampel dan evaluasi diskriminatif terhadap 20 model fonem. Kedua, kontribusi metodologis, yaitu menetapkan *baseline* empiris dan kerangka evaluasi yang dapat direplikasi untuk mengukur kelayakan *forced alignment* dan *goodness of pronunciation* pada bahasa daerah lain di Indonesia. Ketiga, kontribusi praktis, yaitu menghasilkan purwarupa aplikasi berbasis React Native dan FastAPI sebagai bukti konsep arsitektur *client-server* untuk sistem penilaian pelafalan Bahasa Sunda yang dapat menjadi titik tolak pengembangan lanjutan menggunakan pendekatan alternatif seperti model *pretrained* lintas bahasa atau *self-supervised learning*. Ketiga kontribusi tersebut secara kolektif diharapkan menjadi rujukan empiris bagi upaya pelestarian Bahasa Sunda melalui teknologi pembelajaran berbasis suara yang interaktif, akurat, dan adaptif.

2. METODOLOGI PENELITIAN

2.1 Desain dan Kerangka Penelitian

Penelitian ini menggunakan rancangan studi evaluatif dengan pendekatan kuantitatif untuk mengukur akurasi teknis dan validitas fungsional *pipeline* MFA-GOP. Seluruh tahapan penelitian mencakup enam fase. (1) Studi Pendahuluan, (2) Persiapan Data, (3) Pembangunan dan Evaluasi Model MFA, (4) Implementasi dan Validasi GOP, (5) Pengembangan Purwarupa Aplikasi, dan (6) Pembuatan Laporan sebagaimana direpresentasikan dalam Gambar 1 berikut. Kajian sistematis menunjukkan bahwa tahapan penelitian yang terstruktur merupakan kunci dalam membangun sistem penilaian pelafalan yang andal (Liu et al., 2025).



Gambar 1. Tahapan Penelitian Evaluasi MFA dan GOP untuk Penilaian Pelafalan Bahasa Sunda

2.2 Studi Pendahuluan

Tahap pertama adalah Studi Pendahuluan yang meliputi identifikasi masalah dan *research gap* melalui tinjauan literatur mendalam. Kajian pustaka dilakukan terhadap penelitian terkait *Automatic Pronunciation Verification* (APV), MFA, GOP, dan teknologi pembelajaran bahasa daerah. Penelusuran literatur dilakukan pada basis data Scopus, Google Scholar, dan IEEE Xplore dengan kata kunci antara lain *forced alignment*, *goodness of pronunciation*, *low-resource language*, dan *Sundanese speech*. Tinjauan terhadap sebelas penelitian kunci menghasilkan identifikasi tiga celah evaluatif: (1) belum ada pengukuran akurasi *phone alignment* MFA pada data Bahasa Sunda; (2) belum ada pengujian korelasi skor GOP dengan penilaian penutur asli Sunda; dan (3) belum ada eksplorasi kelayakan implementasi *pipeline*

TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2393-2403
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9869

MFA-GOP dalam purwarupa aplikasi pembelajaran berbasis suara. Keluaran tahap ini adalah perumusan masalah, tujuan penelitian, dan desain metodologi evaluasi (Kheir et al., 2023).

2.3 Persiapan Data

Tahap kedua adalah Persiapan Data yang mencakup pemilihan kosakata target, perekrutan partisipan, perekaman suara, dan pra-pemrosesan audio. Kosakata target terdiri dari sepuluh kata dasar Bahasa Sunda tingkat *basa loma* yang dipilih berdasarkan keterwakilan fonem khas Bahasa Sunda, antara lain vokal /*é*/, /*eu*/, serta konsonan retrofleks dan nasal. Sepuluh kosakata tersebut ialah: (1) *abdi*, (2) *beus*, (3) *nginum*, (4) *nyieun*, (5) *leutik*, (6) *gedé*, (7) *leumpang*, (8) *cai*, (9) *imah*, dan (10) *béas*. Kesepuluh kata ini bersama-sama mencakup 20 fonem unik Bahasa Sunda yang menjadi unit analisis dalam evaluasi.

Partisipan yang merupakan penutur asli Bahasa Sunda direkrut berdasarkan kriteria inklusi dan eksklusi yang dirangkum pada Tabel 1. Setiap partisipan diminta mengucapkan setiap kata sebanyak lima kali pengulangan ke dalam mikrofon *smartphone*, sehingga total target sampel adalah 2.500 (50 penutur × 10 kata × 5 repetisi). Penelitian sebelumnya menunjukkan bahwa pengumpulan data terstruktur dari penutur asli merupakan langkah kritis dalam membangun sistem penilaian pelafalan yang andal (Getman et al., 2023).

Tabel 1. Kriteria Responden Penelitian

No.	Kategori	Kriteria	Penjelasan
1.	Inklusi	a. Penutur asli Bahasa Sunda sejak lahir, berdomisili di Jawa Barat. b. Berusia antara 20 hingga 50 tahun. c. Tidak memiliki riwayat gangguan bicara atau pendengaran. d. Bersedia berpartisipasi.	a. Memastikan kompetensi linguistik dan aksen yang otentik sebagai acuan. b. Meminimalisasi variasi pelafalan akibat usia ekstrem. c. Agar sampel suara merepresentasikan produksi ujaran yang normal. d. Memenuhi aspek etika penelitian.
2.	Eksklusi	a. Pengisian data diri atau partisipasi tidak lengkap.	a. Menjaga konsistensi dan kelengkapan data administrasi penelitian.

Tabel 1 merincikan syarat inklusi dan eksklusi bagi partisipan. Seluruh rekaman melalui pra-pemrosesan: (1) konversi format ke WAV mono 16 kHz 16-bit sesuai spesifikasi MFA; (2) normalisasi level suara menggunakan *peak normalization* ke -3 dBFS; dan (3) *silence trimming* otomatis untuk memangkas jeda hening di awal dan akhir rekaman menggunakan *library pydub* dengan ambang batas -40 dBFS. Untuk keperluan validasi, sebanyak 20% dari total rekaman disisihkan sebagai subset data uji final secara acak berstrata per kata.

2.4 Pembangunan dan Evaluasi Model MFA

Tahap ketiga adalah Pembangunan dan Evaluasi Model MFA. MFA merupakan *toolkit* standar untuk *forced alignment* yang memungkinkan pelatihan model akustik baru atau adaptasi model dari bahasa lain ke bahasa target (Mcauliffe et al., 2017). Dalam penelitian ini, model akustik dilatih dari nol (*from-scratch*) menggunakan seluruh data pelatihan (80% dari dataset). Proses pelatihan menggunakan perintah *mfa train* dengan algoritma *Gaussian Mixture Model — Hidden Markov Model* (GMM-HMM) yang merupakan arsitektur standar MFA untuk tugas *forced alignment* (Mcauliffe et al., 2017).

Setelah model terbentuk, proses *forced alignment* dijalankan pada subset data uji. Keluaran *alignment* berupa file *TextGrid* (format Praat) yang berisi batas waktu (*onset* dan *offset*) setiap fonem untuk semua sampel uji. Evaluasi kualitas *alignment* dilakukan secara kuantitatif menggunakan empat metrik: (1) *aligned ratio*, yaitu rasio total durasi fonem terhadap durasi audio keseluruhan; (2) durasi minimum fonem; (3) jumlah fonem dengan durasi di bawah 30 milidetik (ambang batas minimum fonem yang bermakna secara akustik); dan (4) posisi awal fonem pertama relatif terhadap total durasi audio (*late start ratio*). Untuk memastikan objektivitas evaluasi, selain konfigurasi default, dilakukan pula percobaan dengan dua konfigurasi parameter tambahan: pengurangan bobot *silence* (*--silence_weight 0.01*) dan perluasan ruang pencarian *alignment* (*--beam 1000 --retry_beam 10000*).

2.5 Implementasi dan Validasi GOP

Tahap keempat adalah Implementasi dan Validasi GOP. *Goodness of Pronunciation* (GOP) adalah metrik klasik yang mengukur kualitas pelafalan suatu fonem berdasarkan perbandingan kemiripan akustiknya dengan model referensi (Witt & Young, 2000). Formulasi asli GOP yang diusulkan oleh (Witt & Young, 2000) mendefinisikan skor kualitas pelafalan pada level fonem sebagai nilai absolut dari *log* probabilitas posterior yang ternormalisasi terhadap jumlah frame, sebagaimana dinyatakan dalam Persamaan (1).

$$\mathrm{GOP}(p) = \left|\log\left(\frac{p(O^{(p)}|p)}{\max_{q\in Q} p(O^{(p)}|q)}\right)\right| / \mathrm{NF}(p) \quad (1)$$

Dalam Persamaan (1), *p* merupakan fonem target, *O(p)* merupakan segmen akustik yang berkorespondensi, *q* merupakan himpunan seluruh model fonem yang tersedia, dan *NF(p)* merupakan jumlah *frame* dalam segmen tersebut.

TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2393-2403
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9869

Skor GOP yang rendah (mendekati nol) menunjukkan pelafalan yang baik karena fonem target memiliki probabilitas posterior yang tinggi, sedangkan skor yang besar menunjukkan penyimpangan dari pelafalan *native* (Witt & Young, 2000).

Dalam implementasi penelitian ini, formulasi GOP diadaptasi untuk konteks model GMM independen per fonem yang dilatih menggunakan *library scikit-learn*, bukan *framework* Kaldi HMM secara utuh. Alih-alih menggunakan rasio *log-posterior* sebagaimana formulasi asli, digunakan normalisasi *min-max* terhadap *log-likelihood* seluruh model fonem sebagaimana dinyatakan dalam Persamaan (2).

$$\mathrm{GOP}_{\mathrm{norm}}(p) = \frac{LL_{target} - LL_{min}}{LL_{max} - LL_{min}}, \in [0,1] \quad (2)$$

Dalam Persamaan (2), *LL_target* merupakan *log-likelihood* segmen audio terhadap model GMM fonem target *p*; *LL_max* merupakan *log-likelihood* tertinggi dari seluruh K model fonem (K = 20 dalam penelitian ini); dan *LL_min* merupakan *log-likelihood* terendah. Hasil perhitungan di-*clip* ke rentang [0, 1], di mana nilai 1 menunjukkan bahwa fonem target merupakan fonem dengan *likelihood* tertinggi dan nilai mendekati 0 menunjukkan ketidaksesuaian yang besar. Adaptasi ini dipilih karena model GMM per fonem bersifat independen sehingga tidak tersedia probabilitas transisi HMM yang dibutuhkan untuk rasio *log-posterior* pada formulasi asli; namun prinsip fundamentalnya tetap sama, yaitu mengukur seberapa cocok sinyal akustik suatu segmen terhadap model fonem target relatif terhadap model fonem lainnya.

Model GMM dilatih dengan 8 komponen *Gaussian* per fonem. Nilai 8 komponen dipilih berdasarkan pertimbangan keseimbangan antara kapasitas model dan keterbatasan jumlah data pelatihan per fonem. Fitur akustik yang digunakan adalah *Mel-Frequency Cepstral Coefficients* (MFCC) dengan 13 koefisien, yang merupakan representasi standar dalam sistem pengenalan suara dan penilaian pelafalan (Mcauliffe et al., 2017). Kemampuan diskriminatif model dievaluasi menggunakan metrik *top-1 accuracy*, yaitu persentase segmen di mana fonem target merupakan fonem dengan *likelihood* tertinggi dari seluruh 20 model fonem yang tersedia.

Selain itu, dilakukan analisis distribusi statistik skor GOP dari rekaman penutur asli yang dianggap sebagai pelafalan referensi ("benar"). Statistik yang dihitung mencakup rata-rata, simpangan baku, nilai minimum, dan persentil ke-10 (*p10*) skor GOP per fonem. Profil statistik ini menjadi *baseline* untuk menilai deviasi pelafalan pengguna *non-native* pada tahap berikutnya. Pengujian fungsional dilakukan dengan tiga skenario kontrol pada kata target: mengucapkan kata dengan benar, mengucapkan suara acak (*random noise*), dan mengucapkan kata yang berbeda dari target. Ketiga skenario ini dirancang untuk mengukur kemampuan diskriminatif sistem secara langsung.

2.6 Pengembangan Purwarupa Aplikasi

Tahap kelima adalah Pengembangan Purwarupa Aplikasi yang bertujuan mengimplementasikan *pipeline* MFA-GOP yang telah divalidasi dalam sebuah purwarupa (*proof-of-concept*) aplikasi pembelajaran pelafalan Bahasa Sunda berbasis Mobile. Purwarupa dibangun menggunakan arsitektur *client-server* dua lapisan. Aplikasi *client* dikembangkan menggunakan *React Native* dengan *framework* Expo, yang memungkinkan pengembangan lintas platform sekaligus memanfaatkan API perekaman suara perangkat secara langsung. Server *backend* dibangun menggunakan Python dengan *framework* FastAPI dan diakses melalui *tunnel* ngrok selama fase pengembangan. Arsitektur ini memungkinkan pemrosesan model MFA-GOP yang berat dilakukan di sisi server, sehingga tidak membebani perangkat *client*. Sistem CAPT modern sering mengadopsi pola arsitektur serupa untuk skalabilitas dan kemudahan pembaruan model (Kheir et al., 2023).

Alur kerja purwarupa secara fungsional terdiri dari empat tahap: (1) pengguna memilih kata target dan merekam ucapannya melalui antarmuka aplikasi; (2) file audio dikirim ke server *backend* melalui protokol HTTP; (3) server menjalankan proses MFA *alignment* untuk mendapatkan segmentasi fonem, kemudian menghitung skor GOP menggunakan Persamaan (2) untuk setiap fonem; dan (4) hasil berupa skor GOP per fonem beserta *verdict* keseluruhan dikembalikan ke aplikasi *client* untuk ditampilkan sebagai umpan balik kepada pengguna. Pengujian purwarupa dilakukan melalui *black-box testing* terhadap delapan skenario yang mencakup aspek teknis (perekaman, pengiriman, pemrosesan, penanganan *error*) maupun aspek akurasi penilaian (ucapan benar dan ucapan salah).

2.7 Analisis Data

Analisis data dalam penelitian ini dilakukan secara kuantitatif deskriptif. Untuk evaluasi MFA, analisis difokuskan pada empat metrik *alignment* yang telah disebutkan pada subbab 2.4, disajikan dalam bentuk tabel ringkasan per kata dan analisis pola durasi fonem secara keseluruhan. Untuk evaluasi GOP, analisis berupa statistik deskriptif (rata-rata, simpangan baku, distribusi, dan *top-1 accuracy*) dari skor yang dihasilkan oleh penutur asli, disertai analisis distribusi *Log-Likelihood Ratio* (LLR) per fonem. Hasil pengujian fungsional disajikan dalam tabel *black-box* yang memuat skenario, hasil yang diharapkan, status, dan keterangan.

Seluruh komputasi analisis dilakukan menggunakan bahasa pemrograman Python dengan *library* pendukung: *NumPy* dan *SciPy* untuk perhitungan statistik, *Pandas* untuk manajemen data tabular, *scikit-learn* untuk pelatihan model GMM, *librosa* untuk ekstraksi fitur MFCC, dan *matplotlib*/*seaborn* untuk visualisasi distribusi data. Analisis *alignment* TextGrid dilakukan menggunakan *library praatio* yang memungkinkan pembacaan dan inspeksi otomatis terhadap seluruh 2.500 file *TextGrid* hasil *alignment* MFA.

TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2393-2403
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9869

3. HASIL DAN PEMBAHASAN

3.1 Hasil Persiapan Data

Proses pengumpulan data berhasil memperoleh rekaman dari 50 penutur asli Bahasa Sunda yang memenuhi seluruh kriteria inklusi sebagaimana tercantum pada Tabel 1. Setiap penutur mengucapkan sepuluh kata target masing-masing sebanyak lima kali pengulangan, sehingga total diperoleh 2.500 sampel ujaran. Setelah proses verifikasi dan pra-pemrosesan, sebanyak 2.500 sampel dinyatakan valid untuk digunakan dalam tahap selanjutnya

Hasil analisis properti audio menunjukkan bahwa seluruh file memiliki format yang seragam, yaitu WAV mono 16 kHz 16-bit, sesuai dengan spesifikasi yang dibutuhkan oleh MFA. Rata-rata durasi audio per sampel adalah 0,69 detik dengan rentang 0,38 hingga 1,51 detik. Tabel 2 menyajikan ringkasan *dataset* yang berhasil dikumpulkan.

Tabel 2. Ringkasan Dataset Penelitian

Kata	Fonem	Jumlah Sampel	Durasi Rata-rata (detik)
abdi	/a/ /b/ /d/ /i/	250	0,74
béas	/b/ /ɛ/ /a/ /s/	250	0,67
beus	/b/ /ʁ/ /s/	250	0,89
cai	/tʃ/ /a/ /i/	250	0,62
gedé	/g/ /ə/ /d/ /ɛ/	250	0,57
imah	/i/ /m/ /a/ /h/	250	0,74
leumpang	/l/ /ʁ/ /m/ /p/ /a/ /ŋ/	250	1,18
leutik	/l/ /ʁ/ /t/ /i/ /k/	250	0,75
nginum	/ŋ/ /i/ /n/ /u/ /m/	250	0,84
nyieun	/ɲ/ /i/ /ʁ/ /n/	250	0,87
Total	20 fonem unik	2.500	0,69 (rata-rata)

Berdasarkan Tabel 2, teridentifikasi sebanyak 20 fonem unik Bahasa Sunda yang mencakup 7 vokal dan 13 konsonan. Rata-rata durasi audio yang relatif pendek (0,69 detik) menjadi karakteristik penting dataset ini yang berpengaruh signifikan terhadap kinerja *forced alignment* sebagaimana akan dibahas pada subbab berikutnya.

3.2 Evaluasi Model Montreal Forced Aligner

Pelatihan model akustik MFA dilaksanakan menggunakan perintah mfa train dengan seluruh 2.500 sampel sebagai data pelatihan. Model berhasil dibangun dan tersimpan dalam file sunda_model.zip dengan ukuran 4.396 KB. Proses *forced alignment* kemudian dijalankan untuk menghasilkan file *TextGrid* bagi setiap sampel. Untuk mengevaluasi kualitas *alignment* secara komprehensif, dilakukan audit otomatis terhadap seluruh 2.500 file *TextGrid* yang dihasilkan.

3.2.1 Hasil Audit Alignment

Evaluasi kualitas alignment dilakukan dengan mengukur empat metrik utama: *aligned ratio* (rasio total durasi fonem terhadap durasi audio), durasi minimum fonem, jumlah fonem yang terlalu pendek (di bawah 30 milidetik), dan posisi awal fonem pertama relatif terhadap durasi audio. Hasil audit menunjukkan temuan yang sangat signifikan: seluruh 2.500 file (100%) teridentifikasi memiliki masalah *alignment*. Tabel 3 menyajikan ringkasan hasil audit per kata.

Tabel 3. Hasil Audit Alignment MFA per Kata

Kata	Total	Bermasalah	Aligned Ratio	Dur. Min.	Fonem <30ms
abdi	250	250 (100%)	61,8%	10 ms	92–100%
béas	250	250 (100%)	10,7%	10 ms	100%
beus	250	250 (100%)	14,9%	10 ms	100%
cai	250	250 (100%)	6,8%	10 ms	100%
gedé	250	250 (100%)	10,9%	10 ms	100%
imah	250	250 (100%)	7,5%	10 ms	100%
leumpang	250	250 (100%)	9,7%	10 ms	100%
leutik	250	250 (100%)	31,6%	5 ms	95–100%
nginum	250	250 (100%)	7,9%	10 ms	100%
nyieun	250	250 (100%)	8,0%	10 ms	100%

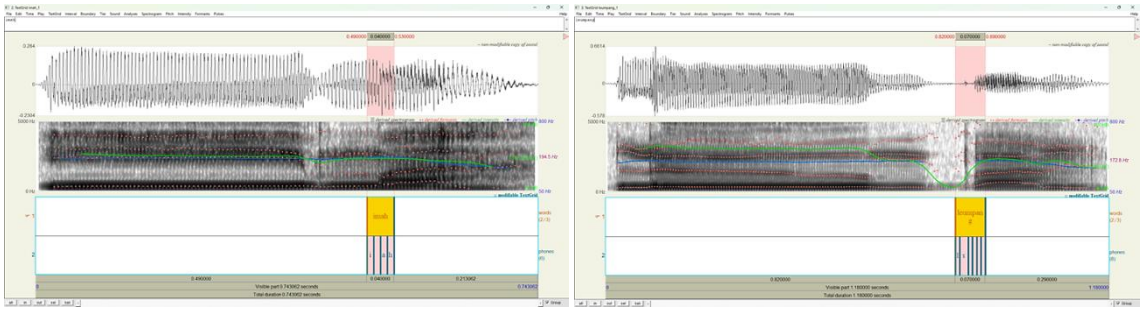
Tabel 3 menunjukkan aligned ratio yang sangat rendah, berkisar antara 6,8% hingga 61,8%. Hal ini mengindikasikan bahwa MFA hanya mampu mencocokkan sebagian kecil dari sinyal audio ke fonem target, sedangkan sebagian besar audio diidentifikasi sebagai interval kosong. Untuk memastikan bahwa kegagalan *alignment* bukan disebabkan oleh konfigurasi parameter yang kurang optimal, dilakukan tiga percobaan dengan konfigurasi berbeda. Konfigurasi pertama (v1) menggunakan parameter --silence_weight 0.01 yang bertujuan untuk mengurangi penyisipan silence. Konfigurasi kedua (v2) menggunakan parameter default tanpa modifikasi apapun. Konfigurasi ketiga (v3) menggunakan parameter --beam 1000 dan --retry_beam 10000 yang secara signifikan memperluas ruang pencarian alignment. Tabel 4 menyajikan perbandingan hasil ketiga konfigurasi.

TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2393-2403
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9869

Tabel 4. Perbandingan Tiga Konfigurasi Parameter MFA

Metrik	v1 (silence_weight)	v2 (default)	v3 (beam besar)
File bermasalah	2.500 (100%)	2.500 (100%)	2.500 (100%)
Aligned ratio	6,5–61,8%	6,6–33,5%	6,5–32,4%
Fonem <30ms	92–100%	88–100%	88–100%
Durasi min. fonem	10 ms	5–10 ms	5–10 ms
Late start (>70%)	84 file	98 file	99 file

Tabel 4 mengkonfirmasi ketiga konfigurasi menghasilkan tingkat kegagalan identik (100%), membuktikan kegagalan bersifat sistemis. Visualisasi *TextGrid* menggunakan perangkat lunak Praat memperjelas pola kegagalan ini sebagaimana ditampilkan pada Gambar 2 berikut.



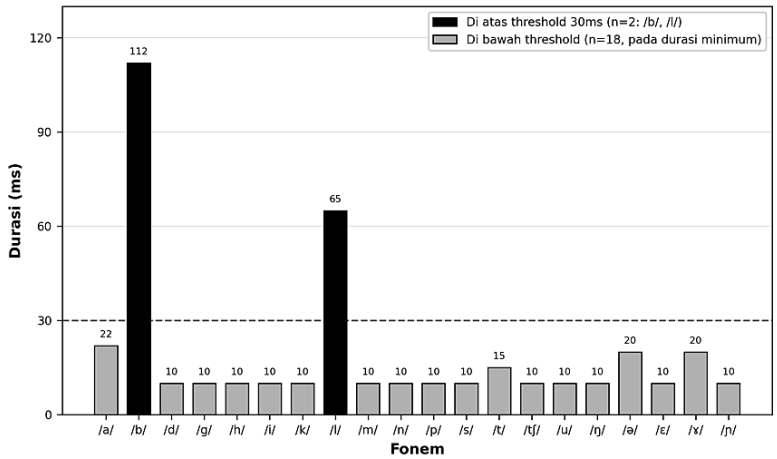
(a) Visualisasi kata "imah"; (b) Visualisasi kata "leumpang"

Gambar 2. Visualisasi TextGrid alignment MFA pada Praat

Pada Gambar 2 terlihat bahwa MFA menempatkan seluruh fonem dalam segmen sangat sempit di satu bagian kecil audio, sementara bagian yang jelas mengandung sinyal *speech* dibiarkan sebagai interval kosong. Tidak ada korelasi antara posisi batas fonem hasil MFA dengan gelombang suara yang terlihat pada spektrogram Praat. Pola ini ditemukan konsisten pada seluruh 10 kata dan 50 penutur.

3.2.2 Analisis Durasi Fonem

Analisis lebih lanjut terhadap durasi fonem dari seluruh 10.492 segmen menunjukkan pola yang sangat homogen dan tidak natural. Sebanyak 17 dari 20 fonem memiliki durasi yang persis sama (10 milidetik) pada seluruh atau hampir seluruh segmennya, dengan standar deviasi mendekati nol. Hanya fonem /b/ (rata-rata 112 milidetik) dan /l/ (rata-rata 64,7 milidetik) yang menunjukkan variasi durasi. Distribusi durasi fonem divisualisasikan pada Gambar 3 berikut.



Gambar 3. Distribusi durasi rata-rata fonem hasil alignment MFA

Gambar 3 merupakan indikasi kuat bahwa MFA menerapkan durasi minimum paksa (*minimum forced duration*) karena tidak mampu menemukan *alignment* yang bermakna. Durasi ini setara dengan satu *frame* MFCC, yang berarti setiap segmen fonem hanya mengandung satu vektor fitur akustik dan tidak cukup untuk merepresentasikan karakteristik fonetik apapun secara bermakna.

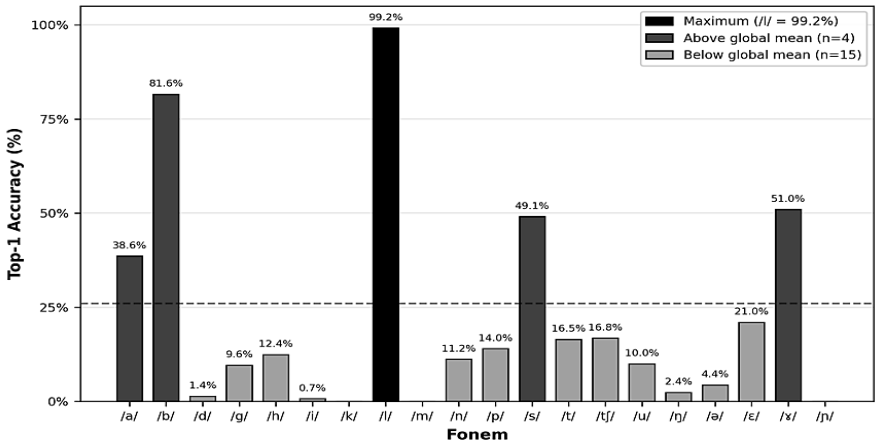
3.3 Evaluasi Goodness of Pronunciation

Model GMM untuk GOP dilatih dari segmen fonem hasil *alignment* MFA menggunakan 8 komponen *Gaussian* per fonem. Meskipun 20 model fonem berhasil dibangun secara teknis, kualitasnya sangat terpengaruh oleh buruknya *alignment*. Tabel 5 menyajikan hasil *top-1 accuracy* per fonem berikut.

TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2393-2403
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9869

Tabel 5. Top-1 Accuracy per Fonem

Fonem	N	Top-1 OK	Top-1 %	Fonem Dominan
/a/	1.250	482	38,6%	/l/ (511 kali)
/b/	749	611	81,6%	/b/ (611 kali)
/d/	500	7	1,4%	/l/ (352 kali)
/i/	1.499	10	0,7%	/l/ (1.168 kali)
/k/	249	0	0,0%	/l/ (242 kali)
/l/	499	495	99,2%	/l/ (495 kali)
/m/	750	0	0,0%	/l/ (563 kali)
Global	10.492	2.741	26,1%	/l/ dominasi 14/20 fonem

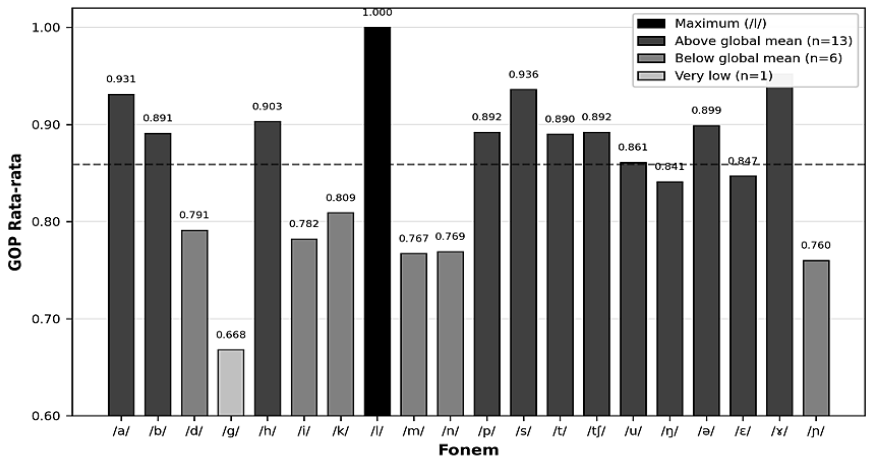


Gambar 4. Top-1 accuracy per fonem

Tabel 5 dan Gambar 4 menunjukkan *top-1 accuracy* global hanya 26,1%. Fonem /l/ mendominasi sebagai *top-1* untuk 14 dari 20 fonem, sementara /k/, /m/, dan /ɲ/ memiliki *top-1 accuracy* 0,0%. Dominasi /l/ merupakan artefak langsung dari *alignment* yang buruk: karena /l/ memiliki variasi durasi lebih besar (64,7 ms) dibanding fonem lain (10 ms), model GMM /l/ dilatih dari segmen yang lebih beragam sehingga mengembangkan distribusi yang sangat luas dalam ruang fitur MFCC, memberikan *likelihood* tinggi terhadap hampir semua *input* akustik. Tabel 6 menyajikan distribusi nilai *Log-Likelihood Ratio* (LLR) dan skor GOP per fonem dari ucapan benar penutur asli sebagai berikut.

Tabel 6. Distribusi LLR dan GOP per Fonem (Ucapan Benar Penutur Asli)

Fonem	N	LLR Mean	LLR Std	GOP Mean	GOP Std	GOP Min	GOP p10
/a/	1.250	−2,09	2,74	0,931	0,094	0,560	0,786
/i/	1.499	−5,61	3,30	0,782	0,127	0,394	0,614
/l/	499	−0,00	0,04	1,000	0,003	0,958	1,000
/m/	750	−5,40	2,89	0,767	0,107	0,331	0,613
/k/	249	−7,04	2,31	0,809	0,098	0,463	0,678
/ɲ/	250	−8,39	2,97	0,760	0,109	0,325	0,628
Global	10.492	−3,63	3,85	0,859	0,154	0,000	0,652



Gambar 5. Distribusi GOP rata-rata per fonem dari ucapan benar penutur asli

TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2393-2403
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9869

Tabel 6 dan Gambar 5 menunjukkan LLR rata-rata global negatif (−3,63), mengkonfirmasi fonem target bukan fonem dengan *likelihood* tertinggi pada mayoritas segmen. Fonem /l/ anomali ekstrem: LLR ≈ 0 dan GOP = 1,000 — model /l/ bersifat *overgeneralized* akibat *alignment* yang buruk. Pengujian fungsional tiga skenario pada kata "*imah*" (Tabel 7) membuktikan ketidakberfungsian sistem berikut.

Tabel 7. Hasil Pengujian Fungsional Sistem MFA-GOP pada Kata "imah"

Skenario	/i/	/m/	/a/	/h/	Overall GOP → Verdict
"imah" (benar)	0,773	0,896	0,912	0,237	0,705 → KURANG
"rrrrr" (salah)	0,731	0,919	0,851	0,960	0,865 → BENAR
"kukuku" (salah)	0,759	0,781	1,000	0,699	0,810 → BENAR

Tabel 7 membuktikan secara definitif bahwa sistem tidak mampu membedakan ucapan benar dari ucapan salah: ucapan "rrrrr" dan "kukuku" mendapat skor GOP lebih tinggi (0,865 dan 0,810) dari ucapan benar (0,705). Fenomena ini terjadi karena normalisasi *min-max* memetakan *log-likelihood* ke [0,1] terlepas dari apakah audio mengandung fonem target, sehingga model yang tidak diskriminatif tetap menghasilkan skor tinggi untuk input apapun.

3.4 Pengembangan Purwarupa Aplikasi

Purwarupa aplikasi berhasil dibangun dengan arsitektur *client-server*: aplikasi *client* berbasis React Native/Expo (diuji mengunakan *device* iPhone) dan server *backend* berbasis Python FastAPI yang diakses via ngrok.

Gambar 6. Antarmuka utama purwarupa

Gambar 6 menunjukkan antarmuka utama purwarupa aplikasi. Pengguna disajikan daftar sepuluh kata target Bahasa Sunda yang dapat dipilih untuk latihan pelafalan. Setiap kata ditampilkan bersama transkripsi fonetiknya untuk memberikan panduan visual kepada pengguna. Setelah memilih kata, pengguna dapat menekan tombol rekam untuk mengucapkan kata tersebut. Audio rekaman kemudian dikirim ke server *backend* untuk diproses melalui *pipeline* MFA *alignment* dan GOP *scoring*.

Gambar 7. Tampilan hasil penilaian

TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2393-2403
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9869

Gambar 7 menunjukkan tampilan hasil penilaian pelafalan setelah pengguna merekam ucapannya. Sistem menampilkan skor GOP untuk setiap fonem beserta status masing-masing (BENAR atau SALAH), serta verdict keseluruhan di bagian atas layar. Umpan balik visual ini dirancang agar pengguna dapat mengidentifikasi fonem mana yang perlu diperbaiki pelafalannya. Waktu pemrosesan rata-rata dari saat pengguna selesai merekam hingga hasil ditampilkan adalah sekitar 5 hingga 6 detik, yang didominasi oleh proses MFA *alignment* di sisi server.

Tabel 8. Pengujian Black-Box Purwarupa Aplikasi

No.	Skenario Pengujian	Hasil Diharapkan	Status	Keterangan
1	Memilih kata target	Kata terpilih ditampilkan	Berhasil	–
2	Merekam audio ucapan	Audio terekam dan terkirim	Berhasil	–
3	Server memproses hasil	Skor GOP ditampilkan	Berhasil	~5–6 detik
4	Audio terlalu pendek (<0,3 dtk)	Pesan error ditampilkan	Berhasil	–
5	Audio tanpa suara	Pesan error ditampilkan	Berhasil	–
6	Skor per fonem ditampilkan	Detail fonem + status	Berhasil	–
7	Ucapan benar → skor tinggi	Verdict BENAR	Gagal	Ucapan benar: KURANG
8	Ucapan salah → skor rendah	Verdict SALAH	Gagal	Ucapan salah: BENAR

Pengujian *black-box* terhadap delapan skenario menunjukkan 6 dari 8 skenario berhasil, sebagaimana yang ditampilkan pada Tabel 8. Skenario teknis (pemilihan kata, perekaman, pengiriman, penanganan *error*, dan penampilan skor) seluruhnya berjalan tanpa *error*. Waktu pemrosesan rata-rata 5–6 detik per *request*, didominasi proses MFA *alignment* di sisi server. Skenario 7 dan 8 (akurasi penilaian) gagal akibat model MFA-GOP yang tidak diskriminatif, bukan akibat arsitektur aplikasi.

3.5 Pembahasan

3.5.1 Analisis Kegagalan Forced Alignment MFA

Kegagalan total MFA bersifat sistemis dan konsisten pada seluruh 2.500 file dan tiga konfigurasi parameter. Temuan ini konsisten dengan keterbatasan yang didokumentasikan (Mcauliffe et al., 2017), yang secara eksplisit menyatakan bahwa MFA tidak dirancang untuk menyelaraskan file tunggal dan membutuhkan kumpulan ujaran dan penutur yang besar. Dokumentasi resmi MFA merekomendasikan minimum 3–5 jam data audio, sementara total kumulatif dataset penelitian ini hanya sekitar 29 menit jauh di bawah ambang minimum.

Perbandingan dengan Tosolini & Bowern (2025) memberikan konteks yang lebih jelas. Dalam studi mereka pada enam bahasa Aborigin Australia dengan data 16–290 menit, model yang dilatih dari nol dengan data terkecil secara konsisten menjadi model berperforma terburuk. Mereka menemukan bahwa model yang diadaptasi dari bahasa Inggris menghasilkan *boundary error* sekitar separuh model *scratch-trained*, dan merekomendasikan penggunaan model *pretrained* sebagai pendekatan paling aksesibel untuk bahasa rendah sumber daya. Mereka juga mengutip studi-studi sebelumnya yang menunjukkan bahwa *forced alignment* lintas bahasa pada bahasa *low-resource* hanya mencapai 61–71% kesesuaian dalam batas 20 milidetik, dibandingkan 93% untuk Bahasa Inggris. Kasus Bahasa Sunda dalam penelitian ini merepresentasikan titik ekstrem dari tren degradasi tersebut.

Tiga faktor spesifik berkontribusi terhadap kegagalan ini. Pertama, durasi audio yang sangat pendek (rata-rata 0,69 detik) memberikan konteks akustik yang terbatas bagi model HMM. Kedua, dataset yang hanya terdiri dari sepuluh kata unik tidak menyediakan variasi fonotaktik yang cukup bagi pelatihan model akustik. Ketiga, ketiadaan model *pretrained* untuk bahasa yang secara fonologis dekat dengan Sunda memaksa pelatihan dari nol, yang menghadapi permasalahan sirkuler: model membutuhkan *alignment* yang baik untuk pelatihan, namun *alignment* yang baik membutuhkan model yang sudah memadai. Hal ini juga dikonfirmasi oleh Cryssiover & Zahra (2024) yang melaporkan bahwa ASR Sunda berbasis Wav2Vec 2.0 pun masih menghasilkan WER 23,5% pada dataset 53 jam, menegaskan tantangan intrinsik bahasa ini sebagai bahasa *low-resource*.

3.5.2 Dampak Kegagalan Alignment terhadap GOP

Hasil penelitian mengkonfirmasi bahwa efektivitas GOP secara fundamental bergantung pada kualitas segmentasi fonem yang mendasarinya (Witt & Young, 2000). Kheir et al. (2023) mengidentifikasi *forced alignment* sebagai *bottleneck* utama dalam *pipeline* penilaian pelafalan, di mana *error alignment* merambat ke skor GOP. Temuan penelitian ini memperkuat observasi tersebut dengan menunjukkan kasus ekstrem: kegagalan *alignment* total menyebabkan model GOP yang sama sekali tidak fungsional.

Dominasi fonem /l/ sebagai *top-1* untuk 14 dari 20 fonem merupakan artefak langsung dari *alignment* yang buruk. Karena /l/ merupakan salah satu dari sedikit fonem yang memiliki variasi durasi pada hasil *alignment* MFA, model GMM /l/ dilatih dari segmen yang lebih beragam dan mengembangkan distribusi yang sangat luas dalam ruang fitur MFCC, memberikan *likelihood* tinggi terhadap hampir semua jenis input akustik. Nilai LLR rata-rata negatif (−3,63) mengkonfirmasi bahwa fonem target bukan merupakan fonem dengan *likelihood* tertinggi pada mayoritas segmen. Kim et al. (2022) menunjukkan bahwa pendekatan berbasis *self-supervised learning* seperti HuBERT Large mampu mencapai PCC 0,82, melampaui GOP (PCC 0,63), mengindikasikan keterbatasan representasi MFCC dan model GMM untuk bahasa *low-resource*.

TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2393-2403
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9869

3.5.3 Implikasi dan Kelayakan Purwarupa

Temuan pengujian fungsional memiliki implikasi penting bagi pengembangan sistem penilaian pelafalan untuk bahasa rendah sumber daya. Getman et al. (2023) melaporkan bahwa bahkan *baseline* berbasis GOP-*like* hanya mencapai UAR 39,05% untuk Bahasa Finlandia dengan data terbatas, dan bahwa penilaian pelafalan untuk bahasa *low-resource* secara inheren sulit bahkan bagi penilai ahli manusia, dengan kesepakatan antar-anotator hanya 44,40%. Konteks ini menunjukkan bahwa kesulitan yang ditemukan bukan kelemahan unik dari implementasi, melainkan refleksi dari tantangan fundamental dalam domain ini.

Dari perspektif teknis, purwarupa berhasil mengimplementasikan seluruh alur kerja yang dirancang. Arsitektur *client-server* menggunakan React Native dan FastAPI merupakan pendekatan yang *viable* untuk sistem penilaian pelafalan. Waktu pemrosesan 5–6 detik per *request* berada di ambang yang dapat ditoleransi untuk aplikasi pembelajaran interaktif, meskipun akan menjadi hambatan pada penggunaan intensif atau *deployment* produksi dengan beban lebih tinggi. Secara keseluruhan, penelitian ini memberikan bukti empiris pertama bahwa *pipeline* MFA-GOP standar tidak dapat diterapkan secara langsung untuk Bahasa Sunda sebagai bahasa rendah sumber daya dengan audio *single-word* berdurasi pendek, dan menetapkan *baseline* empiris untuk penelitian selanjutnya yang mengeksplorasi alternatif seperti model *pretrained* berbasis rumpun Austronesia atau pendekatan *self-supervised learning*.

4. KESIMPULAN

Penelitian ini melaksanakan evaluasi sistematis terhadap *Montreal Forced Aligner* (MFA) dan *Goodness of Pronunciation* (GOP) untuk penilaian pelafalan Bahasa Sunda menggunakan dataset 2.500 sampel ujaran dari 50 penutur asli yang mencakup 10 kosakata dasar dengan 20 fonem unik. Evaluasi MFA menunjukkan kegagalan *alignment* yang bersifat total dan sistemis: seluruh 2.500 *file* (100%) teridentifikasi bermasalah, dengan 17 dari 20 fonem memiliki durasi tepat 10 milidetik (durasi minimum paksa). Tiga konfigurasi parameter yang berbeda menghasilkan tingkat kegagalan identik, mengkonfirmasi bahwa kegagalan bersifat intrinsik akibat kombinasi durasi audio yang sangat pendek (rata-rata 0,69 detik), keterbatasan variasi fonotaktik, dan ketiadaan model *pretrained* untuk bahasa yang secara fonologis dekat dengan Bahasa Sunda. Total data audio (~29 menit) jauh di bawah rekomendasi minimum MFA sebesar 3–5 jam. Evaluasi GOP mengungkapkan bahwa model yang dibangun dari *alignment* yang gagal tidak memiliki kemampuan diskriminatif yang memadai: *top-1 accuracy* hanya 26,1%, fonem /l/ secara anomali mendominasi sebagai fonem *top-1* untuk 14 dari 20 fonem, dan sistem tidak mampu membedakan ucapan benar dari ucapan salah. Pengujian fungsional membuktikan bahwa suara acak mendapatkan skor GOP lebih tinggi daripada ucapan target yang benar, mengkonfirmasi ketidakberfungsian sistem sebagai alat penilaian pelafalan. Purwarupa aplikasi berhasil dari aspek teknis (6 dari 8 skenario *black-box* berhasil), namun nilai praktisnya dibatasi oleh ketidakberfungsian model MFA-GOP. Secara keseluruhan, penelitian ini memberikan tiga kontribusi utama: (1) kontribusi empiris berupa bukti kuantitatif pertama bahwa *pipeline* MFA-GOP standar tidak dapat diterapkan secara langsung untuk Bahasa Sunda sebagai bahasa rendah sumber daya dengan audio *single-word* berdurasi pendek; (2) kontribusi metodologis berupa *baseline* empiris dan kerangka evaluasi yang dapat direplikasi untuk mengukur kelayakan *pipeline forced alignment* dan *goodness of pronunciation* pada bahasa daerah lain di Indonesia; serta (3) kontribusi praktis berupa purwarupa aplikasi *client-server* React Native–FastAPI yang dapat menjadi titik tolak pengembangan sistem penilaian pelafalan Bahasa Sunda berbasis pendekatan alternatif. Untuk penelitian selanjutnya, disarankan perluasan dataset hingga memenuhi rekomendasi minimum MFA (3–5 jam), eksplorasi model MFA *pretrained* dari bahasa yang secara fonologis berkerabat dengan Bahasa Sunda, serta eksplorasi pendekatan alternatif berbasis *self-supervised learning* seperti *wav2vec2* atau HuBERT sebagai kandidat yang lebih sesuai untuk konteks bahasa *low-resource*.

REFERENCES

Arisaputra, P., Handoyo, A. T., & Zahra, A. (2024). XLS-R deep learning model for multilingual ASR on low-resource languages: Indonesian, Javanese, and Sundanese. *ICIC Express Letters, Part B: Applications*, *15*(6), 551–559. https://doi.org/10.24507/icicelb.15.06.551

Ashar, D., & Handayani, R. (2025). Pelestarian Bahasa Daerah Untuk Melestarikan Identitas Di Generasi Muda. *Basaya: Jurnal Bahasa, Sastra Dan Budaya*, *1*(2), 40–43. http://jurnal.inovasipendidikankreatif.com/index.php/BASAYA/article/view/54

Cryssiover, A., & Zahra, A. (2024). Speech recognition model design for Sundanese language using WAV2VEC 2.0. *International Journal of Speech Technology*, *27*(1), 171–177. https://doi.org/10.1007/s10772-023-10066-5

Getman, Y., Phan, N., Al-Ghezi, R., Voskoboinik, E., Singh, M., Grosz, T., Kurimo, M., Salvi, G., Svendsen, T., Strombergsson, S., Smolander, A., & Ylinen, S. (2023). Developing an AI-Assisted Low-Resource Spoken Language Learning App for Children. *IEEE Access*, *11*, 86025–86037. https://doi.org/10.1109/ACCESS.2023.3304274

Kaharuddin, K., Kaharuddin, M. N., & Kaharuddin, N. N. (2024). Penetrasi Bahasa dan Ancaman Kepunahan Bahasa Daerah di Era Komunikasi Digital di Provinsi Sulawesi Selatan. *Jurnal Idiomatik Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, *7*(1), 1–14. https://doi.org/https://doi.org/10.46918/idiomatik.v7i1.2303


TIN: Terapan Informatika Nusantara
Vol 6, No 12, May 2026, page 2393-2403
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v6i12.9869

Kamaly, N., Fuddailah, N., Firsa, P. Z. N., Afrijal, A., & Alqarni, W. (2025). Peran Balai Bahasa Aceh Dalam Meningkatkan Literasi Bahasa Daerah di Kalangan Generasi Muda. *Sosietas: Jurnal Pendidikan Sosiologi*, *15*(1), 101–110. https://doi.org/10.17509/sosietas.v15i1.83694

Kheir, Y., Ali, A., & Chowdhury, S. (2023). Automatic Pronunciation Assessment - A Review. *Findings of the Association for Computational Linguistics: EMNLP 2023*, 8304–8324. https://doi.org/10.18653/v1/2023.findings-emnlp.557

Kim, E., Jeon, J.-J., Seo, H., & Kim, H. (2022). Automatic Pronunciation Assessment using Self-Supervised Speech Representation Learning. *Interspeech 2022*, 1411–1415. https://doi.org/10.21437/Interspeech.2022-10245

Liu, Y., binti Ab Rahman, F., & binti Mohamad Zain, F. (2025). A systematic literature review of research on automatic speech recognition in EFL pronunciation. *Cogent Education*, *12*(1). https://doi.org/10.1080/2331186X.2025.2466288

Mandolang, N. O., Lotulung, D. R., & Ranuntu, G. C. (2024). Reconstruction of the Tontemboan-Indonesian Dictionary: Makela'i and Matana'i. *Santhet (Jurnal Sejarah Pendidikan Dan Humaniora)*, *8*(2), 2510–2516. https://doi.org/10.36526/santhet.v8i2.3287

Mcauliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *Interspeech 2017*, 498–502. https://doi.org/10.21437/Interspeech.2017-1386

Nurjanah, N., Koswara, D., Santosa Nugraha, H., Rukmanah, H. S., & Ruslan, U. (2025). Strategi Inovatif Dalam Pembelajaran Bahasa Sunda: Digitalisasi Materi Ajar untuk Guru Sekolah Dasar. *Jurnal Inovasi Penelitian Pendidikan Dan Pembelajaran*, *5*(2), 579–587. https://doi.org/https://doi.org/10.51878/learning.v5i2.4724

Pratama, R. S. A., & Amrullah, A. (2024). Analysis of Whisper Automatic Speech Recognition Performance on Low Resource Language. *Jurnal Pilar Nusa Mandiri*, *20*(1), 1–8. https://doi.org/10.33480/pilar.v20i1.4633

Rahmah, D. L., & Juhriah, E. (2021). Aplikasi Mengenal Bahasa Sunda Berbasis Android Dalam Dunia Pendidikan. *Jurnal Educatio FKIP UNMA*, *7*(4), 2136–2145. https://doi.org/10.31949/educatio.v7i4.1605

Rusyana, E., & Rohmah, R. U. N. (2024). Interferensi bahasa Indonesia terhadap bahasa Sunda dalam karangan berbahasa Sunda siswa SMP. *Diglosia: Jurnal Kajian Bahasa, Sastra, Dan Pengajarannya*, *7*(2), 237–246. https://doi.org/10.30872/diglosia.v7i2.954

Sulastri Ai, Ali Irfan Muhammad, Adyatma Rafi, Pradana Surya Rahadyan, & Hamidah Siti. (2023). Geolinguistik: Variasi Dialek Dan Lemahnya Pemertahanan Bahasa Sunda Oleh Generasi Muda. *Jurnal Geografi*, *13*(1), 38–46. https://doi.org/10.24036/geografi/vol13-iss1/3970

Tosolini, A., & Bowern, C. (2025). Multilingual MFA: Forced Alignment on Low-Resource Related Languages. In J. Lachler, G. Agyapong, A. Arppe, S. Moeller, A. Chaudhary, S. Rijhwani, & D. Rosenblum (Eds.), *Proceedings of the Eight Workshop on the Use of Computational Methods in the Study of Endangered Languages* (pp. 100–109). Association for Computational Linguistics. https://aclanthology.org/2025.computel-main.11/

Tri Pujiani, N. K. I., & Miftahuddin, Y. (2022). Sistem Automatic Speech Recognition Menggunakan PCA dan VQ Untuk Deteksi Kemiripan Kata Bahasa Sunda. *E-Proceeding FTI*, *1*(1). https://eproceeding.itenas.ac.id/index.php/fti/article/view/964

Witt, S., & Young, S. (2000). Phone-level pronunciation scoring and assessment for interactive language learning. *Speech Communication*, *30*, 95–108. https://doi.org/10.1016/S0167-6393(99)00044-8

Yanti Rut Susanti. (2022). Kurangnya penggunaan dan pemahaman berbahasa Sunda di kalangan remaja. *DEWANTARA: Jurnal Pendidikan Sosial Humaniora*, *1*(3), 74–77. https://doi.org/https://doi.org/10.30640/dewantara.v1i3.403

Zhang, J., Zhang, Z., Wang, Y., Yan, Z., Song, Q., Huang, Y., Li, K., Povey, D., & Wang, Y. (2021). speechocean762: An Open-Source Non-native English Speech Corpus For Pronunciation Assessment. *CoRR*, *abs/2104.01378*. https://arxiv.org/abs/2104.01378