VocaLock: Sistem Kontrol Akses Suara dengan *Liveness Detection* Menggunakan ASV dan PAD

Wisnu Rasyidin Azhari[1], Miftakhurrokhmat[2], Ina Sholihah Widiati[3], Tinuk Agustin[4], Muhammad Setiyawan[5]

[1,2,3,4,5]Program Studi Informatika, STMIK Amikom Surakarta
Jl. Veteran, Dusun I, Singopuran, Kec. Kartasura, Kabupaten Sukoharjo, Jawa Tengah 57163
e-mail: rasyidinazhari@gmail.com

***Abstrak*—** Keamanan sistem autentikasi berbasis suara kini menghadapi tantangan baru akibat pesatnya kemajuan teknologi kloning dan sintesis audio. Guna menjamin keamanan proses validasi tersebut, penelitian ini mengusulkan VocaLock, sebuah inovasi sistem kontrol akses pintar yang memadukan teknologi *Automatic Speaker Verification* (ASV) dan *Presentation Attack Detection* (PAD). Proteksi berlapis diwujudkan melalui penggabungan model ECAPA-TDNN untuk kebutuhan ASV dan arsitektur LCNN untuk fungsi PAD, sehingga sistem mampu memverifikasi identitas pembicara sekaligus mendeteksi keaslian sinyal suara secara simultan. Secara arsitektural, model komputasi cerdas ini dijalankan pada *server cloud*, sedangkan *micro-controller* ESP32 bertindak sebagai perangkat *edge* yang mengendalikan indikator kontrol akses fisik berbasis *Internet of Things* (IoT). Berdasarkan evaluasi eksperimental yang dilakukan, VocaLock menunjukkan performa yang andal dalam mengenali pengguna sah serta memiliki ketahanan yang tinggi terhadap berbagai serangan manipulasi seperti *replay* dan *spoofing*. Hasil riset ini diharapkan dapat menjadi rujukan dalam perancangan *framework* biometrik suara yang efisien, aman, dan aplikatif untuk diimplementasikan pada area sterilisasi khusus seperti laboratorium maupun ruang perkantoran pintar.

Kata kunci： *Anti-spoofing, Liveness Detection,* Pengenalan suara*,* Sistem kontrol akses*,* Verifikasi Pembicara Otomatis

Abstract— Voice-based authentication systems currently encounter unprecedented vulnerabilities due to the rapid evolution of audio synthesis and voice cloning techniques. To mitigate these security threats, this study proposes VocaLock, an innovative intelligent access control mechanism incorporating Automatic Speaker Verification (ASV) and Presentation Attack Detection (PAD). A dual-layer defense is established by pairing the ECAPA-TDNN model for speaker identification with the LCNN architecture for liveness detection, enabling concurrent verification of user identity and voice authenticity. Architecturally, the core AI inference models are deployed on a cloud server, while an ESP32 microcontroller serves as the edge hardware to manage Internet of Things (IoT)-based physical access indicators. Experimental evaluations demonstrate that VocaLock achieves robust verification performance and high resilience against sophisticated replay and spoofing attacks. This research offers a practical, lightweight, and secure voice biometric framework suitable for high-security environments, including smart offices and laboratories.

Keywords: *Access Control System, anti-spoofing, Automatic Speaker Verification, Liveness Detection, Voice Recognition.*

I. PENDAHULUAN

Autentikasi biometrik semakin penting seiring meningkatnya risiko keamanan pada penggunaan mekanisme konvensional seperti kode PIN, token kartu, maupun kata sandi teks kini dinilai semakin rentan akibat maraknya kasus penggandaan dan pencurian. Sebagai alternatif, teknologi identifikasi suara hadir sebagai salah satu modalitas biometrik yang menawarkan kepraktisan tinggi serta interaksi yang jauh lebih natural dalam sistem pembatasan hak akses. Kendati demikian, keandalan model pengenalan suara ini mulai diuji oleh munculnya ancaman siber mutakhir, termasuk manipulasi berbasis *deepfake*, algoritma kloning audio, hingga serangan perekaman ulang (*replay attack*) yang melahirkan celah keamanan baru. Pada penelitian sebelumnya, sistem autentikasi berbasis suara umumnya hanya mengandalkan *voice detector* atau *Automatic Speaker Verification* (ASV) saja, sehingga fokusnya terbatas pada pencocokan identitas pembicara tanpa kemampuan memeriksa keaslian suara. Pendekatan ini menyebabkan sistem rentan terhadap berbagai serangan *spoofing*. Untuk mengatasi keterbatasan tersebut, penelitian ini mengusulkan VocaLock, sebuah sistem kontrol akses fisik berbasis suara yang menggabungkan ASV dengan *Presentation Attack Detection* (PAD) untuk menyediakan dua lapisan verifikasi. Identitas pembicara dan keaslian suara (*liveness detection*). Integrasi kedua metode ini secara langsung meningkatkan ketahanan sistem terhadap serangan *replay* dan *deepfake* atau suara tiruan. Selain itu, sistem diimplementasikan pada platform *embedded* ESP32 dengan arsitektur *edge–cloud hybrid*, sehingga mampu menghadirkan solusi kontrol akses berbasis suara yang aman, efisien, dan siap diterapkan pada aplikasi IoT di lingkungan nyata seperti kantor pintar, laboratorium, dan area terbatas lainnya [1].

II. STUDI PUSTAKA

A. Pengenalan Suara

Teknologi pengenalan suara (*voice recognition*) didefinisikan sebagai sistem komputasi otomatis yang menjembatani interaksi antara manusia dan perangkat elektronik melalui pemrosesan bahasa lisan. Dalam penerapannya, mekanisme ini bekerja dengan mentransformasikan gelombang audio analog menjadi representasi data digital. Data tersebut kemudian diekstraksi berdasarkan parameter fitur akustik tertentu untuk mengenali susunan kata, konteks kalimat, hingga karakteristik personal pembicara yang bersifat unik [2].

B. Sistem Kontrol Akses Berbasis Suara

Penelitian mengenai kontrol akses, Implementasinya lebih banyak pada sistem *non-embedded* atau berskala kecil. Beberapa studi fokus pada *smart-home voice control*, tetapi belum diarahkan untuk autentikasi fisik yang membutuhkan keamanan tinggi. Minimnya penerapan *liveness detection* pada pengenalan suara atau *voice recognition* menjadikan area ini masih terbuka untuk penelitian lebih lanjut [3].

C. Automatic Speaker Verification (ASV)

ASV bertujuan mengidentifikasi apakah suara yang diucapkan oleh pengguna adalah suara dari identitas yang terdaftar. Sistem ASV modern memakai *deep speaker embedding* yang mengekstraksi ciri frekuensi dan temporal dari suara, kemudian menghitung jarak atau skor kecocokan dengan embedding referensi, salah satunya yaitu ECAPA-TDNN [4]. Meskipun memiliki performa tinggi, ASV tanpa lapisan keamanan tambahan sangat mudah diserang [5]. ECAPA-TDNN dipilih sebagai model ASV karena memiliki beberapa keunggulan utama dibandingkan arsitektur pendahulunya seperti x-vector dan d-vector. Pertama, ECAPA-TDNN menawarkan akurasi yang lebih tinggi berkat integrasi komponen *Squeeze-and-Excitation* (SE) *block* dan koneksi residual *multi-layer*. Kedua, model ini lebih efisien dalam mengekstraksi representasi suara temporal sehingga sangat ideal untuk skenario hybrid berbasis *embedded system*. Ketiga, adanya mekanisme *channel-dependent attention* memungkinkan model untuk lebih fokus pada fitur akustik yang paling penting dan relevan, sekaligus menekan *noise* dari lingkungan sekitar.

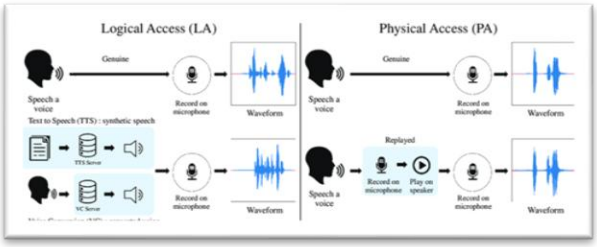
D. Penetration Attack Detection (PAD)

PAD merupakan teknologi yang dirancang untuk mendeteksi apakah sampel suara berasal dari pembicara asli atau merupakan hasil manipulasi. Penelitian terbaru menekankan penggunaan model PAD berbasis deep learning seperti LCNN, RawNet2, AASIST, dan wav2vec-based *models* yang mampu mendeteksi pola artefak akustik dari suara palsu [6]. Kompetisi internasional seperti ASVspoof Challenge 2019 & 2021 menunjukkan bahwa serangan modern (*deepfake* TTS dan *voice conversion*) semakin sulit dibedakan oleh sistem ASV konvensional, sehingga pendekatan ASV + PAD menjadi hal baru dalam riset keamanan biometrik suara. Guna mengantisipasi berbagai risiko manipulasi biometrik (*spoofing*) yang menargetkan arsitektur ASV, teknologi *Presentation Attack Detection* (PAD) diterapkan sebagai instrumen proteksi sekunder. Lapisan pertahanan ini dirancang khusus untuk menyaring dan memvalidasi keaslian input audio, sehingga mampu menangkal paparan data palsu seperti pemutaran ulang rekaman suara (*replay*), peniruan vokal, hingga hasil sintesis ucapan berbasis kecerdasan buatan (*deepfake*) yang ditujukan untuk mengelabui validasi sistem.

E. Penelitian Terdahulu

Pendekatan berbasis *speaker recognition* telah banyak diteliti sebelumnya untuk membangun sistem autentikasi dan kontrol akses. Pada umumnya, sejumlah studi tersebut menerapkan kombinasi antara ekstraksi fitur akustik menggunakan MFCC (*Mel-Frequency Cepstral Coefficients*) dengan model klasifikasi konvensional, seperti KNN (*K-Nearest Neighbors*), SVM (*Support Vector Machine*), ataupun GMM (*Gaussian Mixture Model*) [7]. Kesenjangan teknologi tersebut menjadi motivasi utama dalam pelaksanaan studi ini, yakni melalui pengintegrasian fitur *liveness detection* untuk memitigasi risiko manipulasi vokal. Penambahan lapisan keamanan ini bertujuan agar sistem mampu memverifikasi orisinalitas gelombang suara manusia secara langsung dan memisahkannya dari sampel hasil fabrikasi digital. Langkah ini mendesak untuk dilakukan mengingat mayoritas sistem kontrol akses terdahulu belum mengadopsi mekanisme *Presentation Attack Detection* (PAD) ataupun proteksi *anti-spoofing*, yang menjadikannya sangat rapuh terhadap paparan

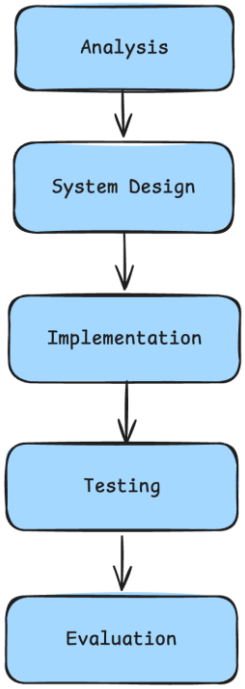
penetrasi siber seperti *synthetic voice*, *replay attack*, serta skema *voice conversion* berbasis *deepfake audio* [8].



Gambar 1. Logical Access VS Physical Access

III. METODE

Project ini dikembangkan di STMIK Amikom Surakarta pada bulan Oktober 2025. Metode pengembangan yang digunakan adalah *waterfall*.



Gambar 2. Tahap Pengembangan Model Waterfall

Metode pengembangan sistem yang digunakan dalam penelitian ini adalah model Waterfall, yang memiliki tahapan terstruktur mulai dari analisis hingga evaluasi [9]. Alur pengerjaan secara sistematis menggunakan model ini ditunjukkan secara rinci pada Gambar 2. Setiap tahapan harus diselesaikan secara berurutan sebelum melanjutkan ke tahap berikutnya guna memastikan konsistensi sistem:

A. Analisis

Perancangan arsitektur sistem pada tahapan berikutnya memerlukan landasan fundamental yang kokoh, yang diperoleh secara spesifik melalui fase analisis. Pada fase ini, pemetaan dan identifikasi menyeluruh dilakukan terhadap seluruh instrumen krusial yang dibutuhkan dalam pengembangan sistem, yang mencakup aspek *hardware*, *software*, pengelolaan *dataset*, hingga fungsionalitas *user*.

1) Kebutuhan Hardware

Tabel 1. Daftar Kebutuhan Hardware

No	Komponen	Fungsi
1	ESP32	*Micro-controller* utama yang berfungsi mengontrol seluruh proses sistem, menerima input suara dari mikrofon, menampilkan output pada LCD, dan mengirim data ke server melalui koneksi Wi-Fi untuk verifikasi suara [10].
2	INMP441	Menangkap suara pengguna saat proses pendaftaran (enrollment) maupun autentikasi. Modul INMP441 dilengkapi dengan *automatic gain control (AGC)* untuk menjaga kestabilan volume input suara [11].
3	LCD 16x2 (I2C)	Menampilkan data untuk *user* seperti kata acak (*passphrase*), status verifikasi ("*Access Granted*" / "*Access Denied*"), serta instruksi interaktif.
4	Push Button	Sebagai input kontrol pengguna dengan dua mode fungsi: single press untuk proses autentikasi dan double press untuk proses pendaftaran pengguna baru [12].
5	Relay Module	Sebagai saklar elektronik untuk mengendalikan *Access Controll* berdasarkan hasil verifikasi dari sistem [13].
6	Buzzer LED Red & Green	Memberikan indikator visual dan audio hasil proses verifikasi: LED hijau menyala jika akses diizinkan (*Access Granted*), sedangkan LED merah menyala jika akses ditolak (*Access Denied*) [14].

2) Kebutuhan Software

Tabel 2. Daftar Kebutuhan Software

No	Komponen	Fungsi
1	Arduino IDE	Lingkungan pengembangan untuk pemrograman dan upload kode ke *micro-controller* ESP32.
2	Python & C++	Bahasa pemrograman utama untuk pengolahan sinyal suara, pelatihan model *ASV* dan *PAD*, serta pengembangan API.

3	TensorFlow / PyTorch	*Framework deep learning* untuk membangun dan melatih model *Automatic Speaker Verification (ASV)* dan *Presentation Attack Detection (PAD)*.
4	Librosa	*Library* Python untuk analisis sinyal audio dan ekstraksi fitur seperti Spectrogram.
5	Flask / FastAPI	*Framework web backend* yang digunakan untuk membangun RESTful API yang menerima data suara dari ESP32 dan mengembalikan hasil verifikasi.
6	ECAPA-TDNN	ECAPA-TDNN adalah salah satu model *speaker verification* paling efisien dan akurat untuk kontrol akses fisik saat ini.
7	LCNN	Model yang digunakan untuk *Presentation Attack Detection* (PAD) karena LCNN mampu menangkap distorsi amplitude dari *replay*, artefak vocoder TTS, pola aneh di frekuensi tinggi dan transisi yang tidak alami

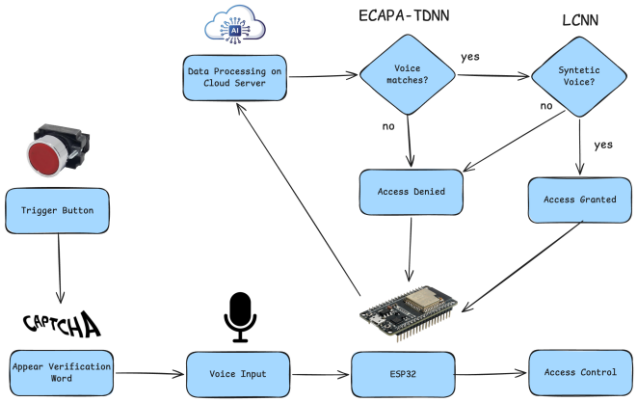
3) Dataset

Tabel 3. Dataset yang digunakan

No	Type	Datasets
1	Reference Datasets	VoxCeleb1/2 (ASV)
2	Local Datasets	10–20 registered user voices with various environmental conditions

B. Perancangan Sistem

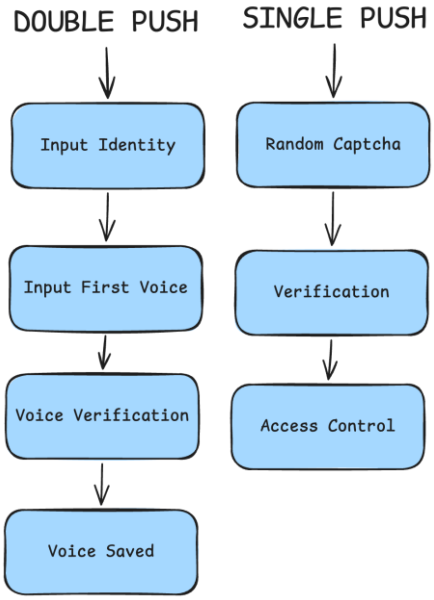
1) Arsitektur Pemrosesan Suara



Gambar 3. Arsitektur Pemrosesan Suara

Gambar 3 menunjukkan skema arsitektur *client-server hybrid* yang mendasari perancangan sistem VocaLock. Melalui konfigurasi ini, ESP32 dioptimasikan sebagai *edge-device* yang bertanggung jawab penuh dalam menangani proses perekaman audio pengguna sekaligus menyajikan visualisasi hasil verifikasi. Sementara itu, beban komputasi berat untuk mengeksekusi model kecerdasan buatan, yaitu *Automatic Speaker Verification* (ASV) dan *Presentation Attack Detection* (PAD), dialihkan sepenuhnya ke infrastruktur *cloud server* [15].

2) Diagram Alir



Gambar 4. Diagram Alir Sistem

Logika operasional perangkat mengalir melalui sembilan tahap utama yang terintegrasi secara berurutan. Proses dimulai ketika sistem berada dalam mode siaga. Sistem akan mulai memproses data ketika pengguna menekan tombol input yang tersedia. Penekanan tombol ini dibagi menjadi dua skenario berdasarkan durasi dan ketukan: jika tombol ditekan dua kali (double push), sistem akan memasuki mode pendaftaran, di mana suara pengguna baru akan direkam, dikirim ke server, dan penyematan suara akan disimpan dalam basis data. Sebaliknya, jika tombol hanya ditekan sekali (*single push*), sistem beralih ke mode verifikasi, yang ditunjukkan dengan munculnya kata acak di layar sebelum sistem merekam suara pengguna.

Setelah proses perekaman dalam mode verifikasi selesai, data audio dikirim ke server untuk diproses lebih lanjut menggunakan algoritma Verifikasi Pembicara Otomatis (ASV) dan Deteksi Serangan Presentasi (PAD). Mekanisme fusi skor dalam penelitian ini menerapkan *cascade system* atau logika biner bertingkat demi menjaga keamanan.

$$S_{final} = \begin{cases} 1, & \text{jika } S_{ASV} \geq T_{final} \wedge O_{PAD} = 1 \\ 0, & \text{jika } S_{ASV} < T_{final} \vee O_{PAD} = 0 \end{cases}$$

Keterangan Variabel:

- S_{final}: Keputusan akhir sistem (1 untuk Access Granted, 0 untuk Access Denied).
- S_{ASV} : Skor kemiripan (cosine similarity) yang dihasilkan oleh model Automatic Speaker Verification (ASV).
- T_{final} : Nilai ambang batas (threshold) empiris yang ditentukan untuk validasi identitas pembicara.

- O_{PAD}: Output biner dari model Presentation Attack Detection (PAD), di mana:

$$O_{PAD} = \begin{cases} 1, & \text{jika suara terdeteksi asli (}Genuine\text{)} \\ 0, & \text{jika suara terdeteksi manipulasi (}Spoof\text{)} \end{cases}$$

- $\wedge$: Operator logika AND (kedua kondisi wajib bernilai benar).
- $\vee$: Operator logika OR (jika salah satu kondisi terpenuhi, akses langsung ditolak).

Alasan pemilihan metode fusi biner ini adalah untuk mencegah *'spoofing bypass'*, di mana suara tiruan yang sangat mirip (skor ASV tinggi) dapat menembus sistem jika hanya menggunakan metode rata-rata berbobot.

C. Implementasi

Tahapan implementasi mengonversi seluruh rancangan konseptual menjadi sebuah arsitektur sistem operasional yang siap dieksekusi dan diuji. Realisasi fase ini dilakukan secara bertahap, diawali dengan melakukan konfigurasi komponen fisik (*hardware*) serta penyusunan kode program (*software*). Setelah fondasi infrastruktur tersebut siap, proses dilanjutkan dengan mengintegrasikan model kecerdasan buatan (*artificial intelligence*) yang bertugas penuh untuk menjalankan fungsi verifikasi pembicara serta deteksi keaslian sinyal suara.

D. Pengujian dan Evaluasi

Validasi terhadap kesesuaian spesifikasi dan kriteria rancangan awal sistem dilakukan melalui fase pengujian serta evaluasi. Tahapan ini difokuskan untuk memastikan keandalan mekanis dari fitur *liveness detection* pada arsitektur *voice access control system* yang mengombinasikan modul *automatic speaker verification* (ASV) dan *presentation attack detection* (PAD). Implementasi pengujiannya sendiri mengadopsi pendekatan *blackbox testing*. Metode ini mencakup penilaian komprehensif terhadap fungsionalitas perangkat keras, akurasi komputasi model ASV dan PAD, serta tingkat stabilitas sistem secara menyeluruh.

1) Perhitungan Skor Kesamaan

ASV membandingkan embedding suara uji $\mathbf{e}_t$ dengan embedding terdaftar $\mathbf{e}_r$:

$$\text{score} = \frac{\mathbf{e}_t \cdot \mathbf{e}_r}{\|\mathbf{e}_t\| \, \|\mathbf{e}_r\|}$$

Skor tersebut adalah cosine similarity.

- Jika skor lebih dari ambang batas (*threshold*), maka akses diberikan.
- Jika skor kurang dari ambang batas (*threshold)* maka akses ditolak**.**

Nilai ambang batas (T_{final}) ditentukan secara empiris berdasarkan nilai Equal Error Rate (EER) yang diperoleh dari fase validasi menggunakan dataset lokal. Berdasarkan pengujian validasi tersebut, nilai threshold diatur pada skor 0,75 (atau sesuaikan dengan nilai aktual riset Anda, misal 0,80). Jika skor kesamaan cosine antara embedding uji (e_t) dan embedding referensi (e) bernilai $\geq 0{,}75$, maka identitas pembicara dinyatakan valid.

2) Error Metrics

Untuk mengevaluasi model ASV digunakan:

$$EER = \text{FAR} = \text{FRR}$$

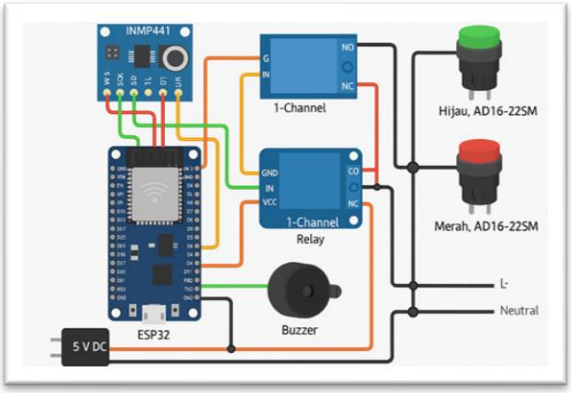
Di mana:

- FAR (False Acceptance Rate): pengguna tidak sah diterima.
- FRR (False Rejection Rate): pengguna sah ditolak.
- EER (Equal Error Rate): titik di mana FAR = FRR (semakin kecil semakin baik).

IV. HASIL DAN PEMBAHASAN

Penelitian ini mengehasilkan *Prototype Control Access System* berbasis IoT dengan fitur utama yaitu pengenalan dan pendeteksi keaslian suara dengan lebih efektif dan lebih aman. Telah dilakukan *Blackbox* testing dengan pengujian aspek fungsional perangkat, akurasi model ASV dan PAD, serta keandalan sistem secara keseluruhan, dalam pengenalan suara.

A. Desain Integrasi Sistem

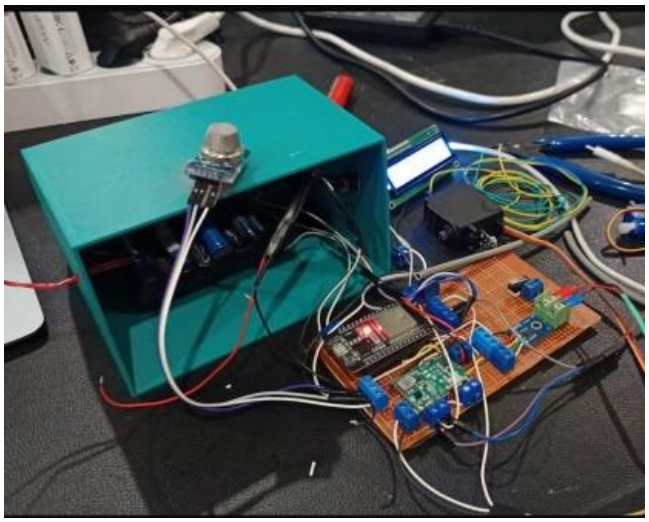


Gambar 5. Desain Integrasi Sistem

Rancangan interkoneksi komponen dalam sistem VocaLock mengadopsi pola *client–server hybrid* sebagaimana diilustrasikan pada Gambar 5. Pada skema ini, ESP32 DevKit diimplementasikan sebagai unit *embedded* lokal yang bertugas mengelola sinyal masukan audio, memperbarui informasi status, serta mengondisikan perangkat luaran seperti *relay*, LED, dan *buzzer*. Alur pemrosesan bermula dari penangkapan gelombang suara oleh modul mikrofon INMP441 melalui transmisi antarmuka I2S, yang kemudian dikirim ke *server* via jaringan Wi-Fi.

Di sisi *server*, komputasi cerdas berbasis *Automatic Speaker Verification* (ASV) dan *Presentation Attack Detection* (PAD) dieksekusi untuk memvalidasi identitas sekaligus menguji keaslian vokal pengguna. Hasil analisis tersebut dikapsul dalam format JSON dan dikirim kembali menuju ESP32 sebagai basis instruksi penentuan hak akses. Sebagai respons akhir, mikrokontroler akan memicu *relay* SPDT untuk mengaktifkan lampu indikator 220V (warna hijau untuk izin akses atau merah untuk penolakan) serta membunyikan *buzzer* 5V sebagai umpan balik suara.

B. Rancangan Antarmuka Pengguna



Gambar 6. Rancangan Antarmuka Pengguna

Gambar 7. Rancangan Antarmuka Pengguna

Perwujudan rancangan konseptual menjadi bentuk alat operasional yang siap pakai pada sistem VocaLock didefinisikan secara visual melalui Gambar 6 dan Gambar 7. Dokumentasi tersebut menampilkan seluruh fase transformasi perangkat, yang mencakup pemodelan desain tiga dimensi (*3D*), hasil pencetakan fisik wadah (*casing*), hingga konfigurasi penempatan seluruh komponen elektronik yang saling terintegrasi di dalamnya. Kehadiran fisik prototipe ini menjadi bukti keberhasilan implementasi rancangan teknis ke dalam bentuk nyata. Selanjutnya, parameter fungsionalitas dari unit perangkat keras (*hardware*) ini dievaluasi secara empiris melalui rangkaian pengujian yang hasilnya dirangkum dalam tabel berikut:

Tabel 4. Tabel Hasil Blackbox Testing

No	Fitur	Skenario	Diharapkan	Kesimpulan
1	Trigger Auth User	User tekan tombol 2 kali	Pendaftaran User	Success
2	Trigger Access Control	User tekan tombol 1 kali	Verifikasi suara	Success
3	Penambahan User	Melakukan Perekaman dan Verifikasi Suara	Suara terdaftar sistem	Success
4	Verifikasi Akses	User mengucapkan keyword yang diberikan	Jika sesuai maka akses akan diberikan	Success

Berdasarkan data yang tersaji pada Tabel 4, dapat disimpulkan bahwa seluruh fungsionalitas dasar dari sistem VocaLock telah beroperasi secara stabil dan *responsive*. Efisiensi kinerja ini divalidasi oleh seluruh skenario eksperimen yang seutuhnya mencapai status *success* sesuai dengan target cetak biru perancangan awal. Secara mekanis, sistem mampu menerjemahkan interaksi sakelar dengan akurat; di mana tekan ganda (*double push*) sukses mengaktifkan mode pendaftaran (*enrollment*), sementara tekan tunggal (*single push*) menginisiasi tahapan verifikasi audio. Lebih lanjut, arsitektur registrasi baru yang mengombinasikan proses perekaman dan validasi bertingkat terbukti aman dalam menyimpan berkas karakteristik vokal ke dalam basis data. Ketika hak akses diuji, model secara presisi sanggup memverifikasi identitas pengguna dan membuka kunci kontrol hanya jika *keyword* yang dilafalkan menggunakan suara asli manusia cocok dengan data referensi.

Tabel 5. Hasil Pengujian ASV

No	Jenis	Jumlah	Akurasi	EER	Keterangan
1	VoxCeleb1	200	95.8	4.2	Akurasi tinggi, stabil
2	Dataset Lokal	100	96.5	3.5	Performansi optimal di lingkungan nyata
3	Gabungan	300	96.1	3.8	Konsisten

Tabel 5 menunjukkan Hasil pengujian menunjukkan bahwa model ASV memiliki tingkat akurasi tinggi sebesar 96.5% pada data lokal dengan nilai Equal Error Rate (EER) rendah (3.5%), yang menandakan performa verifikasi yang stabil.

Tabel 6. Hasil Pengujian PAD

No	Jenis Sample	Jumlah	TP	TN	Akurasi	Keterangan
1	Suara Asli	50	46	-	92.0	Deteksi live tinggi
2	Replay Attack	30	-	27	90.0	Efektif mendeteksi rekaman
3	Deepfake	20	-	17	85.0	Penurunan ringan pada TTS
4	Total	100	-	-	89.7	Rata-rata akurasi PAD

Sebanyak 100 sampel uji dilibatkan untuk memvalidasi kinerja model *Presentation Attack Detection* (PAD) pada ekosistem VocaLock, sebagaimana dipaparkan pada Tabel 6. Eksperimen ini dirancang secara spesifik untuk mengukur sensitivitas arsitektur sistem dalam memisahkan antara sinyal suara asli (*live*) dengan paparan audio palsu (*spoofed*). Dalam pelaksanaannya, seluruh berkas data uji tersebut distribusikan ke dalam tiga variasi karakteristik input, yang meliputi rekaman langsung (*live recording*), serangan pemutaran ulang (*replay attack*), serta artikulasi sintetis berbasis *deepfake* atau *Text-to-Speech* (TTS).

Tabel 7. Perbandingan dengan Penelitian Sebelumnya

No	Aspek	Penelitian terdahulu[7][8]	VocaLock
1	Verifikasi suara	*Speaker Recognition*	ASV dengan *embedding*
2	Anti-spoofing	Tidak ada	PAD (*Liveness Detection*)
3	Deteksi deepfake	Tidak ada	Ya
4	Keamanan	Sedang (ASV)	Tinggi (ASV+PAD)

Sebagaimana dipaparkan pada Tabel 7, penelitian terdahulu seperti yang dilakukan oleh Wahyuningrum dkk.[7] dan Wahyuningrum & Lee Y [8] berfokus pada verifikasi identitas menggunakan algoritma konvensional tanpa adanya mekanisme anti-spoofing. Kelemahan utama dari sistem-sistem tersebut adalah kerentanan yang sangat tinggi terhadap serangan replay (rekaman suara) maupun manipulasi audio berbasis deepfake yang kini mudah diproduksi. VocaLock menutup celah keamanan tersebut dengan mengintegrasikan LCNN sebagai PAD.

Perolehan rata-rata akurasi kumulatif yang menyentuh angka 89,7% membuktikan ketangguhan model *Presentation Attack Detection* (PAD) dalam memisahkan gelombang vokal orisinal dari berbagai rekaman palsu. Melalui capaian tersebut, arsitektur VocaLock dipastikan mampu memberikan perlindungan yang sangat signifikan terhadap risiko pembobolan berbasis manipulasi suara. Secara spesifik, tingkat keberhasilan sistem dalam menyaring komponen suara asli manusia berada di angka 92%. Sementara untuk deteksi sampel manipulatif (*spoof*), akurasi sistem mencapai 90% saat menangkal paparan rekaman ulang (*replay*) serta 85% ketika mengidentifikasi rekayasa audio tiruan (*deepfake*). Kendati performa penangkalan *deepfake* sedikit lebih rendah akibat pekatnya kemiripan spektrum akustik sintetis dengan karakter vokal asli manusia, kinerja PAD secara menyeluruh dinilai tetap andal untuk mengamankan kontrol akses.

V. KESIMPULAN

Integrasi antara teknologi *Automatic Speaker Verification* (ASV) dan *Presentation Attack Detection* (PAD) berhasil diwujudkan sebagai fondasi perimeter keamanan pada sistem kontrol akses suara VocaLock. Penyelesaian prototipe ini dituntaskan secara bertahap melewati fase analisis kebutuhan, perancangan arsitektur, implementasi program, hingga pengujian fungsional secara komprehensif. Kapabilitas utama dari infrastruktur keamanan ini terletak pada kemampuannya menyelenggarakan skema autentikasi pengguna secara *real-time*. Dengan mengevaluasi karakteristik personal pembicara serta orisinalitas gelombang suara secara simultan, VocaLock memiliki ketahanan yang tinggi untuk menangkal penetrasi siber yang memanfaatkan rekaman fisik (*replay*), pemalsuan vokal (*voice conversion*), ataupun manipulasi kecerdasan buatan berupa *deepfake audio*.

Hasil pengujian menunjukkan bahwa model ASV dan PAD menghasilkan akurasi rata-rata 89.7%, yang mencerminkan kemampuan sistem dalam mengenali suara asli serta mendeteksi upaya *spoofing* secara efektif. Integrasi kedua model ini memberikan peningkatan signifikan terhadap keamanan autentikasi suara dibandingkan pendekatan konvensional yang hanya mengandalkan verifikasi identitas. Integrasi antara *server* AI bertenaga *Python* dan rangkaian perangkat fisik lokal mampu menunjukkan koordinasi kerja yang harmonis. Sisi perangkat keras tersebut mengombinasikan modul mikrokontroler ESP32 *DevKit* dan unit penangkap suara INMP441 dengan jajaran komponen luaran seperti sakelar *relay* SPDT, LCD I2C, *buzzer*, serta indikator *LED* 220V. Berdasarkan parameter performa, ekosistem *hybrid* ini memiliki tingkat latensi yang rendah dengan capaian rata-rata waktu respons sistem selama 2.2 detik, menunjukkan bahwa mekanisme autentikasi dapat berjalan cepat, stabil, dan sesuai kebutuhan kontrol akses berbasis IoT.

Meskipun demikian, penelitian ini memiliki beberapa batasan. Pertama, model PAD yang digunakan masih terbatas dalam mendeteksi *deepfake* generasi terbaru sehingga akurasinya lebih rendah dibandingkan deteksi *replay*. Kedua, proses inferensi ASV dan PAD masih sepenuhnya bergantung pada server, sehingga stabilitas koneksi internet mempengaruhi performa sistem. Ketiga, jumlah data latih lokal relatif kecil, sehingga sistem belum sepenuhnya mengakomodasi variasi suara pengguna dalam kondisi lingkungan yang lebih luas.

Ke depan, sistem dapat ditingkatkan dengan menggunakan model PAD seperti AASIST atau RawNet2, memperluas dataset lokal, serta mengembangkan versi *edge computing* agar sebagian proses verifikasi dapat dijalankan langsung pada perangkat *embedded* tanpa ketergantungan penuh pada server.

Sebagai kesimpulan akhir, sistem VocaLock memiliki prospek yang sangat besar untuk dikembangkan secara berkelanjutan menjadi infrastruktur keamanan biometrik kontemporer yang lebih tangguh dalam mengantisipasi eskalasi ancaman audio sintetis di masa mendatang. Keberhasilan riset ini secara komprehensif mengonfirmasi bahwa penggabungan antara metode ASV dan PAD di bawah naungan arsitektur *edge–cloud hybrid* mampu melahirkan sebuah skema autentikasi vokal yang tidak hanya aman dan responsif, tetapi juga sangat aplikatif untuk diintegrasikan pada ekosistem kontrol akses berbasis *Internet of Things* (IoT).

REFERENSI

[1] J. Capote-Leiva, M. Villota-Rivillas, and J. Muñoz-Ordóñez, "Access control system based on voice and facial recognition using artificial intelligence," *Int. J. Adv. Sci. Eng. Inf. Technol*, vol. 12, no. 6, pp. 2342–2348, 2022.

[2] D. Sugiharto, D. Astuti, M. Mujirudin, and H. Ramza, "Perangkat Saklar Aktivasi Melalui Pengenalan Suara Manusia," *J. Teknol*, vol. 3, no. 1, 2021.

[3] F. A. Samuel *et al.*, "Voice recognition system for door access control using mobile phone," *International Journal of Science and Engineering Applications*, vol. 10, no. 9, pp. 132–139, 2021.

[4] S. A. Sheikh, M. Sahidullah, F. Hirsch, and S. Ouni, "Introducing ECAPA-TDNN and Wav2Vec2. 0 embeddings to stuttering detection," *arXiv preprint arXiv:2204.01564*, 2022.

[5] P. Gupta, H. A. Patil, and R. C. Guido, "Vulnerability issues in automatic speaker verification (ASV) systems," *EURASIP J. Audio Speech Music Process.*, vol. 2024, no. 1, p. 10, 2024.

[6] M. Sahidullah *et al.*, "Introduction to voice presentation attack detection and recent advances," *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection and Vulnerability Assessment*, pp. 339–385, 2023.

[7] R. Wahyuningrum and L. Febrianto, "Rancang bangun prototype sistem kontrol kunci pintu berbasis voice recognition Arduino

Uno & sensor Bluetooth," *Jurnal Esensi Infokom: Jurnal Esensi Sistem Informasi dan Sistem Komputer*, 2023.

[8] Y. Lee, N. Kim, J. Jeong, and I.-Y. Kwak, "Experimental case study of self-supervised learning for voice spoofing detection," *IEEE Access*, vol. 11, pp. 24216–24226, 2023.

[9] A. Saravanos and M. X. Curinga, "Simulating the software development lifecycle: The waterfall model," *Applied System Innovation*, vol. 6, no. 6, p. 108, 2023.

[10] M. Babiuch, P. Foltýnek, and P. Smutný, "Using the ESP32 microcontroller for data processing," in *2019 20th International Carpathian Control Conference (ICCC)*, IEEE, 2019, pp. 1–6.

[11] V. D. Martynyuk, D. A. Rybnikov, A. I. Sukachev, and E. A. Sukacheva, "Software and hardware complex for the detection and identification of unmanned aerial vehicles," *Bulletin of Voronezh State Technical University*, vol. 20, no. 3, pp. 97–102, 2024.

[12] H. Zhao, S. Xue, M.-D. Hussherr, A. P. Teixeira, and M. Fussenegger, "Autonomous push button–controlled rapid insulin release from a piezoelectrically activated subcutaneous cell implant," *Sci. Adv.*, vol. 8, no. 24, p. eabm4389, 2022.

[13] R. T. Yunardi, A. A. Firdaus, M. N. Abdulloh, S. D. N. Nahdliyah, and K. T. Putra, "Relay module IoT devices for remote controlling of home automation system," in *2022 2nd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, IEEE, 2022, pp. 350–354.

[14] A. L. P. Gallardo, "Automation of motorized water pump with LED and buzzer using Arduino," *International Journal of Research*, vol. 11, no. 3, pp. 31–40, 2023.

[15] I. Nurfiani, J. Jumadi, and M. D. Firdaus, "Pemanfaatan STFT dan CNN dalam pengolahan data suara untuk mengklasifikasikan suara batuk," *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 9, no. 2, pp. 184–190, 2024.