A Three-Layer Behavioral Authentication Architecture for Resource-Constrained Embedded Devices: Design and Simulation-Based Feasibility Study

Nusratul Islam Mim
Department of Computer Science and Engineering
Jatiya Kabi Kazi Nazrul Islam University
Trishal, Mymensingh, Bangladesh
nusratulislammim@gmail.com

Firuze Tasnim Sneha
Department of Computer Science and Engineering
Jatiya Kabi Kazi Nazrul Islam University
Trishal, Mymensingh, Bangladesh
firuzetasnimsneha@gmail.com

Md. Al Mosabbir Hosen
Department of Computer Science and Engineering
Jatiya Kabi Kazi Nazrul Islam University
Trishal, Mymensingh, Bangladesh
mosabbirsaif@gmail.com

Md. Sujan Ali
Department of Computer Science and Engineering
Jatiya Kabi Kazi Nazrul Islam University
Trishal, Mymensingh, Bangladesh
sujan_cse@jkkniu.edu.bd

***Abstract*—Behavioral authentication has been investigated for physical access control using interaction characteristics such as pressure, motion, timing, and acoustic responses. However, many multimodal behavioral systems target smartphones, wearables, or other platforms with substantially greater memory and computational resources than inexpensive standalone microcontrollers. This paper presents a low-cost three-layer behavioral authentication architecture designed for offline operation on an ATmega328P-class microcontroller with 2 KB SRAM and 32 KB flash. The architecture combines (i) text-dependent spoken-digit verification using MFCC features and Dynamic Time Warping (DTW), (ii) ultrasonic measurement of radial hand-withdrawal distance during inter-press intervals, and (iii) Force–Hold–Gap (FHG) mechanical interaction patterns measured using a single force-sensitive resistor. Sequential screening provides early rejection, followed by weighted score fusion for attempts that pass all screening stages. A Monte Carlo simulation using 50 synthetic user profiles, with a held-out user-disjoint 20-profile test population, produced FAR = 1.67%, FRR = 4.00%, AER = 2.83%, and accuracy = 96.78% for the three-layer system. Ablation analysis indicated that voice and mechanical layers provided the strongest simulated discrimination, while the spatial gap layer showed weak standalone performance (AER = 34.83%) because the synthetic generator did not encode person-specific spatial structure. Although inclusion of the spatial layer changed the simulated decision outcomes, the simulation does not establish that hand-withdrawal distance is an independently discriminative behavioral biometric. The itemized component-level bill of materials is $26.95 (two-layer configuration: $25.45). These results are simulation-based feasibility estimates rather than measurements of a physical prototype or human population. The primary contribution is an embedded architectural design integrating heterogeneous behavioral evidence under a low-cost, memory-constrained, offline-operation design target. Hardware resource usage, power consumption, latency, human variability, and adversarial resistance remain to be experimentally validated.**

***Index Terms*—Behavioral authentication, embedded security, multimodal authentication, voice verification, ultrasonic sensing, force sensing, DTW, resource-constrained systems, offline access control.**

I. Introduction

Physical access-control systems must balance security, usability, cost, and implementation complexity. Conventional PIN-based mechanisms are inexpensive but can expose authentication secrets through observation attacks such as shoulder surfing [1]. Possession-based mechanisms such as RFID introduce additional attack surfaces, including relay and eavesdropping attacks [2]. Physiological biometric systems can provide stronger identity evidence, but they require dedicated sensing hardware and introduce additional cost and biometric-template management requirements [3].

Behavioral authentication provides another design direction by characterizing how a user performs an interaction rather than relying only on a memorized secret. Prior systems have investigated typing pressure, grip force, hand movement, acoustic sensing, and combinations of behavioral signals [4]–[8]. Physical-access systems such as SmartHandle and Step & Turn further demonstrate that behavioral signals can be incorporated directly into access-control interactions [9], [10]. However, relatively little attention has been given to the joint design of heterogeneous behavioral sensing, lightweight fusion, and memory-aware implementation constraints for very low-cost standalone microcontrollers. In particular, it remains unclear which behavioral modalities provide useful complementary evidence when evaluated under a strict embedded resource budget.

979-8-3195-2804-9/26/$31.00 ©2026 IEEE

This paper investigates a three-layer architecture in which the user speaks one digit while simultaneously pressing the force-sensitive resistor, followed by a hand-withdrawal gap between consecutive presses. This press-and-gap interaction is performed across three presses, producing two inter-press gaps. The target is an ATmega328P-class MCU with 2 KB SRAM, 32 KB flash, and 1 KB EEPROM; hence, compact representations, simple similarity functions, sequential screening, and offline operation are emphasized. The study is a simulation-based feasibility case study rather than a validated biometric deployment.

The main contributions are:

1) A three-layer architecture combining spoken-digit MFCC-DTW verification, ultrasonic hand-withdrawal sensing, and single-FSR Force–Hold–Gap interaction patterns.
2) A sequential screening and weighted score-fusion policy with early rejection, evaluated through systematic ablation and sensitivity analysis.
3) A held-out Monte Carlo simulation with 50 synthetic user profiles under controlled sensor-noise and behavioral-variation assumptions, including diagnosis of spatial-layer limitations.
4) A component-level cost analysis and preliminary resource-feasibility assessment targeting a $30 component-cost envelope and ATmega328P-class resources.

The work is a simulation-based feasibility study; no physical prototype or human-subject experiment generated the reported metrics.

II. Related Work

A. Behavioral Authentication for Physical Access

Behavioral authentication has been explored for physical access using motion, pressure, grip, and timing. SmartHandle reported TAR = 87.27% at FAR = 1.39% with 11 users [9], while Step & Turn reported TAR = 97.25% at FAR = 0.01% with 40 participants [10]. Both use sensing configurations and evaluation settings that differ from the proposed ATmega328P-class target.

B. Pressure and Acoustic Behavioral Signals

Pressure-based typing and grip-force characteristics have also been used for authentication [4], [5]. PushPIN reported EER = 0.51% using pressure, wrist motion, and a passcode on smartwatches [6]. HandLock used inaudible acoustic sensing and reported TPR = 96.51% at FAR = 0.82% [7]. The proposed system instead combines spoken-digit verification, ultrasonic hand withdrawal, and FHG mechanical interaction on a standalone MCU.

C. Multimodal Authentication and Score Fusion

Multimodal biometric fusion frameworks have been extensively studied. Information fusion in biometrics can substantially improve decision robustness by combining complementary modalities [8], [11]. Cascade-based early-rejection architectures have also been explored to reduce unnecessary downstream processing [12]. Recent work continues to examine lightweight behavioral authentication for constrained IoT devices [13], and emerging evaluation frameworks emphasize the difficulty of comparing behavioral-authentication systems across heterogeneous sensors and protocols [14]. This work applies these fusion principles within tighter constraints on cost, memory, and computation.

TABLE I
Illustrative Component Cost Estimate

Component	Part	USD
Microcontroller	ATmega328P-PU	4.20
Force sensor	FSR-402	6.50
Microphone	MAX4466	2.80
Ultrasonic ranger	HC-SR04	1.50
Servo	SG90	2.95
Battery	5000 mAh Li-ion	3.50
Misc. electronics/PCB	—	5.50
Total (3-layer / 2-layer)		**26.95 / 25.45**

Estimated component prices are illustrative design estimates and exclude labor, enclosure fabrication, shipping, and manufacturing overhead.

III. System Design

A. Hardware Architecture

The system uses an ATmega328P-class MCU at 16 MHz with 2 KB SRAM, 32 KB flash, 1 KB EEPROM, and a 10-bit ADC. Three behavioral signals are collected during one authentication interaction.

B. Authentication Interaction

The authentication interaction consists of three interleaved steps, where voice and mechanical modalities are collected concurrently:

1) **Step 1:** The user simultaneously speaks Digit 1 and applies Mechanical Press 1 to the FSR.
2) **Gap 1:** The user withdraws their hand; the ultrasonic sensor measures radial distance change between Press 1 and Press 2.
3) **Step 2:** The user simultaneously speaks Digit 2 and applies Mechanical Press 2 to the FSR.
4) **Gap 2:** The user withdraws their hand; the ultrasonic sensor measures radial distance change between Press 2 and Press 3.
5) **Step 3:** The user simultaneously speaks Digit 3 and applies Mechanical Press 3 to the FSR.

This interleaved design is intended to capture all three modalities within a single continuous interaction. In the simulation, the voice energy level is parametrically coupled to the simulated press force via $E_{\text{voice}} = E_{\text{baseline}} + 0.05\sqrt{F_{\text{press}}}$; this is a synthetic modeling assumption, not an experimentally established physiological relationship.

The hardware uses a MAX4466 microphone, HC-SR04, FSR-402, SG90 servo, and battery; the estimated component cost is given in Table I.

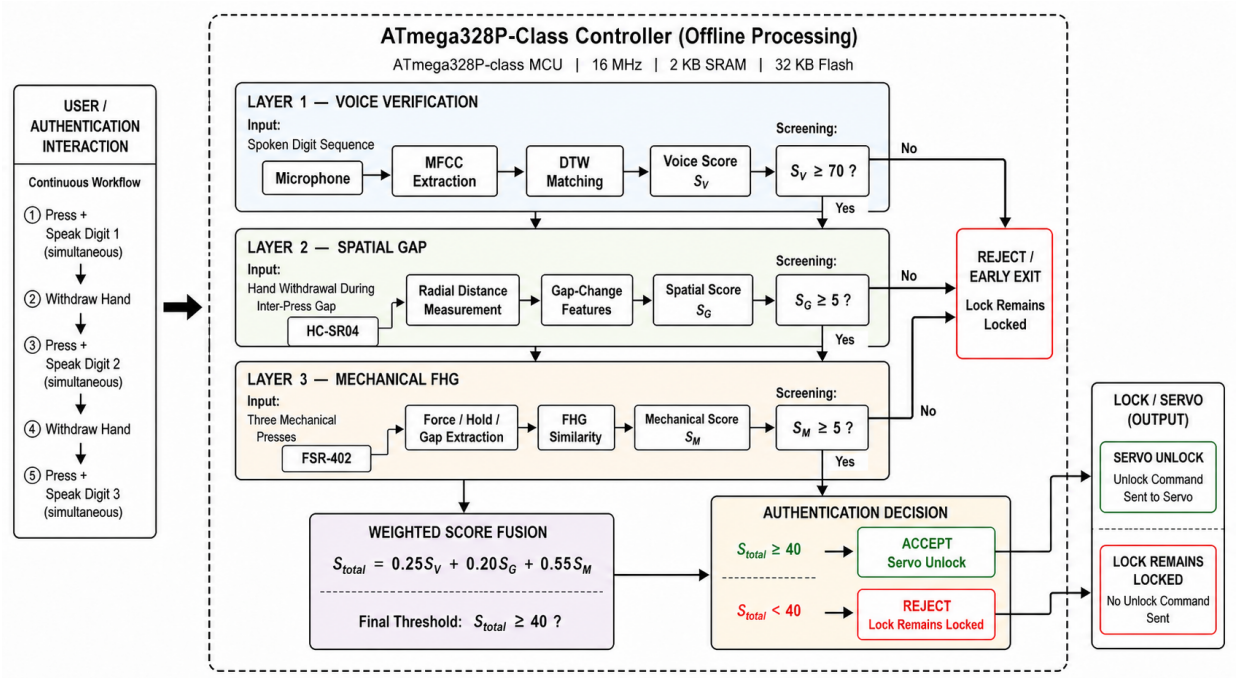


Fig. 1. System architecture of the proposed three-layer behavioral authentication system. Spoken-digit voice verification, ultrasonic hand-withdrawal behavior, and FHG mechanical interaction are evaluated sequentially with early rejection, followed by weighted score fusion and final lock actuation.

C. Authentication Layers

1) Layer 1: Voice Verification: During each of the three presses, the user speaks one digit. The recorded signal is sampled at 8 kHz and represented using MFCC features. DTW measures temporal similarity between the observed digit sequence and the enrolled template. The voice similarity score is normalized to a range of approximately 0–100:

$$S_V = 100 \exp\left(-\frac{D_{\mathrm{DTW}}}{0.3}\right), \tag{1}$$

with screening condition $S_V \geq \tau_V$, $\tau_V = 70$. The voice layer is text-dependent rather than open-vocabulary speaker recognition.

2) Layer 2: Spatial Gap: The HC-SR04 measures hand distance during the two inter-press gaps (between Press 1–Press 2 and Press 2–Press 3). For gap i, the observed distance change is compared with the enrolled reference:

$$S_G = \frac{100}{2} \sum_{i=1}^{2} \exp\left(-\frac{|d_i^{obs} - \bar{d}_i|}{25}\right), \tag{2}$$

with screening condition $S_G \geq \tau_G$, $\tau_G = 5$. The spatial layer captures radial hand-distance changes during the two inter-press intervals rather than performing full 3-D gesture recognition.

3) Layer 3: Mechanical FHG: The FSR records peak force F_j, hold duration H_j, and inter-press gap G_j for each of three presses. The normalized distance from the enrolled reference is

$$d_j = \sqrt{\left(\frac{F_j - \bar{F}_j}{\sigma_{F_j}}\right)^2 + \left(\frac{H_j - \bar{H}_j}{\sigma_{H_j}}\right)^2 + \left(\frac{G_j - \bar{G}_j}{\sigma_{G_j}}\right)^2}. \tag{3}$$

The mechanical similarity score is $S_M = (100/n) \sum_{j=1}^{n} \exp(-d_j^2)$ with $n = 3$, and screening condition $S_M \geq \tau_M$, $\tau_M = 5$.

D. Sequential Screening and Weighted Fusion

The authentication policy consists of two stages. First, the layers are evaluated sequentially:

$$\text{Voice} \rightarrow \text{Spatial Gap} \rightarrow \text{Mechanical}. \tag{4}$$

An attempt that fails a screening threshold is rejected immediately. Thus, later layers do not need to be evaluated for clearly inconsistent attempts. The relatively low screening thresholds for the spatial and mechanical scores ($\tau_G = \tau_M = 5$ on a 0–100 scale) reflect their role as permissive pre-screening gates rather than final authentication criteria; their purpose is primarily to eliminate clearly inconsistent attempts before fusion rather than to provide independent final decisions. The final decision is made only after all three layers pass and the weighted score satisfies $\tau_T = 40$.

Attempts that pass all three screening stages proceed to weighted score fusion:

$$S_{\mathrm{total}} = 0.25S_V + 0.20S_G + 0.55S_M, \tag{5}$$

with the final authentication decision

$$\text{Accept} \iff S_{\text{total}} \geq 40. \tag{6}$$

The screening thresholds, fusion weights, and final decision threshold were selected using only the 30 training users through grid search, minimizing FAR subject to FRR ¡ 5%. All parameters were then fixed and applied unchanged to the 20 held-out test users. No test-set information was used during parameter selection. The FRR constraint is a simulation-design criterion, not an empirical usability target.

E. Embedded Resource Feasibility

The target MCU provides only 2 KB SRAM and 1 KB EEPROM. Storing all MFCC enrollment sequences (3 digits $\times$ 3 repetitions $\times$ 49 frames $\times$ 13 coefficients) requires

$$3 \times 3 \times 49 \times 13 = 5{,}733 \tag{7}$$

coefficient values per user before metadata. If represented using 16-bit values, this corresponds to approximately 11.2 KiB, substantially exceeding the 1 KB EEPROM budget.

The architecture is therefore not yet a fully memory-optimized firmware implementation. Future work must use template compression, reduced frame counts, external non-volatile memory, or similar methods. A preliminary analytical estimate gives a banded-DTW working-buffer requirement of approximately 582 bytes; this estimate has not been verified through compiled firmware or on-device measurement. Because the estimated component cost exceeded the project's available procurement resources, no physical prototype was assembled; consequently, execution time, memory usage, and power consumption remain unmeasured and are reserved for future hardware validation.

F. Parameter Basis

Parameters fall into three categories. *Hardware specifications:* MCU clock, ADC resolution, sensor ranges. *Literature-informed design choices:* 8 kHz sampling, 13 MFCC coefficients, FSR tolerance conventions. *Synthetic simulation assumptions:* noise distributions, $E_{\text{voice}}(F_{\text{press}})$ coupling, impostor observation accuracy. No parameter was treated as experimentally measured unless explicitly stated.

IV. Simulation-Based Performance Evaluation

A. Simulation Methodology

The evaluation is entirely simulation-based. No physical sensors and no human participants were used to produce the reported authentication metrics.

A Monte Carlo simulation generated 50 synthetic user profiles, each assigned user-specific baselines for voice, force, and timing. Genuine attempts applied within-user perturbations plus sensor noise; impostor attempts sampled a different profile's baseline plus execution variability. The 50 profiles were split into 30 training and 20 user-disjoint held-out test profiles. The training set contained 900 genuine and 450 impostor attempts; the held-out test set contained 600 genuine and 300 impostor attempts.

TABLE II
Held-Out Simulation Performance and Ablation Study

Configuration	FAR (%)	FRR (%)	AER (%)	Accuracy (%)
Individual Layer Baselines				
Voice only	4.67	3.67	4.17	96.00
Spatial Gap only†	62.33	7.33	34.83	74.33
Mechanical only	14.67	19.17	16.92	82.33
Two-Layer Combinations				
Voice + Gap	4.67	3.67	4.17	96.00
Voice + Mechanical	2.00	5.17	3.58	95.89
Gap + Mechanical	20.67	8.67	14.67	87.33
Complete Three-Layer System				
All 3 Layers	**1.67**	**4.00**	**2.83**	**96.78**

†Weak standalone spatial-gap performance analyzed in Section IV-D.

B. Sensor and Behavioral Noise Models

The simulation introduces both user-dependent variability and sensor-level noise. For force measurements,

$$F_{obs} = F_{true} + \mathcal{N}(0, \sigma_{user}) + \mathcal{N}(0, 0.03F_{true}) + \mathcal{U}(-0.5, 0.5). \tag{8}$$

For ultrasonic distance,

$$d_{obs} = d_{true} + \mathcal{N}(0, 3) + \mathcal{N}(0, 0.008d_{true}). \tag{9}$$

These models provide controlled perturbations around synthetic templates but do not reproduce effects such as microphone placement, enclosure resonance, temperature drift, sensor aging, hand occlusion, speech accent, fatigue, or learning effects.

C. Overall Authentication Performance

Interpretation: The ablation study indicates that voice and mechanical layers provide the strongest simulated discrimination. Voice+Mechanical achieves AER = 3.58%; adding the spatial layer changes the simulated AER to 2.83%; this does not establish independent spatial biometric discrimination. Because the spatial generator does not model person-specific spatial behavior, this improvement should not be interpreted as evidence that the spatial feature is intrinsically person-specific.

For the complete three-layer system, TP = 576, FN = 24, FP = 5, TN = 295, yielding FAR = 1.67%, FRR = 4.00%, AER = 2.83%, accuracy = 96.78%, and balanced accuracy = 97.17%. Under the present synthetic model, the Voice+Gap configuration produced the same test decisions as voice alone, consistent with the weak person-specific spatial structure.

D. Weight Sensitivity Analysis

The sensitivity analysis is descriptive only; the final configuration was fixed using the training population before held-out evaluation.

E. Spatial Gap Layer: Performance Analysis and Root Cause

The spatial gap layer exhibited weaker standalone discrimination (FAR = 62.33%, AER = 34.83%) compared to voice (AER = 4.17%) and mechanical (AER = 16.92%). To

TABLE III
FUSION WEIGHT SENSITIVITY ANALYSIS (TEST SET, τ_T=40)

Weight config.	Weights ($w_V/w_G/w_M$)	FAR (%)	FRR (%)	Acc. (%)
Equal	0.33/0.33/0.34	3.67	3.67	96.33
Voice-heavy	0.50/0.25/0.25	4.67	3.67	96.00
Spatial-heavy	0.25/0.50/0.25	2.67	3.83	96.56
Mechanical-heavy	0.25/0.25/0.50	2.33	3.83	96.67
Selected	**0.25/0.20/0.55**	**1.67**	**4.00**	**96.78**

understand this weakness, we analyzed the distribution of raw spatial features.

TABLE IV
SPATIAL FEATURE DISTRIBUTION: GENUINE VS. IMPOSTOR POPULATIONS

Feature	Genuine Mean (Std)	Impostor Mean (Std)
Gap distance 1 (mm)	52.5 (22.3)	53.7 (29.6)
Gap distance 2 (mm)	50.7 (18.6)	50.8 (26.6)

The genuine and impostor means differ by only 1.2 mm (Gap 1) and 0.1 mm (Gap 2), with substantial overlap given the large standard deviations. This is a consequence of the synthetic data model: gap distances were drawn independently from a uniform distribution, with no person-specific structure. Consequently, the genuine template and impostor attempts are statistically indistinguishable by construction. This does not establish whether real human users exhibit person-specific hand-withdrawal patterns; that remains an empirical question.

Contribution to fusion. Comparing Voice+Mechanical (FAR = 2.00%, FRR = 5.17%) with the three-layer system (FAR = 1.67%, FRR = 4.00%), adding spatial measurements changes simulated decision outcomes and reduces both FAR and FRR. However, because the synthetic spatial generator does not encode person-specific structure, this should be interpreted as a property of the simulated decision policy rather than evidence of independently discriminative spatial behavior.

F. Comparison with Existing Behavioral Authentication Systems

The primary distinction is the target deployment constraint: the proposed architecture is designed around a low-cost ATmega328P-class platform and offline operation, whereas the cited systems use richer sensing and/or more capable platforms.

V. SECURITY CONSIDERATIONS

The intended threat model is an opportunistic attacker attempting to gain physical access through observation, guessing, imitation, or replay.

The system is not replay-resistant or spoof-proof. Voice replay remains a vulnerability because the current design uses a fixed spoken-digit sequence. Randomized challenge digits could reduce direct replay risk, but would require an enrollment and template-management strategy that fits within the MCU's limited non-volatile memory; this has not yet been designed or evaluated. Trained mimicry, robotic replay, and sensor spoofing also remain unvalidated. These threats require adversarial human-subject experiments before deployment.

VI. LIMITATIONS AND FUTURE WORK

All authentication metrics are derived from Monte Carlo simulation with synthetic user profiles and parametric sensor-noise models. No physical hardware prototype, human participants, or adversarial experiments contributed to the reported results. The simulation does not establish human behavioral consistency across sessions, environmental robustness, usability, replay/spoofing resistance, or actual ATmega328P execution time, memory usage, and power consumption. Sequential screening is evaluated as a decision mechanism, but computational savings and latency remain unmeasured.

Synthetic-user limitations: The 50 synthetic profiles cannot capture real inter-user and intra-user variability, speech characteristics, motor behavior, fatigue, learning effects, or intentional variation. The spatial generator also lacks person-specific structure, making its standalone discrimination uninformative about real hand-withdrawal behavior.

Resource implementation gaps: Storing all MFCC enrollment sequences (3 digits $\times$ 3 repetitions $\times$ 49 frames $\times$ 13 coefficients) requires approximately 5,733 coefficient values per user, or 11.2 KiB at 16-bit precision, exceeding the 1 KB EEPROM. Deployment would require template compression, quantization, or external storage, plus memory-bounded DTW and fixed-point arithmetic. Execution time, memory usage, and power consumption remain unmeasured on compiled firmware; the 582-byte banded-DTW value is analytical only. The $26.95 component cost prevented prototype procurement.

Environmental assumptions: The simulation assumes nominal low-noise indoor conditions; temperature, humidity, acoustic interference, sensor aging, wear, microphone placement, and hand occlusion are not modeled or experimentally evaluated.

Future work: The next step is implementing the architecture on the target MCU using fixed-point MFCC processing and banded DTW, followed by direct measurement of SRAM, flash, latency, and power. Template compression, quantization, or external non-volatile memory will address the EEPROM constraint. A multi-session human-subject study will evaluate authentication performance, behavioral drift, authentication time, and usability. Adversarial experiments will examine voice replay, behavioral imitation, mechanical reproduction, spatial spoofing, and environmental robustness. Thresholds will be recalibrated using physical measurements and human data. These experiments are required before assessing the architecture as a practical authentication system.

VII. CONCLUSION

This paper presented a low-cost three-layer behavioral authentication architecture designed for offline operation on resource-constrained embedded devices. The design combines spoken-digit voice verification, ultrasonic hand-withdrawal

TABLE V
CONTEXTUAL COMPARISON WITH EXISTING BEHAVIORAL AUTHENTICATION SYSTEMS

System	FAR	FRR	EER / AER	Evaluation	Platform / Context
SmartHandle [9]	1.39%	NR	NR	11 users	Physical-access behavioral authentication using inertial hand sensing.
Step & Turn [10]	0.01%	NR	NR	40 participants	Bimodal physical access using footstep and hand-movement signals.
PushPIN [6]	NR	NR	EER 0.51%	User studies	Smartwatch system using pressure-sensitive input and wrist motion.
HandLock [7]	0.82%	NR	NR	45 participants	Acoustic second factor using commodity smartphone audio hardware.
Our Work	**1.67%**	**4.00%**	**AER 2.83%**	**50 synthetic profiles**	**Offline design; 2KB SRAM target; simulation-evaluated.**

NR = Not Reported. Direct numerical comparison should be interpreted cautiously because the systems use different populations, datasets, sensing configurations, and evaluation protocols. In particular, the proposed result is simulation-based and should not be interpreted as experimentally demonstrated biometric performance.

TABLE VI
THREATS AND CURRENT TREATMENTS

Threat	Current Treatment
Shoulder surfing	Behavioral timing and force information are not represented by a simple PIN. Direct human testing is still required.
Voice replay	Not experimentally evaluated. Randomized challenge digits are a future mitigation.
Mechanical imitation	Multiple force, hold, and gap characteristics increase the observation burden, but resistance is unvalidated.
Spatial imitation	Distance sequence is included as an additional behavioral signal; spoofing resistance is not experimentally demonstrated.
Single-layer bypass	Under the selected weights, no individual layer can reach the final threshold by itself; for example, even $S_V = 100$ contributes only 25 points.
Brute-force guessing	A lockout policy can be implemented, but its usability and security impact remain to be experimentally evaluated.

sensing, and Force–Hold–Gap mechanical interaction using commodity sensors and an ATmega328P-class microcontroller.

In a held-out Monte Carlo simulation with 50 synthetic user profiles, including 20 user-disjoint profiles reserved for testing, the complete three-layer architecture achieved FAR = 1.67%, FRR = 4.00%, AER = 2.83%, and accuracy = 96.78%. The Voice+Mechanical configuration achieved AER = 3.58% with an estimated component cost of $25.45 when the HC-SR04 is omitted, compared with $26.95 for the full three-layer hardware design. Ablation analysis indicated that voice provided the strongest standalone discrimination under the synthetic simulation model, while the mechanical layer supplied complementary information in fusion. The spatial gap layer showed weak standalone performance because the synthetic generator did not model person-specific spatial structure; although adding it changed the simulated decision outcomes, this does not establish independent spatial biometric discrimination.

The contribution is therefore architectural and simulation-based: exploring multimodal behavioral fusion under a low-cost, memory-constrained, offline-operation design target. The reported error rates are feasibility estimates and are not experimentally validated biometric performance. Future work will implement the architecture on physical hardware and evaluate it with human participants across multiple sessions and adversarial conditions.

REFERENCES

[1] F. Tari, A. A. Ozok, and S. H. Holden, "A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords," in *Proc. 2nd Symp. Usable Privacy and Security (SOUPS)*, 2006, pp. 56–66.

[2] G. P. Hancke, "Practical eavesdropping and relay attacks on high-frequency RFID tokens," *J. Comput. Secur.*, vol. 19, no. 2, pp. 259–288, 2011.

[3] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino, "Impact of artificial gummy fingers on fingerprint systems," in *Proc. SPIE Optical Security and Counterfeit Deterrence Techniques IV*, vol. 4677, 2002, pp. 275–289.

[4] W. E. Eltahir, M. J. E. Salami, A. F. Ismail, and W. K. Lai, "Design and evaluation of a pressure-based typing biometric authentication system," *EURASIP J. Information Security*, vol. 2008, Art. no. 345047, 2008.

[5] K. Murao, T. Terada, A. Yano, and R. Matsukura, "Detecting smartphone users from the way of gripping," *Int. J. Pervasive Comput. Commun.*, vol. 11, no. 3, pp. 288–301, 2015.

[6] Y. Song and I. Oakley, "PushPIN: A pressure-based behavioral biometric authentication system for smartwatches," *Int. J. Human–Computer Interaction*, vol. 39, no. 4, pp. 893–909, 2023.

[7] S. Zhang and A. Das, "HandLock: Enabling 2-FA for smart home voice assistants using inaudible acoustic signal," in *Proc. Int. Symp. Research in Attacks, Intrusions and Defenses (RAID)*, 2021, pp. 251–265.

[8] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, 1998.

[9] S. Gupta, A. Buriro, and B. Crispo, "SmartHandle–A novel behavioral biometric-based authentication scheme for smart lock systems," in *Proc. 3rd Int. Conf. Biometric Engineering and Applications*, 2019, pp. 15–22.

[10] S. Gupta, M. Kacimi, and B. Crispo, "Step & Turn–A novel bimodal behavioral biometric-based user verification scheme for physical access control," *Computers & Security*, vol. 118, Art. no. 102722, 2022.

[11] A. Ross and A. K. Jain, "Information fusion in biometrics," *Pattern Recognit. Lett.*, vol. 24, no. 13, pp. 2115–2125, 2003.

[12] P. Tomar and R. C. Singh, "Cascade-based multimodal biometric recognition system with fingerprint and face," in *Proc. Int. Conf. Advances in Computing and Information Technology*, 2013, pp. 201–209.

[13] O. Höfener and Q. Wang, "Requirements framework for IoT device authentication using behavioral fingerprinting," *Computers & Security*, vol. 154, Art. no. 104459, July 2025, doi: 10.1016/j.cose.2025.104459.

[14] A. Krašovec, G. Baldini, and V. Pejović, "EBAT: Evaluation of Behavioural Authentication Toolbox," *SoftwareX*, vol. 32, Art. no. 102410, 2025, doi: 10.1016/j.softx.2025.102410.