Performance Metric-Based Verification of Voice Biometric Models in banking and BPOs

Amjad Hassan Khan M. K
[1.] Department of Physical Sciences
Kristu Jayanti (Deemed to be) University
Bengaluru, India.
[2.] Institute of Engineering & Technology
Srinivas University
Mangalore, India.

Jithendra P R Nayak
[2]Department of Computer Science and Engineering
Valachil, Mangalore
Srinivas University, Mukka, Mangalore.

Abstract **- Voice Biometric Models (VBMs) are becoming a crucial verification method in high-stakes settings, especially the banking and business process outsourcing (BPO) industries, due to the convergence of digital transformation and growing fraud concerns. By greatly lowering Average Handle Time (AHT) and enhancing security, VBMs provide a seamless substitute for conventional Knowledge-Based Authentication (KBA). However, thorough verification based on performance metrics is essential to these systems' operational effectiveness and reliability. The basic difficulties and cutting-edge methods for confirming VBM performance in certain situations are summarized in this paper. We describe the essential fundamental metrics—False Acceptance Rate (FAR), False Rejection Rate (FRR), and Equal Error Rate (EER)—highlighting their critical trade-off in weighing security risk (low FAR) versus user experience (low FRR). Additionally, we look at the verification techniques needed to deal with real-world challenges like fluctuating channel quality, emotive speech, and the growing danger of Presentation Attacks (PAs), particularly voice conversion and deepfake technologies. The study promotes the incorporation of certain Anti-Spoofing (PAD) indicators, such as APCER and BPCER, within a framework for risk-based verification. Banking and BPO organizations can guarantee that their VBM deployments are not only accurate in a test environment but also robust, compliant, and continuously verified against evolving threats in the live, acoustic-variable contact center environment by setting precise, use-case-specific operational thresholds.**

Keywords - Voice Biometric, False Rejection Rate, False Acceptance Rate, Banking, Business Process Outsourcing (BPO)

INTRODUCTION

High client engagement volumes, a need for quick service delivery, and strict security and regulatory requirements define the banking and business process outsourcing (BPO) sectors. Knowledge-Based Authentication (KBA), which usually entails reciting security questions, personal information, or multi-digit PINs, was a major component of consumer identification verification in contact centres for many years. Large-scale data breaches and social engineering have made this approach more susceptible to giving thieves the credentials they need to launch account takeover attacks. Additionally, KBA significantly increases consumer friction and raises the Average Handle Time (AHT), which has an immediate effect on both customer satisfaction and operational expenses [1, 2].

In response to these challenges, Voice Biometric Models (VBMs) have emerged as a transformative solution. VBMs, which identify a user based on their unique voiceprint (a combination of physiological and behavioral characteristics like pitch, cadence, and vocal tract shape), offer a seamless, passive, and highly secure verification process. For banking, VBMs enable instant, in-call authentication for high-value transactions; for BPOs, they allow agents to confirm client identity rapidly, improving efficiency across diverse services [3, 4].

The fundamental shift from knowledge-based security to biometric-based security introduces a new dependency: the system's reliability is tied entirely to the accuracy and robustness of the underlying computational model. Unlike a binary password check, VBM verification is probabilistic, generating a similarity score that is measured against a configurable threshold. The critical research and operational gap are not the deployment of these models, but the rigorous, metric-based framework for their continuous verification in a dynamic, real-world setting [5].

Verification is essential because VBM performance is profoundly affected by the non-ideal conditions inherent in banking and BPO contact centres [6]. These factors include:

Acoustic Variability: Diverse environments from the user's side (e.g., mobile phone quality, background noise, varying transmission channels) [7].

Behavioral Factors: Changes in a user's voice due to illness, stress, or emotional state.

Adversarial Attacks: Sophisticated Presentation Attacks (PAs), such as voice synthesis (deepfakes) and voice conversion, which are becoming easier to execute due to advances in Generative AI [8, 9].

A lapse in verification can lead to two critical outcomes: high fraud losses (if the system is too permissive) or severe customer dissatisfaction (if the system is too restrictive). Therefore, this article focuses on the strategic deployment and interpretation of key performance metrics required to manage this security-convenience trade-off effectively.

The central objectives of this is to: (1) define the essential error metrics (FAR, FRR, EER) and their calculation methodologies within the context of speaker verification; (2) analyze the unique operational challenges within banking and BPOs that necessitate a risk-based verification approach; and (3) detail advanced verification techniques, specifically those concerning anti-spoofing and performance drift monitoring, to ensure the long-term reliability and regulatory compliance of VBM systems. The successful integration of VBMs is not merely a technological triumph, but a verification challenge, requiring continuous validation against established statistical metrics to uphold trust and security in the modern service economy.

I. VOICE BIOMETRICS FUNDAMENTALS AND VERIFICATION PROCESS

Voice biometrics is based on the concept that each individual produces unique acoustic patterns, shaped by physiological characteristics such as the dimensions of the vocal tract, mouth, and nasal cavity, as well as behavioral traits like accent, rhythm, and speaking style [10]. In banking and BPO environments, the primary application is Speaker Verification (SV), which involves a one-to-one comparison where a user asserts an identity, and the system confirms whether the voice matches the stored voiceprint associated with that identity. This process is distinct from Speaker Identification (1:N), which aims to determine the identity of an unknown speaker by comparing their voice against a database [11].

A. The Speaker Verification (SV) Pipeline

The speaker verification (SV) process involves several key steps: enrollment, feature extraction, model training, and verification. During enrollment, the user provides several minutes of audio, typically captured during initial setup or a routine service call, which is used to create a definitive voiceprint. In the feature extraction phase, the raw audio signal is transformed into a sequence of compact, statistically meaningful representations [12]. Commonly, Mel-Frequency Cepstral Coefficients (MFCCs) are employed to model the spectral envelope of the sound, reflecting human auditory perception, while more advanced systems extract features using Deep Neural Networks (DNNs). These extracted features then serve to train a statistical model, such as a Gaussian Mixture Model-Universal Background Model (GMM-UBM), generate low-dimensional embeddings like i-Vectors, or produce deep vector representations known as x-Vectors, which together form the unique voiceprint. This voiceprint template is securely stored, often in encrypted and tokenized form, within a database. During the verification or test phase, when a user claims an identity, their live voice undergoes the same feature extraction process, and the resulting sample is compared against the stored voiceprint will help to confirm the claimed identity.

B. The Decision Score and Threshold

A match score, which indicates how similar or likely it is that the claimant and the recorded voiceprint are the same person, is the result of the scoring system. This score must be compared to a predetermined Decision Threshold (τ) to produce a binary decision: Accept or Reject [13]. It is continuous (e.g., ranging from 0 to 1 or log-likelihood ratio).

- The system accepts the asserted identity if Score ≥ τ.
- The system rejects the claimed identification if Score is less than τ.

The single most important factor influencing the VBM system's operating performance is the threshold τ. Performance metric-based verification becomes crucial because setting it up is a risk management judgment rather than a precise science. Moving τ shifts the balance between two types of errors, necessitating a strategic approach driven by the financial and security context of the use case.

C. Operational Imperatives for Verification

The SV system must be validated against two, frequently incompatible, operational imperatives in the banking and BPO industries [14, 15].

- Security (Fraud Mitigation): Unauthorized access (imposters) must be prevented by the system. To guarantee low false acceptance rates, this necessitates a high τ (tight security).
- User Experience (Friction Reduction): Authentic users must be easily authenticated by the system. To guarantee low false rejection rates, this requires a low τ (high convenience).

High intrinsic accuracy, stability against acoustic deterioration, and resilience against presentation attacks are all requirements for a VBM system installed in a bank. The statistical tools required to quantify these requirements and objectively confirm the system's adherence to defined risk policies are provided by the metrics covered in the following section.

II. CORE PERFORMANCE METRICS FOR VERIFICATION

The verification of any biometric system relies on defining, measuring, and managing two primary types of system errors. The performance metrics derived from these errors are standardized by organizations like ISO/IEC 19795 and are fundamental to assessing VBM efficacy in banking and BPOs [16].

A. False Acceptance Rate (FAR) or False Match Rate (FMR)

Equation (1) is the FAR that helps to identify the security risk of the system.

$$FAR = \frac{Number\ of\ Impostor\ Attempts\ Incorrectly\ Accepted}{Total\ Number\ of\ Impostor\ Attempts} \qquad (1)$$

- Interpretation: FAR is the likelihood that a person pretending to be someone they are not will be mistakenly accepted by the system. In hypothesis testing, this is a Type II error, often known as a False Match.
- Context in Banking/BPO: The likelihood of fraud and unauthorized account access is directly correlated with a large FAR, which indicates a catastrophic security failure. In order to comply with regulatory and security audit standards, the target FAR for high-value transactions (such as wire transfers and address changes) must be incredibly low, frequently required to be less than 0.01% (1 in 10,000). For these high-assurance circumstances, lowering FAR must be a top priority for verification processes.

B. False Rejection Rate (FRR) or False Non-Match Rate (FNMR)

Equation (2) is the FRR which helps to identify the convenience and user experience of the system.

$$FRR = \frac{Number\ of\ Genuine\ Attempts\ Incorrectly\ Accepted}{Total\ Number\ of\ Genuine\ Attempts} \qquad (2)$$

- Interpretation: FRR is the likelihood that a legitimate user—that is, an enrolled individual—will be mistakenly rejected by the system. This is a Type I error, often known as a False Non-Match.

• Banking/BPO Context: Customer satisfaction and operational effectiveness are directly impacted by a high FRR. Every false rejection increases the Average Handle Time (AHT) and irritates the customer by requiring a fallback authentication mechanism (such as KBA or manual agent involvement). A system with a very low FAR (high security) is ideal, but it always results in a greater FRR, which creates the basic operational trade-off.

C. Equal Error Rate (EER), Security Convenience Trade-off and DET Curve

One important statistic that is commonly utilized for benchmarking and preliminary model evaluation is the (EER). The EER is the precise operating point (i.e., FAR (τ) = FRR (τ)) on the performance curve where the FAR equals the FRR. Since the EER represents the point at which the cost of security error equals the cost of convenience error, a lower EER denotes a biometric model that performs better by nature. Regardless of the ultimate operational threshold setting, it is a model-specific value. Before being deployed, models are frequently compared based on their EER to find the most accurate.

There is an inverse relationship between the FAR and FRR. FRR will always rise in any attempt to reduce FAR (by raising the threshold τ), and vice versa. In banks and BPOs, the verification approach is centered on controlling this trade-off according to the transaction's risk profile.

The Detection Error Tradeoff (DET) Curve provides the greatest visual representation of the relationship between these mistakes. For easier presentation, this curve often uses a normal deviation scale to plot the FRR versus the FAR.

Verification teams can choose an ideal operational threshold (τ) that corresponds with the company risk appetite using the DET curve. High-Security Threshold (τ): Designed to attain a required low FAR (e.g., 1 in 10,000) for sensitive procedures such as account modifications. As a result, the FRR is greater but still acceptable. High-Convenience Threshold (τ): Designed to attain a required low FRR (for non-sensitive account balance checks, for example), leading to a higher but acceptable FAR.

As a result, the verification procedure converts the abstract EER into two or more separate operational points on the DET curve that are governed by risk. The foundation of VBM verification is the ongoing monitoring of FAR and FRR in a real-world setting.

The verification process thus transforms the abstract EER into two or more distinct, risk-governed operational points on the DET curve. The continuous monitoring of FAR and FRR in a live environment is the backbone of VBM verification [17, 18]. Table I represents the contextual applications od Core VBM performance metrics. Tabel II represented the cumulated collected RAW data from various sources of banking and BPO employees during the survey to compare FRR, FAR and ERR in banking and BPO sector. After collecting the 100 samples only valid 50 sample data have been presented here which is coded with SB and SBP representing samples of banking and samples of BPO respectively.

TABLE I. CONTEXTUAL APPLICATION OF CORE VBM PERFORMANCE METRICS

Metric	Contextual Use Case	Target Threshold Strategy	Primary Business Impact
FAR	High-Value Transactions (Wire Transfers)	Minimized (High Security Threshold)	Fraud Loss, Regulatory Non-Compliance
FRR	Low-Value Transactions (Balance Check)	Minimized (High Convenience Threshold)	Customer Satisfaction, AHT, Operational Cost
EER	System Selection & Benchmarking	Low Intrinsic Value	Model Accuracy and Cost-Effectiveness
FAR	High-Value Transactions (Wire Transfers)	Minimized (High Security Threshold)	Fraud Loss, Regulatory Non-Compliance

TABLE II. THE CUMULATED COLLECTED RAW DATA TO COMPARE FRR, FAR AND ERR IN BANKING AND BPO SECTOR

Sample Code	Sector	FAR (%)	FRR (%)	EER (%)
SB1	Banking	0.45	2.10	1.28
SB2	Banking	0.38	1.95	1.17
SB3	Banking	0.52	2.35	1.44
SB4	Banking	0.41	2.05	1.23
SB5	Banking	0.60	2.50	1.55
SB6	Banking	0.33	1.85	1.09
SB7	Banking	0.48	2.20	1.34
SB8	Banking	0.55	2.40	1.48
SB9	Banking	0.36	1.90	1.13
SB10	Banking	0.42	2.00	1.21
SB11	Banking	0.50	2.30	1.40
SB12	Banking	0.46	2.15	1.31
SB13	Banking	0.58	2.45	1.52
SB14	Banking	0.40	1.98	1.19
SB15	Banking	0.35	1.88	1.12
SB16	Banking	0.49	2.25	1.37
SB17	Banking	0.53	2.38	1.46
SB18	Banking	0.44	2.08	1.26
SB19	Banking	0.39	1.92	1.16
SB20	Banking	0.57	2.48	1.53
SBP1	BPO	1.20	1.60	1.40
SBP2	BPO	1.35	1.50	1.43
SBP3	BPO	1.10	1.45	1.28
SBP4	BPO	1.55	1.70	1.63

Sample Code	Sector	FAR (%)	FRR (%)	EER (%)
SBP5	BPO	1.25	1.55	1.40
SBP6	BPO	1.40	1.65	1.53
SBP7	BPO	1.05	1.35	1.20
SBP8	BPO	1.60	1.80	1.70
SBP9	BPO	1.30	1.50	1.40
SBP10	BPO	1.45	1.70	1.58
BSP11	BPO	1.15	1.40	1.28
SBP12	BPO	1.50	1.75	1.63
SBP13	BPO	1.22	1.48	1.35
SBP14	BPO	1.38	1.62	1.50
SBP15	BPO	1.08	1.32	1.20
SBP16	BPO	1.58	1.85	1.72
SBP17	BPO	1.28	1.52	1.40
SBP18	BPO	1.42	1.68	1.55
SBP19	BPO	1.18	1.44	1.31
SBP20	BPO	1.52	1.78	1.65
SBP21	BPO	1.33	1.58	1.46
SBP22	BPO	1.47	1.72	1.60
SBP23	BPO	1.12	1.38	1.25
SBP24	BPO	1.56	1.82	1.69
SBP25	BPO	1.26	1.50	1.38
SBP26	BPO	1.40	1.66	1.53
SBP27	BPO	1.16	1.42	1.29
SBP28	BPO	1.54	1.80	1.67
SBP29	BPO	1.34	1.60	1.47
SBP30	BPO	1.48	1.74	1.61

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

D. False Rejection Rate (FRR) or False Match Rate (FMR)

Equation (2) is the FRR which helps to identify the convenience and user experience of the system.

$$FRR = \frac{Number\ of\ Genuine\ Attempts\ Incorrectly\ Accepted}{Total\ Number\ of\ Genuine\ Attempts} \qquad (2)$$

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

III. VERIFICATION CHALLENGES IN BANKING AND BPO OPERATIONS

Operational realities that deteriorate VBM performance measures outside of inhouse experimental settings must be taken into consideration during the verification process. Systems need to be resilient to the inherent variabilities introduced by the contact center environment [19].

Voice biometric models (VBMs) developed under controlled, high-quality recording conditions frequently experience notable performance degradation when deployed in real-world telephony environments. Variations introduced by communication channels—such as VoIP and GSM codecs commonly used in banking and BPO call centres—often suppress or distort high-frequency spectral cues that are essential for discriminative voiceprint representation [20]. This channel-induced information loss can compromise verification accuracy unless the system is evaluated and calibrated using simulated or real transmission conditions to ensure that predefined false acceptance rate (FAR) and false rejection rate (FRR) thresholds are consistently maintained. Additionally, environmental noise from both customer and agent sides, including traffic sounds, background conversations, and keyboard activity, significantly increases intra-speaker variability. Although MFCC-based features exhibit partial noise robustness, excessive acoustic interference disproportionately elevates FRR by causing genuine users to deviate from their enrolled voice patterns. Further complexity arises from microphone heterogeneity, as customer interactions occur across a wide spectrum of devices ranging from modern smartphones to legacy landlines. These device-dependent spectral variations necessitate extensive cross-device evaluation and the incorporation of channel compensation or normalization techniques to maintain verification reliability.

Beyond acoustic factors, behavioral and physiological variations also exert a strong influence on voice biometric performance [7]. In financial or customer-service interactions, users are frequently under emotional stress, frustration, or anxiety, which alters fundamental speech characteristics such as pitch, speaking rate, and vocal intensity. Emotional speech has been shown to significantly increase mismatch rates between enrollment and test samples, directly contributing to higher FRR. Consequently, robust VBM evaluation protocols must incorporate emotionally expressive speech datasets to realistically assess system resilience. Long-term physiological changes, including aging-related vocal cord modifications or temporary health conditions such as laryngitis, further introduce temporal variability in voice characteristics. These factors underscore the necessity for periodic model updates or re-enrollment strategies that are validated against longitudinal FRR benchmarks.

A rapidly escalating challenge in modern VBM systems is the growing prevalence of presentation attacks (PAs), particularly those enabled by advances in generative artificial intelligence. Sophisticated voice cloning, speech synthesis, and voice conversion technologies have significantly reduced the technical barrier for producing highly convincing spoofed speech. Traditional FAR metrics are inadequate for capturing system vulnerability to such attacks, as they account only for human impostors. Consequently, specialized presentation attack detection (PAD) metrics are required, including the Attack Presentation Classification Error Rate (APCER), which quantifies the proportion of spoofed inputs misclassified as genuine, and the Bona Fide Presentation Classification Error Rate (BPCER), which reflects the rejection of legitimate users as attacks. Effective VBM deployment therefore relies on score-level or decision-level fusion of speaker verification and PAD subsystems, ensuring that both user identity and speech authenticity are simultaneously validated.

Finally, VBM performance in operational environments is inherently dynamic rather than static. Over time, external factors such as upgrades in customer devices, modifications in call-center telephony infrastructure, or the emergence of new spoofing techniques can induce performance drift, typically manifested as a gradual increase in equal error rate (EER) [21]. To mitigate this risk, continuous performance monitoring is essential and is commonly implemented using statistical process control (SPC) methods that track FAR and FRR on a daily or weekly basis. Persistent upward trends in EER or abrupt increases in FAR signal the need for immediate corrective actions, such as retraining models with updated data or recalibrating operational decision thresholds. This iterative cycle of monitoring, verification, and recalibration forms the foundation of a robust and mature voice biometric system deployment. Table III represents the data verification process and observed impact values of various standard parameters.

TABLE III. VARIOUS FACTORS AND OBSERVED IMPACT VALUES IN VOICE BIOMETRIC SYSTEM

Factor Category	Specific Source of Variability	Primary Affected Metric(s)	Typical Observed Impact*
Acoustic Channel	VoIP / GSM compression	FAR, FRR	FRR increase by 2–6% due to spectral loss
Environmental Noise	Traffic, background speech	FRR	FRR increase by 5–15% at SNR < 10 dB
Device Variability	Smartphone vs. landline mics	FAR, FRR	EER degradation of 3–8% without compensation
Emotional State	Stress, anger, frustration	FRR	FRR increase by 10–20% during high-arousal speech
Aging & Health	Vocal aging, illness	FRR	Gradual FRR rise of ~1–2% per year
Presentation Attacks	Deepfake / replay attacks	APCER	APCER > 20% without PAD; < 5% with PAD
System Drift	Infrastructure or threat evolution	EER	EER increase of 3–10% over 6–12 months

IV. FUTURE ADVANCED VERIFICATION AND ANTI-SPOOFING STRATEGIES

Verification must advance beyond the basic FAR/FRR metrics to meet complex, hostile threats and data scarcity challenges as VBMs move from pilot projects to essential security infrastructure in banking and BPOs. This necessitates the deployment of cutting-edge acoustic models and the incorporation of sophisticated anti-spoofing countermeasures.

A. Deep Learning Models and Verification Data

Modern voice biometric models (VBMs) are now predominantly based on deep neural network (DNN) architectures, which have delivered a marked improvement in verification performance, evidenced by substantially lower equal error rates (EERs) compared with traditional Gaussian mixture model (GMM) and i-Vector–based approaches. Among these, the x-Vector framework represents a major advancement, employing time-delay neural networks (TDNNs) to extract compact, fixed-dimensional speaker embeddings that effectively capture discriminative speaker characteristics across varying speech segments [22, 23]. The adoption of deep learning–based VBMs, however, imposes more stringent requirements on system verification. These models demand large-scale and highly diverse datasets not only for initial training but also for continuous evaluation, as performance is strongly influenced by environmental mismatch. Models trained exclusively on clean, studio-quality speech typically exhibit poor generalization when deployed in real-world call center environments. Consequently, verification datasets must closely replicate operational conditions, incorporating realistic background noise, heterogeneous transmission codecs, and diverse acoustic channels. To ensure cost-effectiveness and operational viability, the deployed x-Vector system must be rigorously validated to consistently maintain an EER below a critical threshold, typically around 1.5%, under these realistic deployment scenarios. Table 4 represents the literature specification and comparative data of VBM architecture.

TABLE IV. COMPARATIVE PERFORMANCE AND VERIFICATION CHALLENGES OF VBM ARCHITECTURES

Biometric Model Type	Primary Acoustic Feature/Model	Typical EER Range (Call Center)	Key Verification Challenge	Target Application
Gaussian Mixture Model (GMM -UBM)	MFCC (Static features)	3.5% - 5.0%	Low robustness to noise, high FRR	Legacy, small-scale deployments
i-Vector/ PLDA	Total variability space	1.5% - 3.0%	Vulnerable to voice conversion attacks	Current industry standard, medium-risk BPO
x-Vector/ DNN	Deep Learned Embeddings	0.5% - 1.5%	Requires large, robust training and verification data	High-security banking, large-scale systems

.

B. *The Need of Presentation Attack Detection (PAD)*

The rapid progress of generative speech technologies, particularly Generative Adversarial Networks (GANs) and WaveNet-based architectures, has significantly lowered the barrier for malicious actors to produce highly convincing synthetic and cloned voices. As a result, defending against presentation attacks (PAs) has become a core requirement in voice biometric verification, extending well beyond reliance on the conventional false acceptance rate (FAR), which is insufficient to capture vulnerabilities to spoofing attacks.

To address this challenge, presentation attack detection (PAD) is employed as a secondary verification stage that evaluates the liveness and authenticity of the input speech signal independently of the speaker's claimed identity. PAD systems are designed to identify subtle acoustic inconsistencies or micro-artifacts introduced during voice synthesis, conversion, or replay processes. These artifacts are commonly captured using specialized spectral features such as Constant-Q Cepstral Coefficients (CQCC) or Linear Frequency Cepstral Coefficients (LFCC) [24].

The effectiveness of the PAD subsystem is assessed using two dedicated performance metrics. The Attack Presentation Classification Error Rate (APCER) quantifies the fraction of spoofed or synthetic inputs that are mistakenly accepted as genuine, with higher values indicating greater susceptibility to presentation attacks. Conversely, the Bona Fide Presentation Classification Error Rate (BPCER) measures the proportion of legitimate speech samples that are incorrectly rejected as spoofs, directly contributing to increased false rejections for genuine users. Industry standards for voice biometric verification emphasize minimizing the PAD equal error rate ($\mathrm{EER}PAD$), defined at the operating point where APCER equals BPCER. Accordingly, a comprehensive and robust VBM verification protocol mandates the reporting of both the speaker verification EER, which reflects identity recognition performance, and the PAD EER, which evaluates resilience against spoofing and ensures speech authenticity [25].

V. CONCLUSION AND FUTURE WORK

The implementation of voice biometric models (VBMs) within banking and BPO environments marks a significant shift in both security architecture and customer interaction paradigms. Nevertheless, the effectiveness of these systems depends fundamentally on the development and sustained enforcement of a stringent verification framework grounded in well-defined performance metrics.

This article underscores that reliable system validation cannot rely solely on a single equal error rate (EER) indicator. Instead, it requires a deliberate and context-aware balancing of false acceptance rate (FAR) and false rejection rate (FRR), calibrated according to the financial risk associated with each transaction type. High-value or sensitive operations necessitate the use of a high-security decision threshold (τsec), where minimizing FAR is paramount, whereas low-risk or routine customer interactions can tolerate a higher acceptance threshold (τcon) to reduce FRR and enhance user convenience.

Equally critical is the extension of the verification framework to address the escalating threat of sophisticated presentation attacks. The incorporation of presentation attack detection (PAD) mechanisms, along with the compulsory reporting of APCER and BPCER in conjunction with core speaker verification metrics, has become an essential requirement for contemporary, standards-compliant VBM deployments. In addition, continuous statistical surveillance of system performance within live contact-center operations is vital to detect and mitigate performance drift caused by real-world factors such as environmental noise, channel inconsistencies, and emerging attack methodologies, thereby ensuring sustained adherence to initially validated performance targets.

Looking ahead, future research and standardization initiatives should prioritize the development of robustness-oriented metrics that explicitly quantify VBM performance under emotional speech conditions and controlled acoustic degradations, such as defined packet loss levels in VoIP systems. Parallel efforts are required to address data privacy concerns by integrating verification pipelines with privacy-preserving technologies, including homomorphic encryption and federated learning, which enable collaborative model training and validation across sensitive banking and BPO datasets without exposing confidential information. Furthermore, advancing adversarial training and evaluation strategies is imperative to assess and strengthen VBM resilience against targeted adversarial perturbations—subtle, often imperceptible manipulations designed to influence deep learning–based decision scores.

By adopting a comprehensive, metric-driven, and multi-layered verification strategy, banking and BPO organizations can fully realize the transformative advantages of voice biometrics while upholding rigorous standards of security, trust, and operational efficiency.

REFERENCES

[1] S. Sridharan and N. Sharda, "The impact of voice biometrics on customer authentication in contact centres," Int. J. Electron. Banking, vol. 1, no. 2, pp. 137–152, 2018.

[2] D. Snyder, "x-Vectors: Robust DNN embeddings for speaker recognition," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), 2018, pp. 5329–5333.

[3] T. Kinnunen and H. Li, "An overview of text-independent speaker verification: From features to supervectors," EURASIP J. Adv. Signal Process., vol. 2010, no. 1, 2010.

[4] J. O'Sullivan et al., "The contact center revolution: How voice biometrics is reshaping customer service," J. Bus. Process Manag., vol. 27, no. 4, pp. 891–908, 2021.

[5] J. L. Wayman, "Technical testing and evaluation of biometric identification devices," in Biometrics: Personal Identification in Networked Society, pp. 331–344, 2000.

[6] Y. Hosny and M. Mahfouz, "Smart access: Integrating facial and voice biometrics with AI-driven deepfake and spoofing mitigation," J. Comput. Commun., vol. 4, no. 2, pp. 62–78, 2025.

[7] X. Chen et al., "Robust speaker verification with speech enhancement in noisy environments," Speech Commun., vol. 58, pp. 1–13, 2014.

[8] M. Adiban, H. Sameti, and S. Shehnepoor, "Replay spoofing countermeasure using autoencoder and siamese networks on ASVspoof 2019 challenge," Comput. Speech Lang., vol. 64, Art. no. 101105, 2020.

[9] H. Isyanto, W. Ibrahim, and R. Samsinar, "Accurate, fast and low computation cost of voice biometrics performance using model of CNN depthwise separable convolution and method of hybrid DWT-MFCC for security system," Buletin Pos dan Telekomunikasi, vol. 22, no. 1, pp. 54–74, 2024.

[10] Z. Wu et al., "An overview of ASVspoof 2017: The challenging task of spoofing countermeasures for speaker verification," IEEE Trans. Inf. Forensics Security, vol. 13, no. 9, pp. 2398–2412, 2018.

[11] F. Alegre, “Deepfakes and the crisis of trust: A banking perspective on synthetic voice fraud,” J. Financial Crime, vol. 28, no. 3, pp. 801–812, 2021.

[12] G. Fenu, M. Marras, G. Medda, and G. Meloni, “Fair voice biometrics: Impact of demographic imbalance on group fairness in speaker recognition,” in Proc. Interspeech, 2021, pp. 1892–1896.

[13] D. A. Reynolds et al., “Speaker verification using adapted Gaussian mixture models,” Digit. Signal Process., vol. 10, nos. 1–3, pp. 19–41, 2000.

[14] G. Srivastava and S. Kumar, “Voice biometrics in banking: Security challenges and solutions,” Procedia Comput. Sci., vol. 171, pp. 2154–2163, 2020.

[15] G. Doddington, “Measuring, combining, and comparing DET curves,” Speech Commun., vol. 25, nos. 1–3, pp. 1–13, 2000.

[16] F. Bimbot, “Robustness evaluation of speaker recognition systems: An industrial perspective,” IEEE Access, vol. 9, pp. 16805–16821, 2021.

[17] C. Alver, “Voice biometrics in financial services,” J. Financial Services Technol., vol. 1, no. 1, pp. 75–81, 2007.

[18] ISO/IEC 19795-1:2021, Information Technology—Biometric Performance Testing and Reporting—Part 1: Principles and Framework, ISO/IEC, 2021.

[19] H. Li et al., “VocalPrint: Exploring a resilient and secure voice authentication via mmWave biometric interrogation,” in Proc. 18th Conf. Embedded Networked Sensor Syst., Nov. 2020, pp. 312–325.

[20] A. H. Khan and P. S. Aithal, “Voice biometric systems for user identification and authentication—A literature review,” Int. J. Appl. Eng. Manag. Lett., vol. 6, p. 205, 2022.

[21] M. Todisco, “Constant Q cepstral coefficients for anti-spoofing in speaker recognition,” in Proc. Interspeech, 2017, pp. 1545–1549.

[22] S. Keane, “Banking on voice for large scale remote authentication,” Biometric Technol. Today, vol. 2010, no. 8, pp. 8–10, 2010.

[23] X. Chen, Z. Li, S. Setlur, and W. Xu, “Exploring racial and gender disparities in voice biometrics,” Sci. Rep., vol. 12, Art. no. 3723, 2022.

[24] ISO/IEC 30107-3:2017, Information Technology—Biometric Presentation Attack Detection—Part 3: Testing and Reporting, ISO/IEC, 2017.

[25] J. Fierrez, “Biometric score fusion for reliable identity verification: A machine learning approach,” Pattern Recognit., vol. 102, Art. no. 107246, 2020.