Acoustic Cryptography: Personalized Microphone Authentication

Summary

Vocal biometrics is a great source of communication for humans. This form of communication establishes trust and can change the course of many situations in this world. One of the situations which is influenced and has caused substantial loss is the financial sector. With the progress of generative AI we have seen a threat in systems that rely on vocal biometrics. To mitigate these problems we can introduce a new layer of authentication before a transaction takes place, in this paper I propose a system to prevent exploitation of Gen AI.

Mike Card, a bank card sized hardware microphone that embeds analog signals to user specific key. By encrypting a user's sound we introduce a new authentication layer in the transaction process. This helps prevent AI deep fakes as it introduces a verification layer similar to our contemporary banking system.

Abstract

AI-powered deep fakes have caused a lot of financial damage, this technology can convincingly impersonate someone and bypass security to harness resources it should not have access to. In a typical transaction, three parties interact—who owns the funds, the bank (which holds them), and the one which receives payment—and current authentication factors (vocal biometrics and human trust) remain vulnerable to cloning, phishing, and replay attacks due to advances in generative AI technology. To address the gaps in security, we can introduce **MikeCard**, a card-like hardware which introduces an extra layer of authentication during transactions. This technology has a microphone exclusively encrypted to a user key which can be used to approve transactions. This idea is similar to our contemporary banking cards and is imagined to be similar in use. Only the factor of authentication changes, we use encrypted voice.

Why this works

Sound consists of longitudinal pressure waves propagating through a medium, generated by vibrating objects that change surrounding air.

A pure tone—an idealized sound containing a single frequency—can be modeled as a sinusoid:

$$s(t)=A\sin(2\pi f\, t+\phi)$$

where A is amplitude (loudness), f is frequency (pitch), t is time, and ϕ is phase (wave's starting point)

Most real-world sounds comprise multiple sinusoids summed together:

$$s_T(t) = \sum_{i=1}^{N} A_i \sin\left(2\pi f_i\, t + \phi_i\right),$$

1. Electromechanical Transduction

Dynamic microphones convert pressure waves into electrical signals via **Faraday's law of induction**, which states that a changing magnetic flux through a conductor induces an electromotive force (EMF).
In a dynamic mic, a diaphragm attached to a coil moves within a permanent magnetic field; diaphragm motion x(t) causes the coil to cut magnetic field lines, inducing a voltage proportional to the velocity of diaphragm displacement:

$$v_{\mathrm{emf}}(t) = -\frac{\mathrm{d}\Phi_B(t)}{\mathrm{d}t}$$

where ΦB is the magnetic flux.
This AC voltage mirrors the original sound wave's amplitude and frequency, producing an electrical waveform e(t) that faithfully represents the acoustic input.

2. Digital Sampling

To store or process sound digitally, continuous voltage e(t) is sampled at discrete intervals T, producing samples e[n]=e(nT). A standard analog-to-digital converter then quantized each sample to a finite precision B bits by rounding:

s[n]= round(e[n] x 2^B)

This pulse-code modulation (PCM) pipeline is universally used in audio devices, making s[n] the basis for all digital sound storage and analysis .

3. User-Specific ADC Encryption

(analog-to-digital conversion)

Rather than using a fixed PCM pipeline, **Mike Card** embeds a **secret, per-device transform** f_k into its ADC.
During quantization, each sample is computed as:

$$\mathbf{S[n] = f_k(e[n])}$$

where K is a 128-bit key stored in a tamper-resistant secure element on the card.
A lightweight instantiation of $f_k(e[n])$ is:

$$\mathbf{f_k(u) = \lfloor (u+a_k u^3) x 2^B \rfloor}$$

with α_k derived from **K** via a cryptographic key-derivation function. Without knowledge of α_k, no generic ADC or software decoder can invert f_k to recover the original signal.

Because only the genuine Mike Card implements f_k in hardware, any attempt to record the passphrase via a standard microphone or replay previously stored samples yields invalid s[n] streams that fail downstream voice-print matching.

4. Integration into Transaction Flow

By embedding this transform at the sensor level, Mike Card adds a **fourth authentication factor—something you have** (the card), **something you are** (your voice), and now at the **hardware level** (the secret ADC mapping).
In practice, during a transaction the customer simply speaks into the card; the client device reads the encrypted digital stream, then sends s[n] to the bank. Because the bank knows **K**, it applies $\mathbf{f_k^{-1}}$ to reconstruct the analog waveform, then runs conventional biometrics on the result.

A mismatch indicates either an illicit deepfake, replay, or microphone substitution attack, causing the transaction to be denied.