Facultad de Ciencias de la Ingeniería

Proyecto de Título – Anteproyecto

Sistema de autenticación biométrica por voz con detección de deepfakes y grabaciones mediante modelos de aprendizaje profundo

Nombre Estudiante: Tomás Fernando Poblete Chamorro
Rut Estudiante: 20.904.540-0
Año de Ingreso: 2020
Carrera: Ingeniería Civil Informática
Email Institucional: tomas.poblete@alumnos.ucm.cl
Teléfono: +569 7964 3904
Actividad Curricular: INF613 – Módulo Integrador de Formación Profesional
Académico/a: Ruben Hernández García

Índice de Contenidos

Introducción

En la actualidad, la transformación digital ha impulsado un crecimiento sostenido en los servicios en línea y las transacciones remotas, lo que ha incrementado la necesidad de implementar mecanismos de autenticación más seguros, confiables y accesibles (Restrepo Suárez, Clavijo López, y Castellanos,). Los métodos tradicionales, basados en contraseñas o códigos de verificación, presentan vulnerabilidades ampliamente conocidas, como el robo de credenciales, los ataques de *phishing* y la suplantación de identidad, lo que motiva la adopción de factores biométricos (Bailón Zambrano y Zambrano Montenegro,).

Entre las tecnologías biométricas, la autenticación por voz destaca por su carácter no intrusivo, facilidad de despliegue y adecuación para escenarios remotos (Harika, Dharmalingam, y Elangovan,). No obstante, el avance de la inteligencia artificial ha habilitado nuevas amenazas, en particular la generación de voces sintéticas y los ataques por reproducción (*replay*), que pueden comprometer sistemas de verificación vocal si no incorporan mecanismos anti-*spoofing* (Kamel, Sood, Dutta, y Aryal,).

Frente a este panorama, se propone el desarrollo de un sistema de autenticación biométrica por voz con detección de falsificaciones, integrable como servicio vía API. La solución combina modelos preentrenados para verificación de hablante y detección de *deepfakes*, con el objetivo de ofrecer una alternativa viable y escalable que responda a requerimientos actuales de seguridad en sectores de alta sensibilidad, como el bancario (Nosrati, Bidgoli, y Haj Seyyed Javadi,). Este enfoque se alinea con la tendencia a incorporar autenticación multimodal y a fortalecer controles anti-suplantación en flujos críticos de negocio (Anwar, Syed Ahmad, Kausar, Stević, y Gaba,).

En síntesis, este proyecto aporta a la integración de técnicas de aprendizaje profundo y detección de falsificaciones en sistemas biométricos de voz, abordando una necesidad real del entorno financiero: garantizar autenticación confiable sin aumentar la fricción de acceso a servicios digitales (Silva, Balamurugan, y Hakim,).

Desarrollo

1. Problemática

El incremento de los servicios digitales y las transacciones en línea ha intensificado la demanda de autenticación segura, superando las limitaciones de esquemas basados exclusivamente en contraseñas o códigos de un solo uso (Restrepo Suárez y cols.,). Estos métodos son vulnerables a ataques de ingeniería social y robo de credenciales, mientras que las soluciones biométricas de voz ofrecen una vía de mejora por su equilibrio entre usabilidad y seguridad (Khan y Aithal, ,).

Sin embargo, la autenticación por voz enfrenta amenazas emergentes derivadas de la IA generativa (p. ej., *voice deepfakes*) y ataques por reproducción (*replay*), capaces de falsificar identidades con alto realismo (Kamel y cols., ,). La literatura reciente subraya que, sin mecanismos de *liveness* y anti-*spoofing*, el riesgo de aceptación fraudulenta aumenta de forma significativa (Pianese, Cozzolino, Poggi, y Verdoliva, ,).

Esta problemática es especialmente crítica en banca y finanzas, donde la verificación remota debe equilibrar seguridad, experiencia de usuario y costos operativos (Amjad Hassan Khan y Aithal, ,). En consecuencia, se identifican tres focos principales:

1. Falta de detección robusta de falsificaciones de voz (grabaciones y *deepfakes*) en soluciones de producción (Kamel y cols., ,).

2. Ausencia de *liveness* y verificación dinámica del contenido leído (texto aleatorio) para mitigar *replay* (Pianese y cols.,).

3. Escasez de soluciones modulares, escalables e integrables vía API que combinen verificación de hablante con anti-*spoofing* en tiempo casi real (Silva y cols., ,).

2. Cliente y/o Público objetivo

El proyecto se desarrolla con el apoyo del **Banco Ripley**, específicamente con el área de **Autenticación Digital**, perteneciente al **Value Stream de Riesgo y Fraude**. Esta área ha manifestado interés en la exploración de nuevos métodos de inicio de sesión que complementen los mecanismos actuales de seguridad y fortalezcan la *dimensionalidad de la autenticación*, es decir, la combinación de múltiples factores y tecnologías que permitan verificar de manera más robusta la identidad de los usuarios. En este contexto, la autenticación biométrica por voz se presenta como una alternativa innovadora y no intrusiva que podría integrarse como un factor adicional dentro de sus procesos de validación digital.

El interés del Banco Ripley radica en evaluar la factibilidad técnica y la aplicabilidad práctica de este tipo de autenticación en entornos bancarios, donde la detección temprana de intentos de fraude y suplantación de identidad resulta crítica. La posibilidad de incorporar un factor biométrico vocal permitiría reforzar la seguridad sin comprometer la experiencia de usuario, al ofrecer un método natural, rápido y de bajo costo operativo.

De manera más general, el proyecto está orientado al **sector financiero**, abarcando a instituciones que gestionan procesos de autenticación y verificación de identidad digital para el acceso a sus servicios en línea. Estas entidades buscan soluciones que mitiguen los riesgos asociados a

las amenazas emergentes, como la suplantación de identidad mediante grabaciones o la generación de voces sintéticas (*deepfakes*).

Dentro de estas organizaciones, el trabajo se enfoca particularmente en las áreas de **Autenticación Digital**, **Seguridad de la Información** y **Experiencia de Usuario (UX)**, las cuales enfrentan el desafío de equilibrar la seguridad con la usabilidad y la eficiencia operativa. Dichas áreas son responsables de adoptar tecnologías que reduzcan la fricción en los procesos de inicio de sesión y aumenten la confianza del usuario final en los servicios digitales.

El **público objetivo** del sistema corresponde a los usuarios finales de plataformas digitales que acceden regularmente a servicios bancarios, financieros o de atención al cliente de manera remota. Estos usuarios requieren mecanismos de autenticación seguros, rápidos y confiables, que no dependan exclusivamente de contraseñas o dispositivos físicos, sino que aprovechen rasgos biométricos personales como la voz, ofreciendo una experiencia más natural e intuitiva.

Además, el sistema propuesto posee potencial de aplicación en otros sectores donde la identificación remota confiable es esencial, tales como entidades gubernamentales, servicios de salud, educación en línea y plataformas de atención ciudadana, ampliando su alcance más allá del entorno bancario.

3. Carta de Compromiso

UNIVERSIDAD CATOLICA DEL MAULE

Talca 10-10-2025

Yo Félix Mauricio Troncoso Arcos, Technical Manager del equipo de Autenticación Digital en la empresa Banco Ripley declaro que la empresa es del rubro Bancario enfocado al **sector consumo / retail financiero,** y valido la necesidad de desarrollo del proyecto del estudiante Tomás Poblete Chamorro, comprometiendo mi participación, respaldo y compromiso durante el desarrollo del proyecto de título, asociado a la actividad curricular: **INF-613 – Módulo Integrador de formación profesional**, desarrollado en el semestre en curso.

Validando que esta solución mejorará las próblemáticas anteriormente descritas.

Nombre y firma los estudiantes Tomas Poblete Chamorro	Nombre y Firma por parte de la Empresa Félix Troncoso Arcos

4. Propuesta de Solución

La propuesta consiste en el desarrollo de un sistema de autenticación biométrica por voz, expuesto mediante una API para su integración con aplicaciones cliente. El sistema permitirá: (i) registrar la voz del usuario leyendo frases breves para crear su firma de voz; (ii) verificar la identidad del usuario comparando una nueva grabación con su firma almacenada; y (iii) detectar intentos de suplantación debidos a grabaciones reproducidas o voces sintéticas generadas por inteligencia artificial (*deepfakes*).

La solución prioriza la viabilidad en tiempo de tesis: emplea modelos preentrenados de reconocimiento de hablantes y de detección de falsificaciones, y realiza un ajuste fino ligero con bases de datos públicas y un conjunto reducido de muestras locales, con el fin de mejorar el desempeño en español y en el entorno de uso.

4.1. Aspectos técnicos

Flujo funcional

- **Registro de voz (enrolamiento):** el usuario lee 4–6 frases cortas propuestas por el sistema. Cada audio se valida en calidad (duración mínima de voz útil, relación señal-ruido, ausencia de saturación). A partir de estas muestras se obtiene una firma de voz (vector numérico representativo) y se almacena su promedio y estadísticas asociadas.
- **Desafío de texto para el inicio de sesión:** al autenticarse, el sistema entrega una frase aleatoria que el usuario debe leer. Esto dificulta el uso de una grabación previa.
- **Verificación de identidad por voz:** se extrae la firma de voz de la nueva grabación y se compara la similitud con la firma almacenada; si supera un umbral calibrado, el sistema considera que corresponde a la misma persona.
- **Validación de la frase leída (verificación del texto):** se utiliza un sistema de reconocimiento automático del habla de tamaño reducido para comprobar que el contenido de la grabación coincide con la frase solicitada.
- **Detección de falsificaciones (grabaciones y voces sintéticas):** un modelo especializado analiza la señal para identificar indicios de audio reproducido o sintetizado. Si el puntaje de "probable falsificación" supera un umbral definido, la autenticación se rechaza.

Modelos preentrenados seleccionados

- **Reconocimiento de hablantes (firma de voz):** se utilizarán **ECAPA-TDNN** y **x-vector**, ampliamente empleados para embeddings de identidad vocal; disponibles en `SpeechBrain` y `pyannote.audio`.
- **Detección de falsificaciones: AASIST**, **RawNet2** y variantes **ResNet** anti-spoofing.
- **Verificación de la frase:** un modelo ASR ligero para validar el contenido leído.

Calibración de umbrales y ajuste ligero

Se calibrarán los umbrales de decisión (similitud mínima de identidad, puntaje máximo permitido de "audio falso" y coincidencia textual de la frase) utilizando conjuntos de datos públicos y muestras recolectadas localmente en idioma español. Para el ajuste fino del detector de falsificaciones se emplearán las bases de datos **VoxCeleb** y **ASVspoof 2019/2021**, con el objetivo de mejorar la sensibilidad del modelo al español y al entorno de grabación previsto para la demostración.

Métricas de evaluación

- **EER (Equal Error Rate):** punto donde FAR y FRR son iguales.
- **FAR (False Acceptance Rate):** intentos no autorizados aceptados.
- **FRR (False Rejection Rate):** intentos legítimos rechazados.
- **Latencia de inferencia:** tiempo promedio por autenticación.

4.2. Aspectos tecnológicos

- **Lenguaje y entorno:** Python.
- **API:** `FastAPI` con endpoints para registro, desafío y verificación.
- **Procesamiento de audio:** lectura, normalización, recorte de silencios y validación de calidad.
- **Modelos:** integración con `PyTorch`; exportación opcional a formatos optimizados.
- **Almacenamiento:** `PostgreSQL` + extensión vectorial para similitud; cifrado opcional de audios.
- **Despliegue:** contenedores `Docker`; HTTPS.
- **Privacidad y cumplimiento:** consentimiento, cifrado en tránsito y reposo, minimización de datos y derecho al olvido.

4.3. Aspectos de arquitectura

La arquitectura sigue un esquema **cliente–servidor** y contempla tres componentes principales:

Aplicación cliente (demostración):

- Captura la voz del usuario, muestra la frase a leer y envía el audio a la API.
- Presenta el resultado y, en caso de rechazo, la causa principal (baja similitud, frase incorrecta o audio sospechoso).

API biométrica (núcleo):

- Preprocesa el audio.

- Valida el texto leído frente al desafío.
- Genera y compara la firma de voz.
- Evalúa probabilidad de falsificación.
- Fusiona decisiones con umbrales calibrados.
- Registra auditoría (puntajes, latencia, versiones).

Capa de datos y seguridad:

- Base de datos para usuarios, firmas, desafíos y auditoría.
- Almacenamiento cifrado de audios cuando aplique; opción de almacenar solo firmas.
- Políticas de retención y eliminación acordes a normativa.

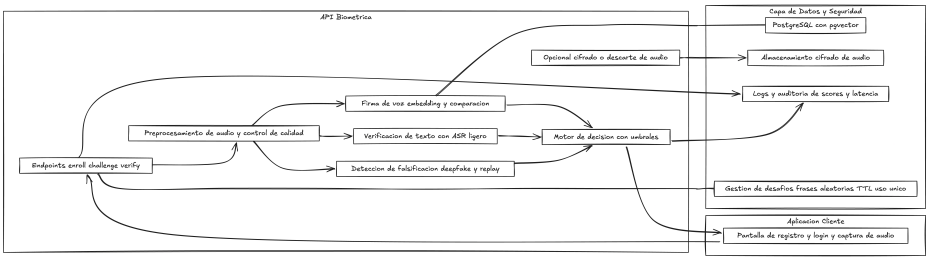

Figura 1: Diagrama de Arquitectura.

Resultado esperado

Con esta solución, el sistema permitirá autenticar por voz de forma práctica y confiable, rechazando intentos de suplantación basados en audios reproducidos o voces generadas por inteligencia artificial, y validando que el usuario leyó la frase solicitada. El uso de modelos preentrenados y el ajuste fino ligero garantizan viabilidad en tiempos de tesis, a la vez que mantienen rigurosidad técnica y evaluaciones cuantitativas.

5. Objetivos

5.1. Objetivo General

Desarrollar un sistema de autenticación biométrica por voz, expuesto mediante una API, que permita verificar la identidad de los usuarios de manera segura y eficiente, incorporando mecanismos de detección de falsificaciones de audio (*deepfakes* y grabaciones reproducidas), con el fin de fortalecer los procesos de autenticación digital en entornos financieros y de alta sensibilidad.

5.2. Objetivos Específicos

- Implementar un sistema de autenticación por voz que integre el registro de usuarios, la generación de firmas vocales, el desafío de texto y la detección de falsificaciones.
- Diseñar una arquitectura modular basada en `FastAPI` que permita la comunicación entre los módulos de procesamiento, verificación y almacenamiento de datos.
- Evaluar el desempeño del sistema mediante métricas especializadas (*EER*, *FAR*, *FRR*, *t-DCF*) considerando precisión, latencia y robustez frente a ruido.
- Documentar el proceso de desarrollo, configuración y validación del sistema, estableciendo lineamientos para su mejora e integración futura en entornos reales.

6. Alcance del Proyecto

El presente proyecto contempla el diseño e implementación de un sistema de autenticación biométrica por voz, orientado a su uso en entornos de prueba o demostración funcional, con proyección hacia su integración en sistemas de autenticación reales. El alcance se centra en el desarrollo del **núcleo funcional** del sistema, considerando los siguientes componentes principales:

- **Módulo de registro (enrolamiento):** permite capturar y procesar muestras de voz de un usuario para generar y almacenar su firma de voz.
- **Módulo de autenticación:** gestiona la verificación a partir de una nueva grabación, incluyendo la validación del texto leído y la comparación de similitud vocal.
- **Mecanismo de detección de falsificaciones:** implementa un modelo capaz de identificar intentos de suplantación mediante grabaciones o voces sintéticas.
- **Interfaz API:** endpoints REST para registro, desafío de texto y verificación.
- **Base de datos y almacenamiento:** estructura relacional y vectorial; buenas prácticas de seguridad y privacidad.
- **Evaluación experimental:** pruebas con un conjunto controlado de usuarios y audios; tasas de error, latencia y robustez.

El proyecto no incluye la puesta en producción del sistema ni la integración directa con infraestructuras bancarias reales. Sin embargo, se dejarán establecidas las bases técnicas y arquitectónicas necesarias para su escalamiento y adaptación futura en entornos empresariales o institucionales.

En términos de tiempo, el desarrollo se circunscribe al período académico correspondiente al módulo integrador, considerando las etapas de análisis, diseño, implementación, calibración y validación experimental. El resultado esperado es un **prototipo funcional** que demuestre la viabilidad técnica y la efectividad de la autenticación biométrica por voz con detección de suplantaciones en idioma español.

7. Planificación Inicial

La planificación del proyecto se ha estructurado considerando un semestre académico comprendido entre el 1 de septiembre y el 20 de diciembre de 2025, abarcando desde la etapa de análisis hasta la validación final del sistema. A continuación se presenta la planificación propuesta:

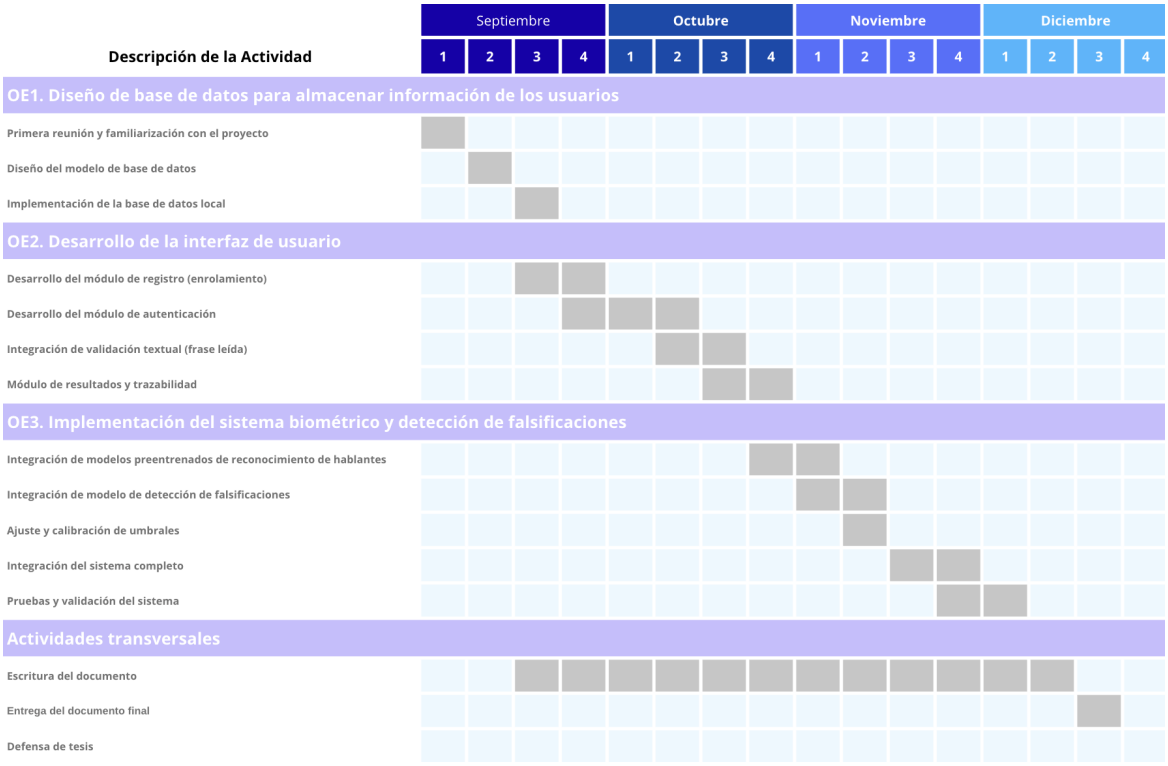

Figura 2: Carta Gantt de Planificación

Referencias

Amjad Hassan Khan, M. K., Aithal, P. S. (2024, abril). Abcd analysis of voice biometric system in banking. *International Journal of Management, Technology and Social Sciences (IJMTS)*, *9*(2), 1–17. Descargado de https://supublication.com/index.php/ijmts/article/view/1186 doi: 10.47992/IJMTS.2581.6012.0342

Anwar, N., Syed Ahmad, S. S., Kausar, N., Stević, , Gaba, Y. U. (2025, septiembre). Multiple biometric authentication for online banking system based on multiple fuzzy approach. *Scientific Reports*, *15*, 1–20. doi: 10.1038/s41598-025-13571-6

Bailón Zambrano, G. S., Zambrano Montenegro, D. F. (2024, septiembre). Revisión sistemática de la literatura sobre métodos de autenticación biométrica en aplicaciones móviles. *Revista Científica Multidisciplinar G-nerando*, *5*(2), 1233–1245. Descargado de https://revista.gnerando.org/revista/index.php/RCMG/article/view/313 doi: 10.60100/rcmg.v5i2.313

Harika, L., Dharmalingam, V., Elangovan, P. (2023, diciembre). Voice authentication system. En (pp. 1–6). doi: 10.1109/ICDSAAI59313.2023.10452482

Hernández Nava, C. A., Rincón García, E. A., Lara Velázquez, P., de los Cobos Silva, S. G., Gutiérrez Andrade, M. , Martínez Licona, F. M., … Montes Orozco, E. (2024, diciembre). El peligro de la suplantación de la identidad por medio de audio. *Contactos, Revista de Educación en Ciencias e Ingeniería*(137), 43–52. Descargado de https://contactos.izt.uam.mx/index.php/contactos/article/view/443

Kamel, K., Sood, K., Dutta, H. S., Aryal, S. (2025). *A survey of threats against voice authentication and anti-spoofing systems*. Descargado de https://arxiv.org/abs/2508.16843

Khan, A. M., Aithal, S. (2022, abril). Voice biometric systems for user identification and authentication – a literature review. *International Journal of Applied Engineering and Management Letters*, 198–209. doi: 10.47992/IJAEML.2581.7000.0131

Nosrati, L., Bidgoli, A. M., Haj Seyyed Javadi, H. (2024, noviembre). Machine learning and metaheuristic algorithms for voice-based authentication: A mobile banking case study. *International Journal of Computational Intelligence Systems*, *17*. doi: 10.1007/s44196-024-00690-7

Pianese, A., Cozzolino, D., Poggi, G., Verdoliva, L. (2024, junio). Training-free deepfake voice recognition by leveraging large-scale pre-trained models. En (pp. 289–294). doi: 10.1145/3658664.3659662

Restrepo Suárez, J., Clavijo López, F., Castellanos, J. G. (2025). Biometría en el contexto de la ciberseguridad, retos empresariales. *En Contexto*, *13*(23), 163–182. Descargado de https://doi.org/10.53995/23463279.1631 doi: 10.53995/23463279.1631

Silva, F., Balamurugan, B., Hakim, J. (2025, junio). An architecture for voice-based authentication and authorization with deepfake detection. *European Conference on Cyber Warfare and Security*, *24*, 425–435. doi: 10.34190/eccws.24.1.3567