

# Verificación del Hablante Mediante Dispositivos Móviles en Entornos Ruidosos

Graciela Etchart, Silvia Ruiz, Ernesto Miranda, Juan José Aguirre, Mauro Herlein, Carlos Alvez,

Facultad de Ciencias de la Administración - Universidad Nacional de Entre Ríos

Av. Tavella 1424, Concordia, Entre Ríos - CP 3200

getchart@fcad.uner.edu.ar, sruiz@fcad.uner.edu.ar, emiranda@fcad.uner.edu.ar, juaagu@fcad.uner.edu.ar, herlein.mauro@gmail.com, caralv@fcad.uner.edu.ar

## Resumen

El interés por las aplicaciones biométricas ha crecido considerablemente en los últimos años debido a que han surgido nuevos servicios que necesitan formas confiables para autenticar a sus usuarios. En este trabajo se abordará la biometría de voz, a través de dispositivos móviles, para la verificación de la identidad del hablante. Actualmente, los dispositivos móviles proporcionan una plataforma conveniente para la captura de voz y su transferencia para la verificación remota. Sin embargo, debido a limitaciones de almacenamiento y de capacidad computacional de estos dispositivos se requiere un adecuado conjunto de parámetros para llevar a cabo la autenticación. Además, por las características del entorno donde se realizan la mayoría de las transacciones de verificación, es necesario trabajar en el fortalecimiento de sistemas de reconocimiento cuando la voz es capturada en entornos ruidosos. El objetivo general de este trabajo es realizar un análisis comparativo de diferentes enfoques para contrarrestar desajustes entre la fase de entrenamiento y prueba en la verificación del hablante a través de dispositivos móviles, analizando su adecuación principalmente en entornos ruidosos.

**Palabras clave:** biometría de voz, verificación, dispositivos móviles, seguridad.

## Contexto

Este trabajo se da en el marco del Proyecto PID 07/G044 “*Gestión de datos biométricos en base de datos objeto - relacionales*”, que da continuidad al Proyecto PID 07/G035 “*Identificación de personas mediante Sistemas Biométricos. Estudio de factibilidad y su implementación en organismos estatales*” [1][2][3].

El mismo constituye un futuro trabajo de tesis de la Maestría en Sistemas de Información (MSI) dictada en la Facultad de Ciencias de la Administración perteneciente a la Universidad Nacional de Entre Ríos (UNER). El objetivo general es realizar un análisis comparativo de diferentes enfoques para contrarrestar desajustes entre la fase de entrenamiento y prueba en la verificación del hablante a través de dispositivos móviles, analizando su adecuación principalmente en entornos ruidosos.

## Introducción

El interés por las aplicaciones biométricas ha crecido considerablemente en los

últimos años debido a que han surgido nuevos servicios que necesitan formas confiables para autenticar a sus usuarios. Actualmente, los usuarios deben recordar decenas de contraseñas y números de identificación personal, debido a que resulta poco seguro utilizar la misma contraseña para autenticar su identidad en cada servicio al que se pretende acceder. Además, las contraseñas pueden ser olvidadas o robadas por diferentes medios. La biometría ofrece mecanismos para resolver estos problemas.

En este trabajo se abordará la biometría de voz para la verificación de la identidad del hablante.

La verificación mediante el rasgo biométrico de la voz no es un nuevo campo de investigación. Los primeros intentos de construir sistemas de verificación del hablante se hicieron a principios de la década de 1950 [4, 5].

En biometría, verificar un hablante consiste en autenticar una identidad reclamada mediante el análisis de una muestra de voz del reclamante. Es un problema de detección binaria donde el reclamante debe ser clasificado como el verdadero o como un impostor. De esta manera, pueden ocurrir dos tipos de error: el falso rechazo de un usuario genuino o la falsa aceptación de un impostor [5].

En los servicios de acceso remoto, como ser a través de dispositivos móviles, la verificación de identidad utilizando la voz del usuario presenta numerosas ventajas, debido a la seguridad que ofrece la biometría con respecto a métodos tradicionales de autenticación.

Además, el número de teléfonos móviles ha superado el número de personas en el mundo. Los usuarios se muestran más dependientes de sus teléfonos inteligentes. Su mayor uso se debe a las aplicaciones móviles en creciente aumento. Varias aplicaciones requieren la verificación del usuario para acceder a los servicios.

En este contexto, los dispositivos móviles proporcionan una plataforma conveniente para la captura de voz y la transferencia a distancia para la verificación remota. Sin embargo, debido a limitaciones de almacenamiento y de capacidad computacional de estos dispositivos se requiere un adecuado conjunto de parámetros para llevar a cabo el proceso de autenticación. Además, por las características del entorno donde se realizan la mayoría de las transacciones de verificación, es necesario trabajar en el fortalecimiento de sistemas de reconocimiento cuando la voz es capturada en entornos ruidosos. Estos aspectos serán tratados en este trabajo.

La investigación continua en este campo ha estado en curso durante los últimos veinticinco años [4, 5, 6, 7]. Estudios recientes han encontrado que el desajuste de entrenamiento (enrolamiento) y prueba (verificación) afectan significativamente el desempeño de la verificación del hablante [8, 9, 10].

Las aplicaciones in situ de verificación del hablante, como control de acceso, control de fronteras, domótica, pueden ofrecer mucha libertad en la elección del equipo de captura de datos, el diseño de interfaces de usuario, medios para controlar el comportamiento del usuario y el entorno de fondo. Esto facilita la recolección de muestras de voz de alta calidad bajo condiciones controladas para permitir una verificación exacta del hablante. Sin embargo, con los usuarios físicamente presentes existen otras modalidades biométricas que podrían utilizarse, como huella digital, iris y rostro [11].

Por otro lado, con las aplicaciones basadas en transmisiones remotas, como el caso de los dispositivos móviles, la verificación del hablante presenta ventajas sobre otras modalidades biométricas. Debido a que en cada dispositivo se dispone de un micrófono y un canal de transmisión de voz, mientras que el uso de

otros rasgos biométricos puede requerir instrumentos de captura de datos no disponibles o de costosa incorporación. Sin embargo, existe poca posibilidad de influir en los instrumentos telefónicos que se utilizan en los sistemas de verificación de hablantes, lo que da lugar a diferentes características de señal de entrada entre los usuarios y posiblemente incluso entre capturas de un mismo usuario. También se tienen poca posibilidad de controlar a los usuarios o el entorno acústico. Por lo tanto, en comparación con las aplicaciones in situ, las aplicaciones remotas de la verificación de hablantes generalmente necesitan contrarrestar una gama más amplia de variabilidad entre sesiones que no esté directamente relacionada con las características de voz de los hablantes.

## Líneas de Investigación y Desarrollo

Una de las líneas más importantes, consideradas en este trabajo, es la biometría de voz para la verificación de la identidad del hablante, utilizada en diversas aplicaciones o sistemas relacionados con la seguridad.

Entre los principales desafíos que debe enfrentar la verificación del hablante en muchas situaciones prácticas se encuentran [7, 12]:

- Deben desarrollarse métodos eficaces para combatir el desajuste de entrenamiento y prueba que está presente debido a las condiciones acústicas adversas que reducen la precisión de la verificación del hablante.
- Debe reducirse la cantidad de datos necesarios para diseñar sistemas de verificación de hablantes de última generación, ya que es difícil recopilar y registrar gran cantidad de datos de voz.
- Debe adecuarse la duración del discurso requerido para entrenar

modelos de hablantes durante la etapa de entrenamiento así como la duración de los enunciados de prueba o test.

La compensación del canal es un enfoque que se utiliza para reducir la falta de coincidencia entre el entrenamiento y las pruebas. Esta compensación se produce en los diferentes niveles, como el dominio de la característica, el dominio del modelo y el dominio de la puntuación. En el dominio de la característica, la supresión adaptable del ruido, la substracción media cepstral (CMS), el filtrado de RASTA y el warping de la característica se utilizan para compensar la variabilidad del canal. Mientras que para compensar la falta de concordancia en el dominio del modelo se utilizan los enfoques JFA, JFASVM e i-vector. Por otra parte, en el dominio de la puntuación, para compensar la variabilidad de la sesión se utilizan los métodos de normalización, tales como normalización de prueba (norma T), simétrica (norma S) y de prueba cero (norma ZT)[13].

## Objetivos

El objetivo general de este trabajo es realizar un análisis comparativo de diferentes enfoques para contrarrestar desajustes entre la fase de entrenamiento y prueba en la verificación del hablante a través de dispositivos móviles, analizando su adecuación principalmente en entornos ruidosos.

Con este propósito se explorarán diferentes modelos y métodos estudiando el impacto de las duraciones de entrenamiento y prueba, el efecto del ruido de fondo, la longitud de la sesión y los canales en el sistema de verificación del hablante.

Para llevar a cabo las pruebas experimentales, se realizará una investigación exploratoria de diferentes herramientas open source utilizadas para el reconocimiento biométrico de personas

mediante voz. Las pruebas se efectuarán sobre una base de datos propia.

## Formación de Recursos Humanos

La estructura del equipo de investigación del PID 07/G035 está formada por el Director, Co-directora y cuatro integrantes docentes. El Director del proyecto dirige la tesis de MSI de la Facultad de Ciencias de la Administración de la UNER de la Co-directora y tres integrantes, que se encuentran realizando sus trabajos de investigación en el área del proyecto.

Además, el proyecto cuenta con un becario de Iniciación en la Investigación y un integrante alumno de la Licenciatura en Sistemas, cuyas tareas están relacionadas con la captura, registración y almacenamiento de datos biométricos.

## Referencias

1. Etchart G., Luna L., Leal C., Benedetto M., Alvez C., "Sistemas de reconocimiento biométricos, importancia del uso de estándares en entes estatales". CGIV - XIII Workshop de Investigadores en Ciencias de la Computación (WICC 2011), 5 y 6 de Mayo de 2011. Universidad Nacional de Rosario. Rosario – Argentina. pp 339-343.
2. Etchart G., Alvez C., Benedetto M., "Gestión de Datos Biométricos en Bases de Datos Objeto-Relacionales". XV Workshop de Investigadores en Ciencias de la Computación (WICC 2013). Paraná Entre Ríos. pp. 97-101.
3. Ruiz S., Etchart G., Alvez C., Miranda E., Benedetto M., Aguirre J., "Representación e interoperabilidad de imágenes biométricas". XVII Workshop de Investigadores en Ciencias de la Computación (WICC 2015). Salta.
4. Furui S., "50 years of progress in speech and speaker recognition research", ECTI Transaction on Computer and Information Technology, Vol. 1, 2005.
5. Campbell J., "Speaker recognition: A tutorial", Proceedings of the IEEE, Vol. 85, N° 9, 1997.
6. Furui S., "An overview of speaker recognition technology", Kluwer international series in Engineering and Computer science, pp. 31-56, 1996.
7. Naufal A., Phaklen E., Badlishah A., Naseer S., "Speaker Recognition System: Vulnerable and Challenges", International Journal of Engineering and Technology (IJET), Vol. 5, N° 4, 2013.
8. Chen Y., Hong Q., Chen X., Zhang C., "Real-time speaker verification based on GMM-UBM for PDA" 5° IEEE International Symposium on Embedded Computing, 2008.
9. Kinnunen T., Li H., "An overview of text-independent speaker recognition: From features to supervectors", Speech Communication, Vol. 52, N° 1, pp. 12-40, 2010.
10. Ding P., He L., Yan X., Zhao R., Hao J., "Robust technologies towards automatic speech recognition in car noise environments", 8° International Conference on Signal Processing, Vol. 1, 2006.
11. Beranek B., "Voice biometrics: success stories, success factors and what's next", Biometric Technology Today, pp. 9-11, 2013.
12. Ming J., Hazen T. J., Glass J. R., Reynolds D. A., "Robust Speaker Recognition in Noisy Conditions", IEEE Transactions on Audio, Speech and Language Processing, Vol. 15, N° 5, pp. 1711-1723, 2007.
13. Kanagasundaram A., "Speaker Verification using I-vector Features" (Tesis doctoral). Queensland University of Technology, Speech and Audio Research Laboratory, Science and Engineering Faculty, 2014.