

1 INTRODUCCIÓN

1.1 Motivación

En la sociedad actual el uso de técnicas de reconocimiento biométrico va adquiriendo cada vez más importancia debido a la necesidad creciente de contar con herramientas poderosas que incrementen la seguridad en la identificación de personas. La implantación de las tecnologías de la información en nuestra cultura y el hecho de que actualmente prácticamente todo el mundo dispone de Internet han afectado a la práctica totalidad de los quehaceres cotidianos. Hoy en día se puede llevar a cabo una transacción bancaria o realizar una compra sin movernos de casa, enviar correos de forma instantánea o mantener una conferencia a una distancia de miles de kilómetros en tiempo real.

Sin embargo, esta libertad de movimientos y consecuente mejora en el confort arrastra inevitablemente consigo nuevas formas de delincuencia y estafa. Es por ello que han proliferado los sistemas de seguridad basados en reconocimiento biométrico. Tradicionalmente, los sistemas basados en clave eran relativamente fáciles de asaltar con las técnicas de descifrado existentes. Pero un sistema de seguridad basado en un parámetro biométrico es mucho más difícil de interceptar con éxito. Esto justifica el hecho de que progresivamente se han ido imponiendo este tipo de sistemas para garantizar la seguridad en transacciones económicas, intercambios de información o bien para tener acceso a información reservada o bienes preciados.

El reconocimiento de voz es una de las tecnologías biométricas más populares gracias al desarrollo de la telefonía y de las redes informáticas, y muy aceptada socialmente ya que se trata de una técnica no invasiva que cuenta con numerosas aplicaciones y cuyo proceso de adquisición no supone grandes costes. Entre las muchas aplicaciones que tiene el reconocimiento de voz, podríamos destacar la banca telefónica, donde con una aplicación de reconocimiento de locutor se garantiza muchísima seguridad, ya que un supuesto impostor no podrá realizar ningún tipo de transacción sin la voz del dueño de las cuentas. También en entornos domóticos o para discapacitados, se puede conseguir dar instrucciones a un sistema automático para que realice una determinada función con tan sólo pronunciar la instrucción. Y por supuesto en entornos de seguridad o control de acceso la voz es un sistema muy cómodo y natural de identificación.

El reconocimiento de voz se divide en dos áreas: reconocimiento de habla y reconocimiento de locutor. Mientras que al primero le concierne la extracción del mensaje lingüístico de una locución, el segundo trata sobre la extracción de la identidad de la persona que pronuncia la locución. El eje de este proyecto se cierce sobre la segunda opción, y en concreto, sobre el campo o área de la Verificación Automática de Locutor.

1.2 Objetivo y metodología

El presente proyecto abordará las diferentes técnicas de Verificación Automática de Locutor que se han estudiado y analizado en la última década hasta aproximadamente nuestros días. El objetivo fundamental consistirá en comparar los resultados obtenidos con las diferentes técnicas para una misma base de datos proveniente del Instituto Nacional de Tecnología de Estados Unidos (NIST). Para la implementación, el departamento de Tratamiento Digital de Voz e Imagen de la Universidad de Swansea me ha permitido hacer uso del software ALIZE, utilizado por algunas universidades que participan en las evaluaciones del NIST en temas de reconocimiento de voz. Con este proyecto se pretende por tanto comprobar y describir cómo ha ido mejorando progresivamente la eficiencia de los sistemas de Verificación Automática de Locutor en los últimos años, lo cual nos permitirá realizar un recorrido cronológico hasta alcanzar el estado del arte en este campo y al mismo tiempo proporcionar una base sólida para entender e investigar futuras técnicas.

Los dos objetivos principales que se persiguen con este proyecto son los siguientes:

- Por una parte, desarrollar un análisis teórico, completo y detallado sobre los cuatro sistemas más importantes utilizados en el área de la Verificación Automática de Locutor en los últimos diez años. Ello nos permitirá entender la estructura básica de cada una de las técnicas y conocer con total detalle los algoritmos estadísticos empleados para el entrenamiento y el test.
- Implementar las diferentes técnicas en ALIZE, que nos permitirá llevar a cabo todas las simulaciones con la base de datos perteneciente al NIST, con el fin de poder comparar las diferencias en la eficiencia de cada uno los sistemas y captar las ventajas y desventajas proporcionadas por cada uno de ellos.

Para lograr dichos objetivos se ha seguido la siguiente metodología y plan de trabajo, organizada en diferentes fases:

- **Documentación.** Antes de comenzar a trabajar con los diferentes sistemas de Verificación Automática de Locutor es necesario formarse al respecto por lo que es útil estudiar la bibliografía básica (publicaciones científicas y libros) sobre el estado del arte actual en esta técnica de reconocimiento de locutor. Además, será necesario documentarse con la base de datos utilizada en este proyecto (NIST).
- **Estudio de ALIZE.** Un porcentaje importante de la elaboración del proyecto se ha dedicado al entendimiento del software empleado para poder ejecutar las simulaciones. En concreto, se han estudiado aquellas funciones que incorpora ALIZE que sean necesarias para las pruebas experimentales.

- **Evaluación de los resultados y elaboración de la memoria.** Se ha llevado a cabo un análisis de los resultados obtenidos en las pruebas experimentales así como una comparativa entre las diferentes técnicas utilizadas según los resultados arrojados por las mismas. Estos análisis, junto con la revisión del estado del arte y un estudio completo del proyecto llevado a cabo, servirán para elaborar la memoria que pone punto y final al proyecto fin de carrera.

1.3 Organización de la memoria

La presente memoria se encuentra estructurada de la siguiente forma:

El proyecto comienza en el capítulo 2 con una explicación general y superficial sobre los aspectos generales de un sistema de Verificación Automática de Locutor.

Ya en el capítulo 3 se explica y describe con total profundidad la extracción de características, etapa fundamental en cualquier sistema de reconocimiento de voz. Nunca se trabaja directamente con la señal de voz digitalizada, sino que se lleva a cabo un determinado procesamiento que nos permite obtener unos parámetros característicos de una grabación de voz. En este capítulo por tanto se explicarán las diferentes fases necesarias para extraer dichas características.

En el capítulo 4 se analiza con total detalle las cuatro técnicas que se han analizado e implementado en este proyecto fin de carrera, haciendo especial hincapié en la primera pues se considera como base o punto de partida de las otras tres.

Durante el capítulo 5 se describen las distintas técnicas de normalización de puntuaciones que se han tratado en este proyecto.

La base de datos empleada para la realización de los experimentos y los protocolos de evaluación utilizados para llevar a cabo los mismos se ha explicado durante el capítulo 6. También se describe el software empleado al igual que las principales funciones o paquetes de los que se han hecho uso para poder ejecutar las pruebas experimentales.

En el capítulo 7 se presentan y analizan los resultados obtenidos en todas las simulaciones, donde se podrá percibir la diferencia en la eficiencia de las distintas técnicas que nos permitirá resaltar ciertos detalles de especial importancia.

Finalmente, en el último capítulo se extraerán una serie de conclusiones y directrices relativas a los resultados de este proyecto. También se analizarán posibles líneas de trabajo futuro para la construcción de nuevos sistemas más precisos haciendo uso de la aportación realizada con este proyecto.