

UNIVERSIDAD DE SALAMANCA

Escuela Politécnica Superior de Zamora

Ingeniería Técnica Informática de Gestión

Departamento de Informática y Automática, Área de Lenguajes y Sistemas Informáticos



RESUMEN DEL PROYECTO FINAL DE CARRERA

SERAL: Sistema Experimental de Reconocimiento Automático de Locutor

Autor:

Manuel Montero Villar

Tutor:

D. Hernando Silva Varela

Departamento de Informática y Automática

Área de Lenguajes y Sistemas Informáticos

Cotutores:

D. José Luis Pérez Iglesias

Departamento de Informática y Automática

Área de Lenguajes y Sistemas Informáticos

D^a. María Luisa Pérez Delgado

Departamento de Informática y Automática

Área de Lenguajes y Sistemas Informáticos

Adjudicación: Octubre 2006

Presentación: Junio 2007

1. Introducción

El reconocimiento automático de locutor es un sistema de Identificación Biométrica Personal, denominación que reciben aquellas técnicas para identificar alguna de las características intrínsecas de la persona (como por ejemplo huellas dactilares, iris o voz) y que supone únicas en cada persona.

Un sistema de reconocimiento automático de locutor trata de llevar a cabo una identificación de la persona usando la voz como medio de identificación, presentando la gran ventaja de que esta se puede realizar a distancia (Ej. por teléfono), pudiéndose aplicar a sistemas control de acceso o de identificación personal.

Los sistemas de reconocimiento automático de locutor funcionan en dos fases bien diferenciadas:

- Fase de entrenamiento: En esta fase los usuarios del sistema introducen muestras de voz en el mismo, este extrae sus características y construye los patrones de voz de los usuarios.
- Fase de prueba (o reconocimiento): En esta fase un usuario introduce una muestra de voz en el sistema, este extrae las características, construye un patrón de voz y lo compara con el que generó en la fase de entrenamiento.

Dentro del reconocimiento automático de locutor se pueden encontrar a su vez dos disciplinas claramente diferenciadas como son la verificación automática de locutor y la identificación automática de locutor.

La verificación automática de locutor consiste en decidir si una muestra de voz desconocida pertenece o no a un locutor que proporciona su identidad al sistema.

Un sistema de verificación automática de locutor puede producir dos resultados:

- Aceptación: Se decide que la muestra de voz pertenece al locutor. La persona es quien dice ser.
- Rechazo: Se decide que la muestra de voz no pertenece al locutor. La persona no es quien dice ser.

La identificación automática de locutor consiste en decidir a que locutor de los reconocidos por el sistema pertenece una muestra de voz desconocida.

Un sistema de identificación automática de locutor puede producir dos resultados:

- Identificación: Se indica a que usuario pertenece la muestra de voz introducida en el sistema.
- No identificación: Se indica que la muestra de voz introducida en el sistema no pertenece a ningún usuario reconocido.

Un aspecto muy importante en los sistemas de reconocimiento automático de locutor es si el reconocimiento del locutor se hace a partir de una palabra o frase establecida de antemano o si se hace a partir de cualquier palabra o frase elegida por el locutor de manera unilateral.

Aquellos sistemas en los que hay que introducir siempre la misma frase son denominados Dependientes del Texto, por el contrario en aquellos que se puede elegir la frase a introducir libremente son denominados Independientes del Texto.

La dependencia del texto es muy empleada en verificación automática de locutor y muy raramente en identificación automática de locutor donde se usa más la independencia del texto.

Los sistemas Dependientes del Texto presentan mejores resultados y requieren menos tiempo en las fases de entrenamiento y prueba que los Independientes del Texto, ya que el vocabulario es conocido, sin embargo también son más susceptibles al fraude.

2. Objetivos

Los objetivos del proyecto, se pueden clasificar en dos grupos, por una parte están los objetivos definidos en la especificación de requisitos del software, por otra parte, están los objetivos técnicos planteados por el alumno.

OBJETIVOS DEL SOFTWARE

Los objetivos que se han marcado durante la especificación de requisitos son los siguientes:

- Obtener un software que pueda ser utilizado para diseñar y experimentar con sistemas de Reconocimiento Automático del Locutor (RAL).
- Implementar procedimientos para el alta, baja, verificación y reconocimiento de locutores a partir de señales de voz provenientes de ficheros o del micrófono.
- Diseñar y evaluar sistemas de RAL con tasas bajas de FA y FR que puedan implementarse posteriormente en sistemas de control de acceso o ser aplicados en criminalística.
- Conseguir un software cuya utilización sea sencilla, a pesar de la dificultad de las tareas que este pueda realizar, ya que puede estar destinado a la utilización por usuarios no expertos, cuyo nivel de conocimiento en la materia puede ser bastante bajo.
- Permitir ver las estadísticas de la utilización del software y de las operaciones fundamentales asociadas al RAL con el fin de conocer, por ejemplo, las veces que se ha utilizado, el número de aceptaciones y rechazos, etc.
- Implementar funcionalidades tales como la grabación de voz, visualización de señales de voz y su espectrograma asociado, selección, reproducción y etiquetado de segmentos de la señal de voz, localización automática de bordes.

OBJETIVOS TÉCNICOS

Para el desarrollo del proyecto, se han aplicado muchos de los conocimientos adquiridos durante los tres cursos de carrera, sin embargo, en alguna de las ocasiones, no han sido suficientes, por lo que se ha desarrollado un trabajo de investigación en campos como el procesado de señales digitales, nuevos lenguajes de programación y nuevos entornos de desarrollo.

Los objetivos técnicos que son necesarios para la realización del proyecto son los siguientes:

- Profundizar en el conocimiento del sistema del cuerpo humano que se usa para la generación de la voz.
- Estudiar los parámetros y procedimientos mediante los cuales se modela el sistema humano que genera la voz.
- Conocer y aplicar las técnicas mediante las cuales se crean patrones o modelos de referencia que permiten representar a un locutor en particular.
- Analizar los conocimientos necesarios para implementar un sistema de RAL así como la toma de decisiones que se llevan a cabo.
- Estudiar y aplicar herramientas o lenguajes específicos para el procesamiento y modelado de señales de voz que permitan abordar los objetivos de manera realista.

3. Desarrollo

Durante el desarrollo del proyecto, se ha empleado un ciclo de vida que consta de las siguientes cuatro fases cíclicas:

Análisis -> Diseño -> Implementación -> Pruebas

A las mencionas fases, hay que añadir, la investigación que se realizó previamente al análisis, así como el mantenimiento del sistema que se realiza una vez que este supera las pruebas a las que es sometido.

Como primer paso del desarrollo, se lleva a cabo un estudio de viabilidad económica, con el fin de determinar la posible rentabilidad que conlleva el desarrollo del proyecto, también se realiza una planificación temporal, para determinar los tiempos asignados a cada fase del desarrollo.

Las primeras decisiones importantes se toman durante el análisis de requisitos, donde se decide sobre el tipo de técnica que se empleará para la tarea de reconocimiento automático de locutor y sobre las modalidades de reconocimiento que se implementarán en el sistema.

Como técnica para el reconocimiento automático de locutor, se decide implantar las redes neuronales artificiales, y como modalidades de reconocimiento, se implementan tanto la verificación automática de locutor, como la identificación automática de locutor.

Para el diseño se emplean diagramas de flujo de datos, con los que se define las conexiones entre los distintos módulos que conforman el sistema y un diccionario de datos, para describir los datos empleados.

Antes de iniciar la fase de implementación, se decidió el lenguaje de desarrollo, optando por utilizar MATLAB por las siguientes razones:

- Está disponible para muchas plataformas.
- Implementa diversas operaciones para el procesado de señales.
- Implementa diversas topologías de redes neuronales artificiales.
- Implementa diversas operaciones sobre las redes neuronales artificiales.

La implementación ha sido desarrollada en dos fases diferentes.

Una fase en la que se implementan todos los módulos relacionados con la interfaz de usuario, tales como los módulos de zoom, de reproducción de señales de voz o de grabación.

En otra fase diferente, se implementan los módulos relacionados con la lógica necesaria para la tarea de reconocimiento automático de locutor, como son los módulos de creación de redes neuronales artificiales, de entrenamiento o de simulación de las redes neuronales artificiales.

Las pruebas a las que fue sometido el sistema se pueden agrupar en dos categorías.

Por una parte, las pruebas unitarias, aplicadas a cada módulo por separado, para comprobar el correcto funcionamiento de los mismos.

Por otra parte, las pruebas de integración, que se realizan una vez que han concluido las anteriores. Estas pruebas se aplican a todo el sistema en conjunto, para verificar que el comportamiento del sistema es el deseado.

La documentación se ha ido desarrollando de manera paralela a la realización del sistema, documentando todas y cada una de las fases del desarrollo.

La documentación ha sido desarrollada basándose en la guía propuesta por la Universidad de Salamanca en el documento “Proyecto de final de carrera en la Ingeniería Técnica en Informática, Guía de realización y documentación”.

Este desarrollo se ha conseguido repitiendo un número finito de veces las cuatro fases del ciclo de vida expuestas al principio de esta sección.

4. Conclusiones

Este proyecto es el resultado de 9 meses de trabajo, durante los cuales, ha sido indispensable la colaboración de varias personas.

Durante la realización del proyecto, se han podido aplicar los conocimientos adquiridos durante los tres cursos de la carrera, y además se han adquirido otros nuevos, que se han orientado hacia la realización de un proyecto de tamaño medio.

Con este proyecto, se ha conseguido una formación sobre el entorno de desarrollo de MATLAB, así como sobre procesamiento de señales digitales y redes neuronales artificiales.

Este proyecto ha permitido ampliar los conocimientos sobre las técnicas utilizadas para tareas de reconocimiento automático de locutor, a la vez que se ha conseguido un software que implementa una de ellas.

5. Fuentes de información

La mayor parte de la información utilizada para el desarrollo del presente proyecto, se ha obtenido de artículos publicados por diversas universidades e instituciones de estandarización europeas y estadounidenses, presentes en Internet.

El trabajo publicado en el artículo “Mejoras en un Sistema de Reconocimiento de Locutor Basado en RNA Mediante entrenamiento con Normalización de Canal” de la Universidad de Valladolid, determinó la arquitectura a implantar en la red neuronal artificial.

El artículo “An Improved Endpoint Detector for Isolated Word Recognition”, publicado por la IEEE, supuso la base en la que se apoya todo el procesado de señales que se implementa con este proyecto.

El documento “Proyecto de final de carrera en la Ingeniería Técnica en Informática, Guía de realización y documentación”, definió la estructura que se ha seguido en la documentación.

Tutoriales on-line sobre MATLAB, facilitaron la labor de desarrollo.

La cooperación del profesorado ha supuesto una ayuda en la resolución de dudas, así como en la toma de decisiones en todas las fases del proyecto.