

UNIVERSIDAD DE SALAMANCA

Escuela Politécnica Superior de Zamora

Ingeniería Técnica Informática de Gestión

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



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

Agradecimientos

A la Universidad de Salamanca, que además de acogerme como estudiante, puso a mi disposición los medios necesarios para la realización de este proyecto.

A D. Hernando Silva Varela, por haber accedido a dirigir el presente proyecto de fin de carrera.

A D. José Luis Pérez Iglesias y D^a. María Luisa Pérez Delgado, por participar como cotutores del presente proyecto.

A D. Jaime Calvo Gallego, por la ayuda prestada tanto en la selección del entorno de desarrollo, como en su utilización.

Al Departamento de Informática y Automática de la Universidad de Salamanca y a sus profesores por su ayuda prestada.

A todas las personas que participaron en la fase de pruebas del sistema desarrollado en el proyecto.

A todas las demás personas que participaron de alguna manera en la realización de este proyecto, haciéndolo posible.

A todos los profesores que me dieron clase en alguno de los cursos estudiados en la Escuela Politécnica Superior de Zamora.

A todas las personas que como evaluadores o miembros de un tribunal analizarán este proyecto.

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

D. Hernando Silva Varela, Profesor del Departamento de Informática y Automática (Área de Lenguajes y Sistemas Informáticos) de la Universidad de Salamanca.

HAGO CONSTAR: que D. Manuel Montero Villar, ha realizado bajo mi dirección el Proyecto de Fin de Carrera que lleva por título “SERAL: Sistema Experimental de Reconocimiento Automático de Locutor”, con el fin de optar al título de Ingeniero Técnico en Informática de Gestión.

Y para que surta los efectos oportunos, firmo el presente en Zamora, a 25 de Junio de 2007.

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

Índice

Lista de tablas	5
Lista de figuras	7
Prólogo	11
1. Introducción	13
1.1 El Reconocimiento Automático de Locutor (RAL)	14
1.1.1 Verificación e Identificación del locutor	15
1.1.2 Dependencia e Independencia del Texto	18
1.2 Contenido y Estructura de la memoria	19
2. Objetivos del proyecto	20
2.1 Objetivos especificados por los requisitos	20
2.2 Objetivos técnicos.	20
3. Conceptos teóricos.	21
3.1 Características de la voz	21
3.2 Coeficientes de Predicción Lineal (CPL).	23
3.3 Energía por bandas de frecuencia (EBF)	25
3.4 Información transicional (parámetros Delta)	26
3.5 Redes Neuronales Artificiales (RNA)	27
3.5.1 Topología de una RNA	29
3.5.2 Aprendizaje en las RNA	30
4. Técnicas y herramientas	31
4.1 Técnicas	31
4.1.1 Ciclo de vida en espiral	31
4.2 Herramientas	33
4.2.1 MatLab.	33
4.2.2 Image Processing Toolbox.	34
4.2.3 Signal Processing Toolbox	34
4.2.4 Neural Network Toolbox.	34
4.2.5 HTM2CHM	34
5. Aspectos relevantes del desarrollo del proyecto.	35
5.1 Aspectos relevantes del análisis	35
5.2 Aspectos relevantes del diseño.	35
5.3 Aspectos relevantes de la implementación.	38
6. Trabajos relacionados.	39
7. Conclusiones y líneas de trabajo futuras	40
7.1 Conclusiones.	40
7.2 Líneas de trabajo futuras.	40

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

Anexo 1: Plan del proyecto software.	42
A1.1 Planificación temporal del proyecto	49
A1.1.1 Tabla de precedencias.	49
A1.1.2 Diseño del Grafo PERT	50
A1.1.3 Calculo de los tiempos	51
A1.1.4 Calculo de los caminos críticos	52
A1.2 Estudio de viabilidad del proyecto	53
A1.2.1 Determinación de gastos y beneficios	53
A1.2.2 Determinación de la rentabilidad	55
A1.2.3 Conclusiones	57
 Anexo 2: Especificación de requisitos del software	60
Lista de cambios	67
A2.1 Introducción	68
A2.2 Lista de usuarios participantes	68
A2.3 Catálogo de requisitos del sistema	69
A2.4 Especificación de los requisitos	70
A2.4.1 Requisitos funcionales	70
A2.4.2 Requisitos no funcionales	76
A2.4.3 Requisitos de información	78
A2.5 Interfaces de usuario	80
A2.5.1 Menú Archivo.	81
A2.5.2 Menú Operaciones	82
A2.5.3 Menú Estadística	83
A2.5.4 Menú Ayuda	83
A2.6 Pruebas	84
 Anexo 3: Especificación de Diseño	87
A3.1 Introducción.	94
A3.2 Ámbito del software	95
A3.3 Diseño de datos	96
A3.4 Diseño arquitectónico	101
A3.5 Diseño de la interfaz	120
A3.6 Diseño procedimental	135
A3.7 Referencia cruzada a los requisitos	143
A3.8 Pruebas	146
A3.9 Entorno tecnológico del sistema	148
A3.10 Plan de desarrollo e implantación	149
 Anexo 4: Documentación Técnica de Programación.	153
A4.1 Introducción.	160
A4.2 Documentación de las bibliotecas	161
A4.3 Código fuente	165
A4.4 Manual del programador	184
A4.5 Pruebas unitarias	185

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

Anexo 5: Manuales de Usuario	189
A5.1 Documento de instalación y configuración	196
A5.2 Manual de usuario	199
A5.2.1 Configurar el sistema.	199
A5.2.2 Abrir un fichero de voz.	200
A5.2.3 Reproducir un intervalo	201
A5.2.4 Grabar un fichero de voz.	202
A5.2.5 Zoom.	203
A5.2.5.1 Acercar.	203
A5.2.5.2 Alejar.	204
A5.2.5.3 Intervalo.	205
A5.2.5.4 Vista inicial	206
A5.2.6 Ver estadísticas.	207
A5.2.7 Borrar estadísticas.	208
A5.2.8 Entrenamiento.	209
A5.2.9 Borrar locutor.	211
A5.2.10 Verificación	212
A5.2.10.1 Por fichero.	212
A5.2.10.2 Por micrófono.	214
A5.2.11 Identificación.	216
A5.2.11.1 Por fichero	216
A5.2.11.2 Por micrófono.	218
A5.2.12 Detección de inicio y fin de palabras.	220
A5.2.13 Etiquetado	221
A5.2.14 Alternancia del espectrograma.	222
A5.2.15 Ayuda.	223
 Acrónimos.	 226
Glosario.	228
Bibliografía.	230
Índice alfabético	233

Lista de tablas

Tabla 3.1. Características de la voz	22
Tabla A1.1. Tabla de precedencias de las tareas del proyecto.	49
Tabla A1.2. Resumen de los gastos e ingresos del proyecto	54
Tabla A2.1. Lista de cambios.	67
Tabla A2.2. RF-1: Entrenamiento del sistema.	70
Tabla A2.3. RF-2: Verificación de locutor	71
Tabla A2.4. RF-3: Identificación de locutor	71
Tabla A2.5. RF-4: Eliminación de locutor	72
Tabla A2.6. RF-5: Ver estadísticas	72
Tabla A2.7. RF-6: Borrar estadísticas	73
Tabla A2.8. RF-7: Grabar voz	73
Tabla A2.9. RF-8: Reproducir voz	74
Tabla A2.10. RF-9: Visualizar señal de voz	74
Tabla A2.11. RF-10: Cambiar la configuración	75
Tabla A2.12. RNF-1: Bajo consumo de recursos	76
Tabla A2.13. RNF-2: Interfaz de usuario basada en ventanas.	76
Tabla A2.14. RNF-3: Existencia de menú	76
Tabla A2.15. RNF-4: Sistema operativo	77
Tabla A2.16. RNF-5: Documentación.	77
Tabla A2.17. RI-1: Fichero de estadísticas	78
Tabla A2.18. RI-2: Fichero de configuración	78
Tabla A2.19. RI-3: Fichero de patrones de voz.	78
Tabla A2.20. RI-4: Fichero de etiquetas	79
Tabla A3.1 controles de la ventana principal de SERAL	122
Tabla A3.2 controles del cuadro de diálogo de salida	123
Tabla A3.3 controles del cuadro de diálogo de abrir fichero	124
Tabla A3.4 controles del cuadro de diálogo de mostrar estadísticas	125
Tabla A3.5 controles del cuadro de diálogo de borrar estadísticas	126
Tabla A3.6 controles del cuadro de diálogo de ayuda	127
Tabla A3.7 controles del cuadro de diálogo de reconocimiento.	128
Tabla A3.8 controles del cuadro de diálogo de mostrar resultados.	129
Tabla A3.9 controles del cuadro de diálogo de nombrar locutor	130
Tabla A3.10 controles del cuadro de diálogo de inicio de grabación.	131
Tabla A3.11 controles del cuadro de diálogo de etiquetar	132
Tabla A3.12 controles del cuadro de diálogo de ver etiquetas	133
Tabla A3.13 controles del cuadro de diálogo de fichero de etiquetas.	134
Tabla A3.14 Referencia cruzada a los requisitos	144
Tabla A4.1. función endpoint	165
Tabla A4.2. función parametros	166
Tabla A4.3. función reproducir.	167
Tabla A4.4. función reproducirinterval	168
Tabla A4.5. función zoommas	169
Tabla A4.6. función zoommenos	170
Tabla A4.7. función zoominterval	171
Tabla A4.8. función vertodo.	172
Tabla A4.9. función grabar.	173

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

Tabla A4.10. función crearred	174
Tabla A4.11. función entrenar	175
Tabla A4.12. función simular.	176
Tabla A4.13. función bordes	177
Tabla A4.14. función parametros2.	178
Tabla A4.15. función parametros3.	179
Tabla A4.16. fichero seral	180
Tabla A4.17. fichero acercade	180
Tabla A4.18. fichero dialogo	181
Tabla A4.19. fichero reconocimiento	181
Tabla A4.20. fichero verestadística.	182
Tabla A4.21. fichero nombrarloqu岸	182
Tabla A4.22. fichero mostrarmensaje	182
Tabla A4.23. fichero etiquetar	183
Tabla A4.24. fichero etiquetado.	183

Lista de figuras

Figura 1.1	Descomposición del RAV en sus dos disciplinas	13
Figura 1.2	Funcionamiento de un sistema RAL	14
Figura 1.3	Descomposición del RAL en sus dos disciplinas	15
Figura 1.4	Funcionamiento de un sistema VAL	15
Figura 1.5	Funcionamiento de un sistema IAL	16
Figura 1.6	Decisión en un sistema VAL	16
Figura 1.7	Decisión en un sistema IAL	16
Figura 3.1	Tracto vocal	21
Figura 3.2	Diagrama de un sistema de producción de voz	23
Figura 3.3	Aproximación de los CPL a la TRF.	24
Figura 3.4	Esquema de filtros pasa-banda para el calculo de la EBF.	25
Figura 3.5	Contexto de una trama de voz.	26
Figura 3.6	Esquema de una RNA	27
Figura 3.7	Funciones de transferencia de las neuronas	28
Figura 3.8	Percetrón unicapa (a) y perceptrón multicapa (b)	29
Figura 4.1	Ciclo de vida en espiral	31
Figura 4.2	Etapas de cada iteración.	32
Figura 5.1	Topología de la RNA implantada	37
Figura A1.1.	Grafo PERT del proyecto	50
Figura A1.2.	Grafo PERT del proyecto con tiempos	51
Figura A1.3.	Camino crítico del proyecto	52
Figura A1.4.	Ingresos, gastos y beneficio acumulado del proyecto.	55
Figura A2.1.	Apariencia de la ventana inicial de la aplicación	81
Figura A2.2.	Menú “Archivo”	81
Figura A2.3.	Menú “Operaciones”	82
Figura A2.4.	Menú “Estadísticas”	83
Figura A2.5.	Menú “Ayuda”	83
Figura A3.1	estructura del fichero de configuración	96
Figura A3.2	estructura del fichero de estadísticas	97
Figura A3.3	estructura del fichero de patrones	97
Figura A3.4	estructura del fichero de voz	98
Figura A3.5	estructura del fichero de etiquetas.	99
Figura A3.6	estructura del vector endpoint	100
Figura A3.7	estructura de la matriz de parámetros	100
Figura A3.8	Diagrama de contexto	101
Figura A3.9	DFD 0 SERAL.	103
Figura A3.10	DFD 1 Entrenar	104
Figura A3.11	DFD 2 Procesar señal de voz	105
Figura A3.12	DFD 1.1 extraer parámetros.	107
Figura A3.13	DFD 1.2 construir patrón de voz	108
Figura A3.14	DFD 1.3 almacenar patrón de voz	109
Figura A3.15	DFD 2.3 leer patrón de voz	110
Figura A3.16	DFD 2.4 comparar patrones de voz	111
Figura A3.17	DFD 3 eliminar locutor	112
Figura A3.18	DFD 4 visualizar señal de voz	113
Figura A3.19	DFD 5 reproducir señal de voz	114

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

Figura A3.20 DFD 6 escribir fichero de voz	115
Figura A3.21 DFD 7 leer fichero de voz	116
Figura A3.22 DFD 8 ver estadísticas	117
Figura A3.23 DFD 9 borrar estadísticas.	118
Figura A3.24 DFD 10 actualizar configuración	119
Figura A3.25 Vista de la venta principal de SERAL	121
Figura A3.26 Menús de la ventana principal de SERAL	121
Figura A3.27 Vista del cuadro de diálogo de salida.	123
Figura A3.28 Vista del cuadro de diálogo de abrir fichero.	124
Figura A3.29 Vista del cuadro de diálogo de mostrar estadísticas	125
Figura A3.30 Vista del cuadro de diálogo de borrar estadísticas.	126
Figura A3.31 Vista del cuadro de diálogo de ayuda	127
Figura A3.32 Vista del cuadro de diálogo de reconocimiento	128
Figura A3.33 Vista del cuadro de diálogo de mostrar resultados.	129
Figura A3.34 Vista del cuadro de diálogo de nombrar locutor	130
Figura A3.35 Vista del cuadro de diálogo de inicio de grabación.	131
Figura A3.36 Vista del cuadro de diálogo de etiquetar	132
Figura A3.37 Vista del cuadro de diálogo de ver etiquetas	133
Figura A3.38 Vista del cuadro de diálogo de fichero de etiquetas.	134
Figura A3.39 Algoritmo ENDPOINT	136
Figura A3.40 Algoritmo ENDPOINT (continuación).	137
Figura A3.41 Algoritmo PARAMETROS.	138
Figura A3.42 Algoritmo ZOOMMAS.	139
Figura A3.43 Algoritmo ZOOMMENOS	140
Figura A3.44 Algoritmo ZOOMINTERVALO.	141
Figura A3.45 Algoritmo REPRODUCEINTERVALO.	142
Figura A3.46 Árbol de directorios del sistema	149
Figura A4.1 Cambio del directorio actual	184
Figura A4.2 Ejecución del sistema.	184
Figura A4.3 Descenso del error en las RNA	185
Figura A4.4 Curva DET.	186
Figura A5.1 Menú Archivo -> Configurar.	199
Figura A5.2 Menú Archivo -> Abrir.	200
Figura A5.3 Cuadro de diálogo de abrir fichero	200
Figura A5.4 Menú Operaciones -> Reproducir.	201
Figura A5.5 Osciloscopio	201
Figura A5.6 Menú Operaciones -> Grabación	202
Figura A5.7 Cuadro de diálogo de grabar fichero.	202
Figura A5.8 Menú Operaciones -> Zoom -> Acercar.	203
Figura A5.9 Menú Operaciones -> Zoom -> Alejar.	204
Figura A5.10 Menú Operaciones -> Zoom -> Intervalo	205
Figura A5.11 Selección de un intervalo en el osciloscopio	205
Figura A5.12 Menú Operaciones -> Zoom -> Vista inicial.	206
Figura A5.13 Menú Estadísticas -> Ver	207
Figura A5.14 Cuadro de diálogo de mostrar estadísticas.	207
Figura A5.15 Menú Estadísticas -> Borrar	208
Figura A5.16 Cuadro de diálogo de borrar estadísticas	208
Figura A5.17 Menú Operaciones -> Entrenamiento	209

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

Figura A5.18 Cuadro de diálogo de introducción de nombre de locutor. .	209
Figura A5.19 Cuadros de diálogo de introducción de un fichero de voz del impostor (izquierda) y del locutor auténtico (derecha)	209
Figura A5.20 Menú Operaciones -> Eliminar locutor.	211
Figura A5.21 Cuadro de diálogo de introducción de nombre de locutor. .	211
Figura A5.22 Menú Operaciones -> Reconocimiento -> Verificación . .	212
Figura A5.23 Cuadro de diálogo de introducción de nombre de locutor. .	212
Figura A5.24 Cuadro de diálogo de selección de opción.	212
Figura A5.25 Cuadro de diálogo de selección de fichero de voz.	213
Figura A5.26 Menú Operaciones -> Reconocimiento -> Verificación . .	214
Figura A5.27 Cuadro de diálogo de introducción de nombre de locutor. .	214
Figura A5.28 Cuadro de diálogo de selección de opción.	214
Figura A5.29 Cuadro de diálogo de inicio de grabación	215
Figura A5.30 Menú Operaciones -> Reconocimiento -> Identificación . .	216
Figura A5.31 Cuadro de diálogo de selección de opción.	216
Figura A5.32 Cuadro de diálogo de selección de fichero de voz.	217
Figura A5.33 Menú Operaciones -> Reconocimiento -> Identificación . .	218
Figura A5.34 Cuadro de diálogo de selección de opción.	218
Figura A5.35 Cuadro de diálogo de inicio de grabación	218
Figura A5.36 Detección de inicio y fin de palabras.	220
Figura A5.37 Selección de dos puntos de la señal de voz	221
Figura A5.38 Cuadro de diálogo de etiquetar	221
Figura A5.39 Señal de voz etiquetada.	221
Figura A5.40 Botones Espectro TRF y Espectro LPC	222
Figura A5.41 (a)Espectrograma calculado con TRF (b)Espectrograma calculado con LPC.	222
Figura A5.42 Menú Ayuda	223
Figura A5.43 Cuadro de diálogo de Ayuda Acerca de....	223

Prólogo

Este proyecto de fin de carrera, es fruto del trabajo e investigación realizado durante los últimos meses en el campo del reconocimiento automático de la voz (RAV) y más concretamente en una de las dos disciplinas en las que se descompone como es el Reconocimiento Automático del Locutor (RAL).

El Reconocimiento Automático del Locutor (RAL), es una disciplina que presenta gran interés como campo de investigación, debido a que es una área que no está aún lo suficientemente desarrollada a pesar del esfuerzo realizado durante los últimos años por diversos grupos de investigadores.

El Reconocimiento Automático del Locutor (RAL), junto con otras técnicas de reconocimiento biométrico (como reconocimiento por iris o por huellas dactilares) han cobrado gran importancia en los últimos años gracias a sus aplicaciones a sistemas de control de acceso a equipos informáticos o control de acceso físico a distintas instalaciones.

Desde el punto de vista organizativo, este documento se divide en seis partes según la estructura propuesta por Francisco José García Peñalvo, Jesús Manuel Maudes Raedo, Mario Gerardo Piattini Velthius, José Rafael García-Bermejo Giner y María N. Moreno García en el documento titulado “Proyecto de final de carrera en la Ingeniería Técnica en Informática, Guía de realización y documentación”:

Parte 1: Compuesta por los 7 primeros puntos de la memoria, está dedicada a la descripción del problema al que se intenta dar solución, esbozando también la solución dada. Esta parte finaliza con una conclusión y unas líneas de trabajo futuras.

Parte 2: Formada por el Anexo1, que incluye un estudio de la viabilidad del proyecto y un calendario para llevarlo a cabo.

Parte 3: Formada por el Anexo 2, incluye los requisitos planteados, que debe cumplir el sistema desarrollado

Parte 4: Formada por el Anexo 3, que contiene los detalles sobre el diseño del sistema, aquí se describe la solución que se va a emplear, y se marca el camino a seguir durante el desarrollo del sistema.

Parte 5: Formada por el Anexo 4, que incluye documentación asociada al código fuente desarrollado y a bibliotecas utilizadas para su desarrollo

Parte 6: Formada por el Anexo 5, que contiene manuales, tutoriales o guías de usuario, necesarias para el manejo de la aplicación.

1. Introducción

El Reconocimiento Automático de la Voz (RAV) ha crecido en importancia durante los últimos años, sin embargo no puede tratarse desde un punto de vista general, sino que ha de tratarse desde las dos disciplinas en las que se descompone: el Reconocimiento Automático de la Habla (RAH) y el Reconocimiento Automático de Locutor (RAL).

El RAH es el método por el cual una máquina es capaz de reconocer todas o alguna de las palabras pronunciadas por un locutor.

El RAL, por el contrario es el método que permite a una máquina saber quién es el que está pronunciando las palabras. Es en esta disciplina donde se centrará el proyecto.

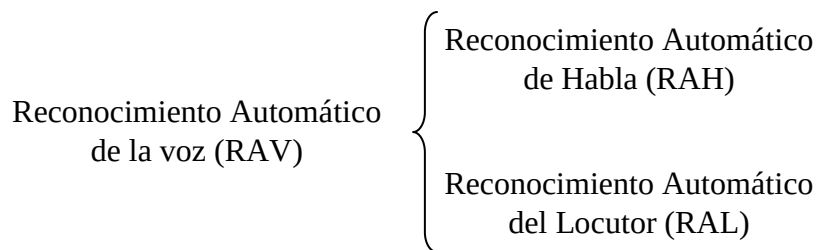


Figura 1.1. Descomposición del RAV en sus dos disciplinas

1.1 El Reconocimiento Automático de Locutor (RAL)

El RAL es un sistema de Identificación Biométrica Personal (IBP), 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 [Doddington, 1985].

Un sistema de RAL 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 RAL 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.

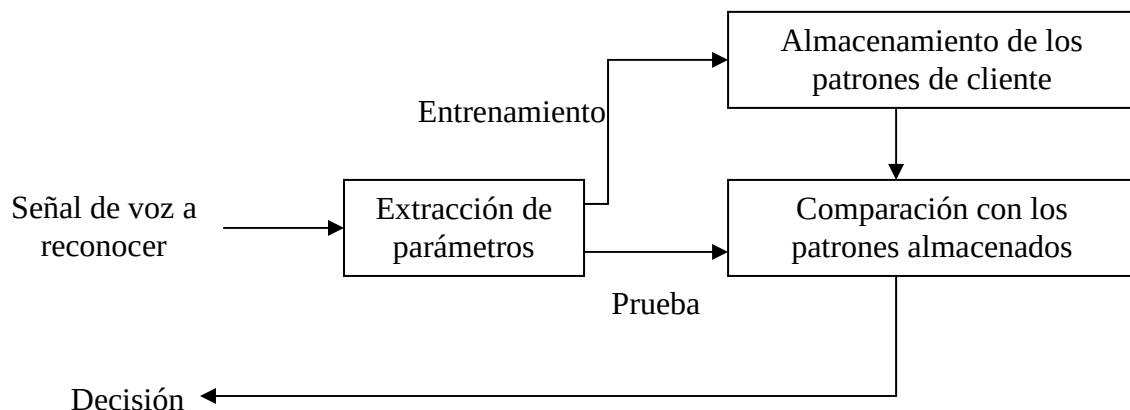


Figura 1.2. Funcionamiento de un sistema RAL

1.1.1 Verificación e Identificación del locutor

Dentro del RAL se pueden encontrar a su vez dos disciplinas claramente diferenciadas como son la Verificación Automática de Locutor (VAL) y la Identificación Automática de Locutor (IAL).

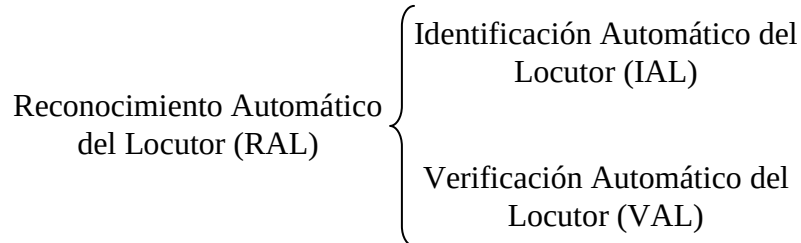


Figura 1.3. Descomposición del RAL en sus dos disciplinas

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

Un sistema de VAL 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.

Como consecuencia de estas decisiones se producen dos tipos de errores:

- Falso Rechazo (FR): Un usuario auténtico es rechazado por el sistema.
- Falsa Aceptación (FA): Un usuario falso (impostor) es aceptado por el sistema.

El esquema de funcionamiento de un sistema de VAL se muestra en la figura 1.4

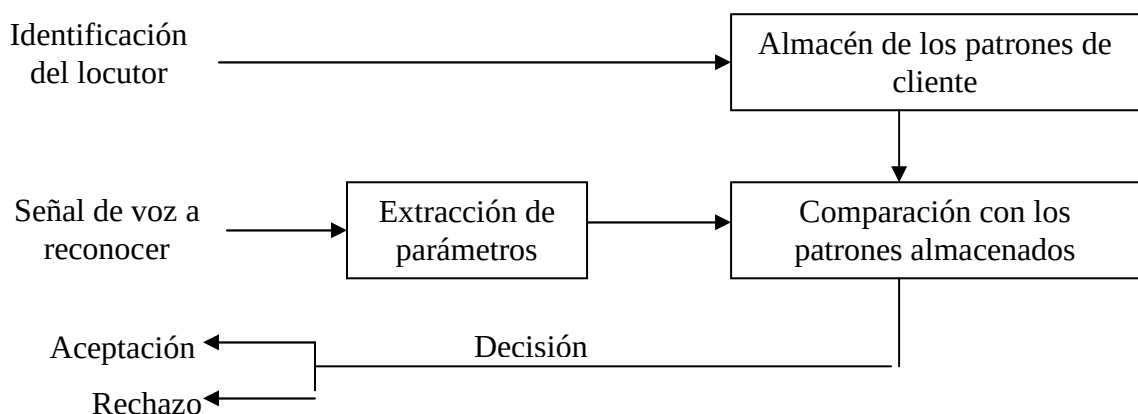


Figura 1.4. Funcionamiento de un sistema VAL

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

El IAL consiste en decidir a que locutor de los reconocidos por el sistema pertenece una muestra de voz desconocida.

Un sistema de IAL puede producir dos resultados [Soong, Rosenberg y Juang, 1987]:

- 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.

Como consecuencia de estas decisiones se producen dos tipos de errores:

- Falsa Identificación (FI): Un usuario o impostor es identificado erróneamente, es decir, se le asigna al locutor una identidad que no le corresponde.
- Falsa No Identificación (FNI): Un usuario del sistema no es identificado por el mismo.

El esquema de funcionamiento de un sistema de IAL se muestra en la figura 1.5

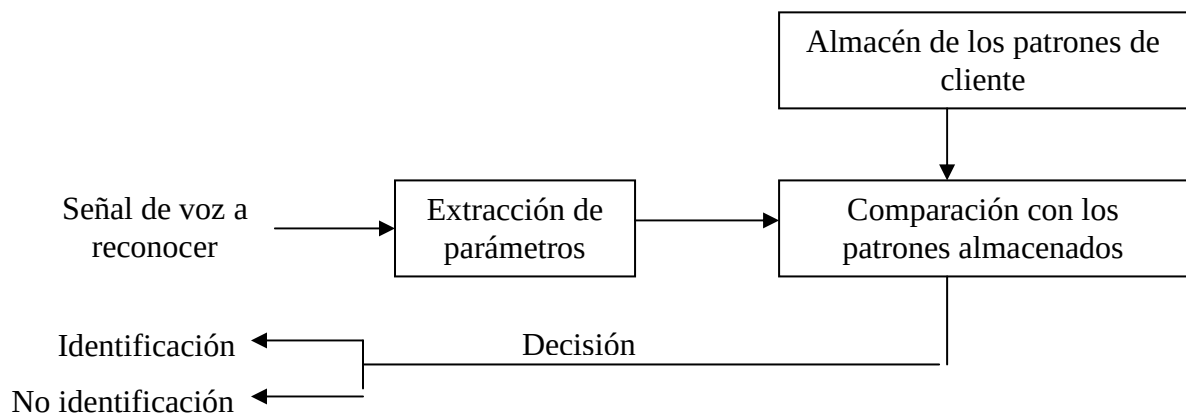


Figura 1.5. Funcionamiento de un sistema IAL

La toma de decisiones en un sistema de VAL se realiza de la siguiente manera:

$$\text{Decisión} \begin{cases} \text{Aceptación: } D(m, P) < U \\ \text{Rechazo: En otro caso} \end{cases}$$

Figura 1.6. Decisión en un sistema de VAL

Siendo $D(m, P)$ la distancia entre la muestra de voz (m) introducida en el sistema y el patrón (P) almacenado en él y U un umbral de decisión que normalmente es la Tasa de EquiError (TEE), que en un VAL es el umbral para el cual las FA son iguales a los FR.

Por el contrario en un sistema de IAL las decisiones se toman de la siguiente manera:

$$\text{Decisión} \begin{cases} \text{Locutor } i: & D(m, P_i) < D(m, P_j) \text{ y } D(m, P_i) < U \\ \text{No identificado: En otro caso} \end{cases}$$

Figura 1.7. Decisión en un sistema de IAL

Siendo $D(m, P_i)$ la distancia entre la muestra de voz (m) introducida en el sistema y el patrón del usuario i (P_i) almacenado en él, $D(m, P_j)$ la distancia entre la muestra de voz (m) introducida en el sistema y el patrón del usuario j (P_j) almacenado en él y U un umbral de decisión que normalmente es la Tasa de EquiError (TEE), que en un IAL es el umbral para el cual las FI son iguales a los FNI.

La utilización de la TEE como umbral U tanto en la VAL como en la IAL, se debe a un intento por normalizar dicho umbral, siendo más eficiente un sistema cuanto más baja es su TEE, siendo más baja generalmente para un sistema VAL que para uno IAL [Doddington, 1985].

1.1.2 Dependencia e Independencia del Texto

Un aspecto muy importante en los sistemas de RAL 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 RAL-Dependientes del Texto (RAL-DT), por el contrario en aquellos que se puede elegir la frase a introducir libremente son denominados RAL-Independientes del Texto (RAL-IT).

La dependencia del texto es muy empleada en VAL y muy raramente en IAL donde se usa más la independencia del texto.

Los sistemas RAL-DT presentan mejores resultados y requieren menos tiempo en las fases de entrenamiento y prueba que los RAL-IT, ya que el vocabulario es conocido, sin embargo también son más susceptibles al fraude [O'Shaughnessy, 1987].

1.2 Contenido y Estructura de la memoria

El contenido que se desarrolla a lo largo de esta memoria pretende servir a modo de guía o manual para los posibles usuarios del sistema que se desarrolla en este proyecto fin de carrera.

Así mismo también se pretende que esta memoria sirva como documentación clara para la realización de nuevas versiones, actualizaciones o modificaciones del sistema por parte de personas que no estaban involucradas en su desarrollo inicial.

Con vistas en poder cumplir con estos objetivos se desarrollará la memoria conforme a estos puntos:

Descripción del proyecto:

1. Introducción
2. Objetivos del proyecto
3. Conceptos teóricos
4. Técnicas y herramientas
5. Aspectos relevantes del desarrollo del proyecto
6. Trabajos relacionados
7. Conclusiones y líneas de trabajo futuras

Documentación técnica:

- Anexo 1. Plan del Proyecto Software
- Anexo 2. Especificación de Requisitos del Software
- Anexo 3. Especificación de Diseño
- Anexo 4. Documentación Técnica de Programación
- Anexo 5. Manuales de Usuario

2. Objetivos del proyecto

En este apartado de la memoria se especificarán los objetivos especificados por los requisitos y de carácter técnico que se marcan en el proyecto.

2.1 Objetivos especificados por los requisitos

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.

2.2 Objetivos técnicos

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. Conceptos teóricos

3.1 Características de la voz

La señal de voz, se produce cuando un flujo de aire iniciado en los pulmones, atraviesa ciertos obstáculos y cavidades situados entre la glotis y la boca (o nariz) conocidos como tracto vocal [Tapiador y Sigüenza, 2005].

El tracto vocal esta formado por los siguientes componentes:

- Traquea
- Cuerdas vocales
- Velo del paladar
- Tracto nasal
- Tracto oral

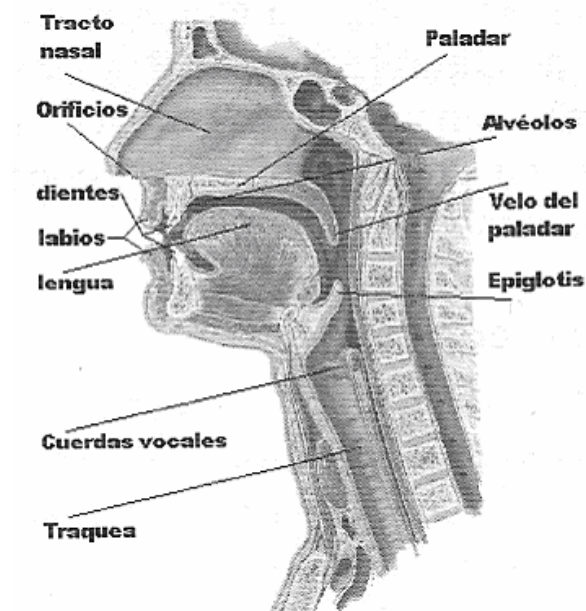


Figura 3.1 Tracto vocal

La señal de voz depende de las características físicas y emocionales del locutor como son las dimensiones de las cuerdas y el trato vocal, el estado de salud, el estado de ánimo, los hábitos lingüísticos, etc. Además, hay que tener en cuenta el entorno en el que se produce la señal de voz ya que las condiciones del entorno pueden distorsionar la señal.

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

Las características de la voz de una persona se pueden clasificar en de alto nivel o de bajo nivel, en la siguiente tabla se detallan cada una de ellas

	Características	Causas
Alto nivel	pronunciación, entonación, duración	socio-económicas, educación, lugar de nacimiento
	ritmo, velocidad, volumen	personalidad, influencia paterna
Bajo nivel	tono, frecuencias	estructura anatómica del aparato vocal.

Tabla 3.1. Características de la voz

Hasta la fecha los sistemas que han obtenido los mejores resultados usan características denominadas de bajo nivel, sin embargo se sabe que las características de alto nivel también son capaces de discernir locutores.

Cabe destacar que estas características son las que determinan los parámetros de la voz, sin embargo estos parámetros no se pueden extraer directamente de una muestra de voz, sino que previamente la muestra de voz se divide en fragmentos denominados ventanas de pocos milisegundos cada uno, para posteriormente analizar cada ventana y extraer los parámetros presentes en ellas.

3.2 Coeficientes de Predicción Lineal (CPL)

Es un modelo para extraer los parámetros de una muestra de voz.

Considera el sistema de producción de voz como un sistema de la forma:

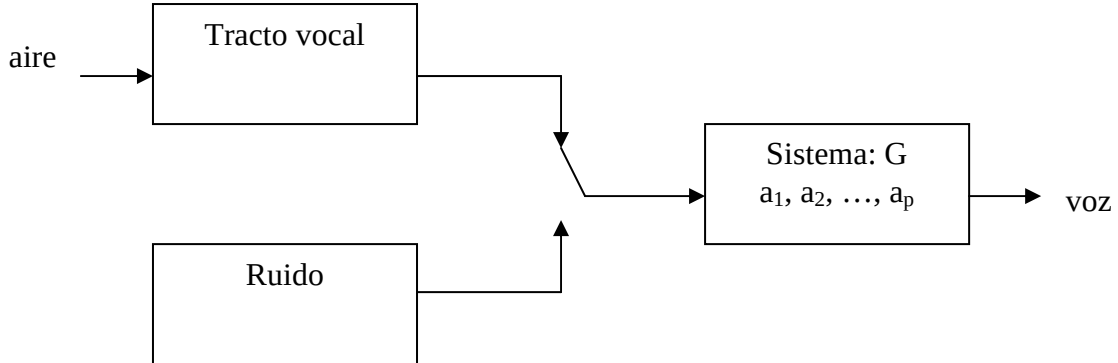


Figura 3.2 Diagrama de un sistema de producción de voz

Este sistema se puede expresar también con la ecuación:

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 + \sum_{K=1}^p a_k z^{-k}}$$

Donde los parámetros a considerar son la ganancia (G) y los coeficientes de predicción lineal (a_k), y donde $S(z)$ es la salida y $U(z)$ es un ruido aleatorio.

Hay que determinar el orden p de los coeficientes, para ello hay que tener en cuenta que si se quiere un sistema perfecto, el orden tiene que ser infinito, pero esto es lógicamente imposible, en la práctica, el orden p depende de la frecuencia de muestreo, y se elige un orden al menos dos veces superior al tiempo necesario para que las ondas sonoras viajen desde la glotis hasta los labios [Atal y Hanauer, 1971].

Teniendo en cuenta que la distancia entre glotis y labios son aproximadamente 17 cm, el tiempo que necesitan las ondas para recorrerlo es:

$$t = \frac{0.17m}{340m/s} = 0.5ms$$

Por tanto el orden p será:

$$p = 2 \cdot 0.0005 \cdot F_m$$

Donde F_m es la frecuencia de muestreo, y como esta suele estar entre 8000 y 10000 Hz, se elige un orden $p = 10$, con el que se obtienen unos resultados buenos [O'Shaughnessy, 1987].

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

La siguiente figura muestra la aproximación conseguida con un orden 12 en los CPL (línea negra), respecto de la Transformada Rápida de Fourier de 512 puntos (línea roja).

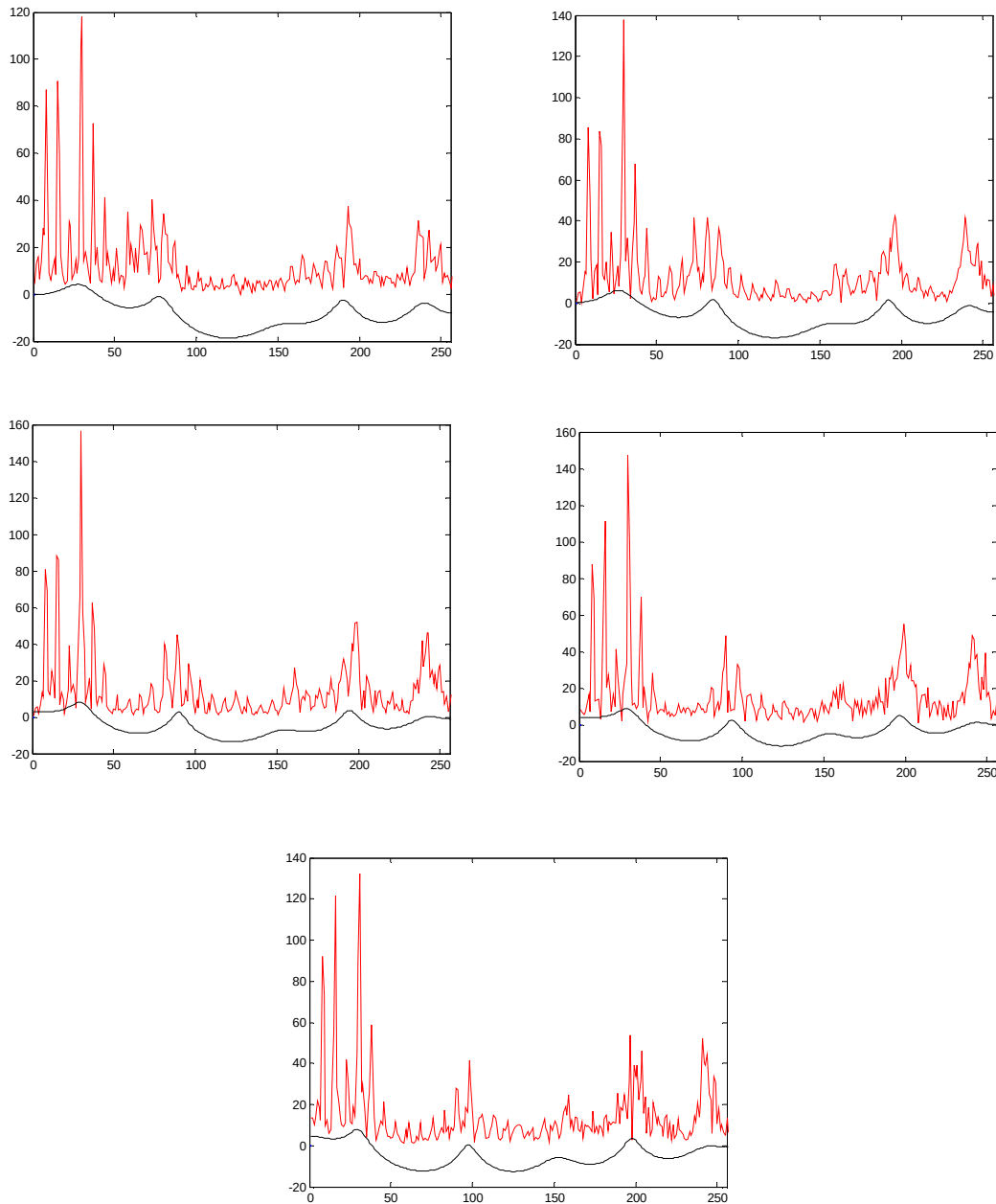


Figura 3.3 Aproximación de los CPL a la TRF

Existen varios métodos para el cálculo de los coeficientes de predicción lineal como son la varianza, malla, autocorrelación, etc., pero de entre ellos destaca el de la autocorrelación por presentar las siguientes ventajas [Rabiner y Schafer, 1978]:

- Los requisitos de memoria son de los más bajos
- Los requisitos de cómputo son de los más bajos

Por ello, el método de la autocorrelación, será el que se utilice en el sistema SERAL.

3.3 Energía por bandas de frecuencia (EBF)

La idea de este tipo de parámetros, es dividir la señal de voz en bandas de frecuencia con una serie de filtros pasa-banda, que cubren todo el ancho de banda de la señal de entrada como muestra la siguiente figura:

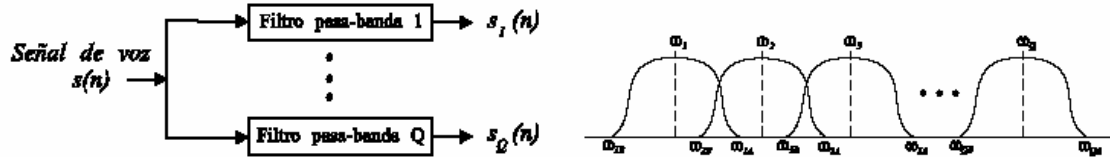


Figura 3.4 Esquema de filtros pasa-banda para el calculo de la EBF

Después de la división de la señal se realiza una suma de la energía acumulada para cada una de las bandas, de la siguiente forma

$$E(i) = \sum_{j=1}^{n(i)} s(j)$$

Donde $E(i)$ es la energía de la banda i , $n(i)$, el número de componentes de la banda i y $s(j)$, el j -ésimo componente de la banda i .

Posteriormente, la energía de cada una de las bandas tiene que ser normalizada, para ello, se calculan los valores mínimo y máximo de las energías de todas las bandas:

$$\begin{aligned} Min &= \min\{E(i)\} \\ Max &= \max\{E(i)\} \end{aligned}$$

La normalización se lleva a cabo con la siguiente ecuación:

$$EN(i) = \frac{E(i) - Min}{Max - Min}$$

3.4 Información transicional (parámetros Delta)

Todos los parámetros que se obtienen segmentado la señal de en tramas y extrayendo de ellas información representativa de la misma, se considera información instantánea. [Furi,1993]

La información transicional, se obtiene calculando la primera derivada de cada uno de los parámetros que constituyen el vector de parámetros de la trama, está información, es también conocida como parámetros delta (o delta-delta, si se calcula la segunda derivada).

Para el cálculo de este tipo de parámetros, es necesario tener en cuenta los parámetros de las tramas adyacentes, como muestra la figura:

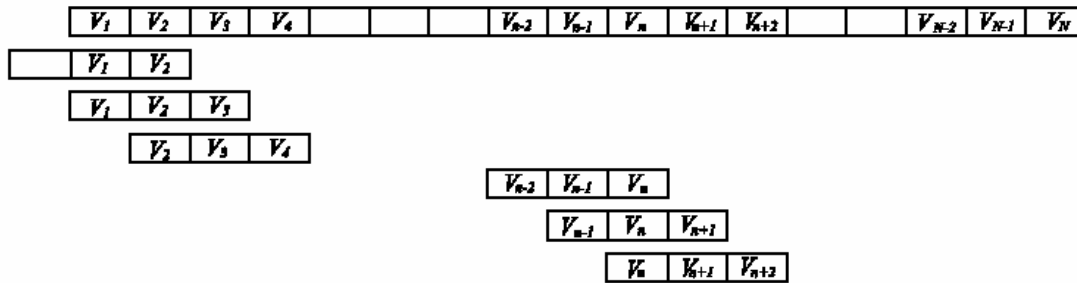


Figura 3.5 Contexto de una trama de voz

Una de las formas más extendida de calcular dichos parámetros es la siguiente:

$$\delta_i(n) = \frac{\sum_{j=-k}^k k \cdot v_i(n+k)}{\sum_{j=-k}^k j^2}$$

donde k, es el número de tramas adyacentes que se tienen en cuenta por cada lado de la trama analizada, $\delta_i(n)$, es el parámetro delta correspondiente al parámetro i-ésimo de la trama n-ésima, y $v_i(n)$, es el parámetro i-ésimo de la trama n-ésima

3.5 Redes Neuronales Artificiales (RNA)

Las redes neuronales artificiales son una de las últimas incorporaciones al proceso de reconocimiento automático de voz.

En las RNA las características de las señales se distribuyen a través de unidades de procesamiento (neuronas) que se encuentran conectadas entre si mediante una red, cuyas características dependen de su arquitectura [Dayhoff, 1990].

El objetivo de los sistemas de entrenamiento de las RNA es el de optimizar éstas redes.

Una red neuronal artificial estará compuesta por un conjunto de unidades de procesamiento o neuronas, que poseen varias entradas y una o más salidas y que se comunican mandando señales a través de conexiones que tienen unos pesos que se han establecido en la fase de entrenamiento, por tanto los principales elementos de una RNA son:

- Las unidades de procesamiento o neuronas.
- Un estado de activación (a_i), que depende del valor de la salida de cada neurona.
- Conexiones entre las neuronas i y j que tienen asociado un peso w_{ij} .
- Una regla que determine la entrada (i_i) a una neurona a partir de las salidas de las neuronas con las que se conecta y de los pesos w_{ij} .
- Una función F que determine la salida de una neurona a_i a partir de su entrada i_i .
- Una entrada de polarización (θ_i) por cada neurona.

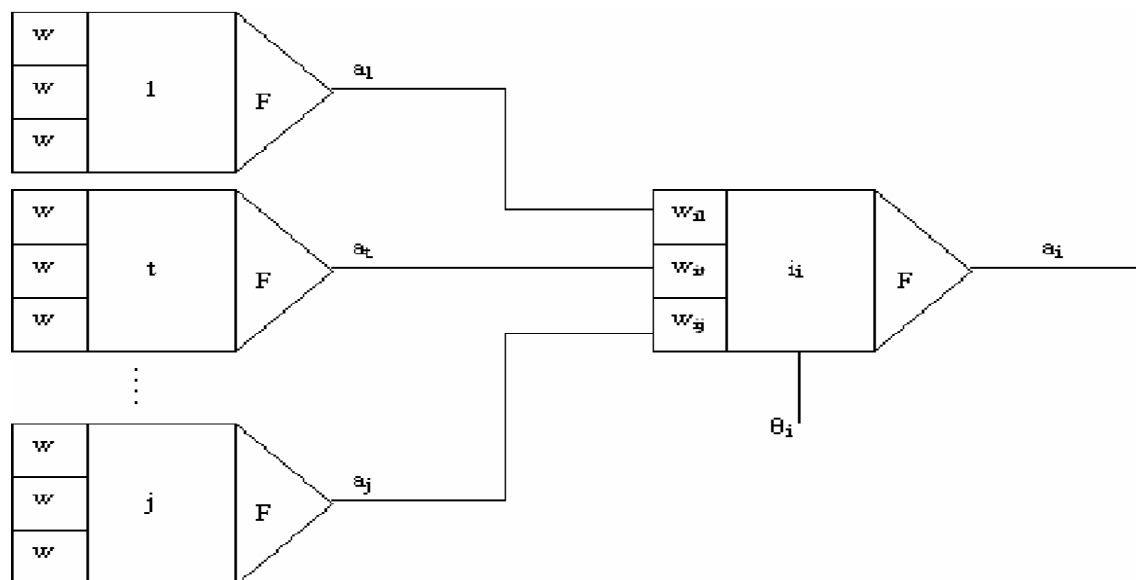


Figura 3.6 Esquema de una RNA

Se puede distinguir dentro de una RNA tres tipos de neuronas, las unidades de entrada, que reciben datos del exterior del sistema, las unidades de salida que envían datos hacia el exterior del sistema y las unidades oculta, cuyas entradas y salidas permanecen dentro del sistema [Kröse y Smagt, 1993].

De forma general, las neuronas tienen en cuenta sus entradas de forma aditiva para calcular la entrada total de tal manera que:

$$i_i(t) = \sum_j w_{i,j}(t) \cdot a_j(t) + \theta_i(t)$$

Para obtener la salida (estado de activación) de la neurona, se aplica una función de transferencia (F) a la entrada i_i .

$$a_i(t+1) = F_i(i_i(t)) = F_i\left(\sum_j w_{i,j}(t) \cdot a_j(t) + \theta_i(t)\right)$$

Se suele utilizar como función de transferencia F, funciones umbral, como la función signo o semi-lineal, o una que se aproxime suavemente a la semi-lineal como es la sigmoide (que es la más empleada), cuya ecuación se muestra a continuación:

$$F(i_i) = \frac{1}{1 + e^{-i_i}}$$

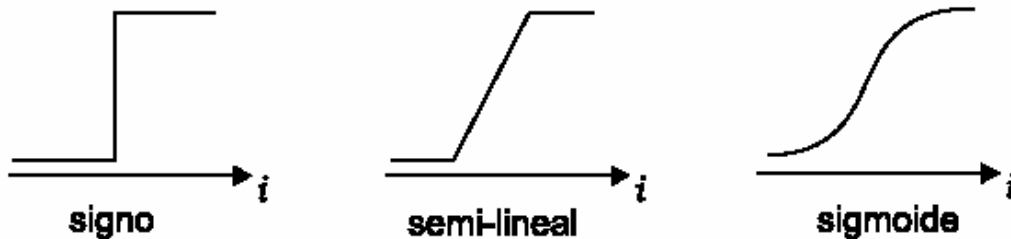


Figura 3.7 Funciones de transferencia de las neuronas

3.5.1 Topología de una RNA

Las conexiones entre las distintas neuronas de una RNA se pueden realizar de manera arbitraria, no obstante existen tres formas de conexión o topologías que tienen especial interés [Rabiner y Juang, 1993]:

1. Perceptrones multicapa o unicapa.
2. RNA recurrentes o de Hopfield.
3. RNA de Kohonen o auto-organizadas.

En los perceptrones multicapa o unicapa, las salidas de las neuronas de una capa son las entradas de las neuronas de la capa siguiente, siendo imposible la realimentación, es decir que la salida de una neurona de una capa sea la entrada a otra de esa misma capa o otra anterior.

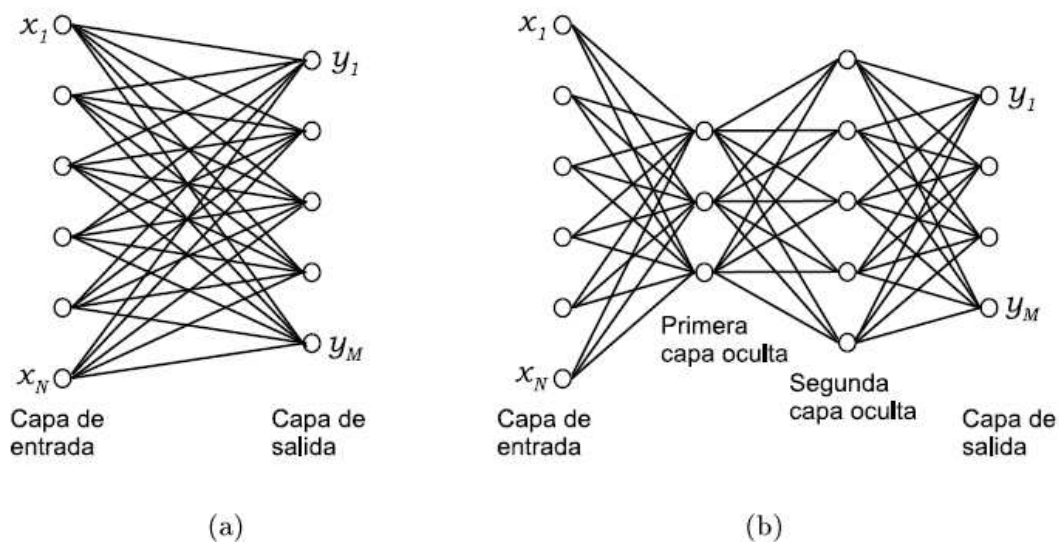


Figura 3.8 Perceptrón unicapa (a) y perceptrón multicapa (b)

Las RNA recurrentes o de Hopfield, incluyen la salida de una neurona como una de las entradas a la misma, es decir, están realimentadas.

Tanto los perceptrones como las redes de Hopfield se entrenan presentándole pares de patrones de entrada y salida (X^p , O^p), donde X^p es el p-ésimo patrón de entrada y O^p es el patrón de salida esperado para esa entrada.

Las RNA de Kohonen o auto-organizadas se emplean en aquellos casos que no tenemos pares de patrones asociados (X^p , O^p).

3.5.2 Aprendizaje en las RNA

El aprendizaje en una RNA consiste en calcular los pesos w_{ij} tal que un conjunto de entradas produzcan las salidas adecuadas.

Existen dos formas de realizar este aprendizaje, en ambas el objetivo es el de ajustar los pesos de las distintas conexiones:

1. Aprendizaje supervisado o asociativo en el que la red se entrena dándole patrones de entrada-salida, cada una de las salidas de la red es una categoría y se debe aprender cuales de los patrones de entrada pertenecen a cada una de ellas.
2. Aprendizaje no supervisado o auto-organizado en el que cada neurona de salida debe responder a ciertos grupos de patrones de los dados en el entrenamiento.

En ambos paradigmas de aprendizaje se produce un ajuste de los pesos asociados a cada una de las conexiones de acuerdo con alguna regla de aprendizaje.

Para cada uno de los dos paradigmas, existen gran cantidad de reglas de aprendizaje que son susceptibles de ser utilizadas.

4. Técnicas y herramientas

En este punto de la memoria se describirán las diferentes técnicas y herramientas empleadas para el desarrollo de este proyecto, explicando algunos de los aspectos más importantes de las mismas, así como las razones para su utilización.

4.1 Técnicas

Este apartado se dedicará a la descripción de las técnicas de desarrollo de software empleadas en el presente proyecto.

4.1.1 Ciclo de vida en espiral

El ciclo de vida empleado para el desarrollo del sistema desarrollado con el presente proyecto ha sido un ciclo de vida en espiral.

La razón para la elección de este ciclo de vida ha sido que no presenta los problemas de los ciclos de vida secuenciales, que no se ajustan a la realidad del desarrollo del software.

El ciclo de vida en espiral elegido ha sido uno con cuatro tareas en cada iteración como el que se representa en la siguiente figura:

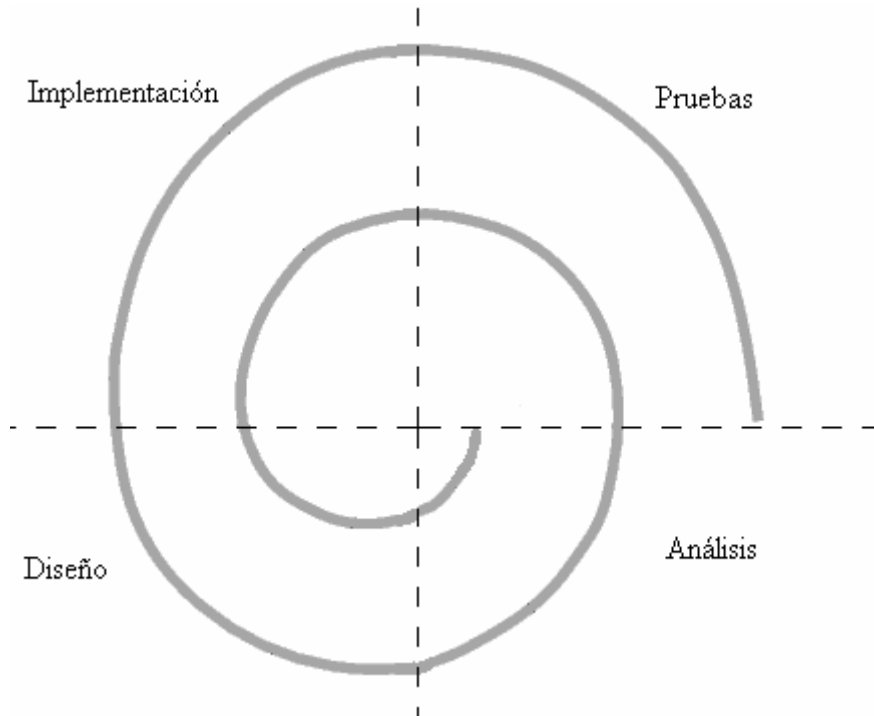


Figura 4.1 Ciclo de vida en espiral

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

Más concretamente la sucesión de tareas de cada iteración se puede observar en la siguiente figura:

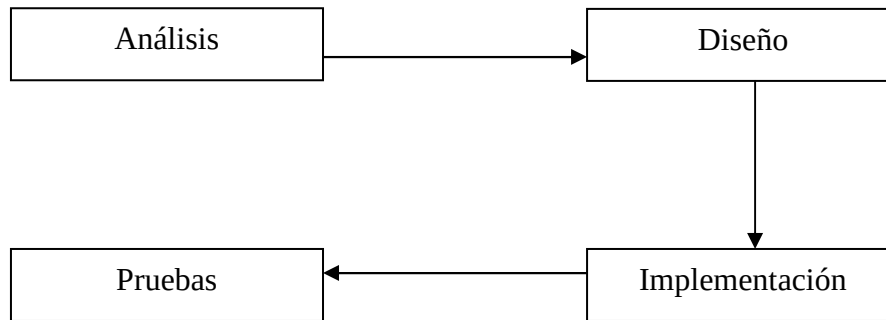


Figura 4.2 Etapas de cada iteración

Las ventajas que presenta este ciclo de vida sobre otros son:

- Mayor realismo en el desarrollo del software.
- Gran flexibilidad.
- Eliminación de errores desde el principio.
- Favorece el crecimiento del sistema

En el caso concreto de este proyecto, el sistema creció de un módulo para reproducir ficheros de audio, desarrollado con la primera iteración hasta llegar al sistema completo en la última iteración.

4.2 Herramientas

Este apartado se dedicará a la descripción de las herramientas y librerías empleadas para el desarrollo de este proyecto.

4.2.1 MatLab

El nombre de MATLAB proviene de la contracción de los términos MATrix LABoratory.

MatLab es un programa interactivo para computación numérica y visualización de datos. Es ampliamente usado por ingenieros debido a que posee una gran versatilidad y capacidad para resolver problemas en matemática aplicada, física, química, ingeniería, finanzas y muchos otros ámbitos. Está basado en un sofisticado software de matrices para el análisis de sistemas de ecuaciones. Permite resolver complicados problemas numéricos de manera fácil y rápida.

MatLab es un entorno de computación y desarrollo de aplicaciones totalmente integrado orientado para llevar a cabo proyectos en donde se encuentren implicados elevados cálculos matemáticos y la visualización gráfica de los mismos.

MatLab es un sistema de trabajo interactivo cuyo elemento básico de trabajo son las matrices. El programa permite realizar de un modo rápido la resolución numérica de problemas en un tiempo mucho menor que si se quisiesen resolver estos mismos problemas con lenguajes de programación tradicionales como pueden ser los lenguajes Fortran, Basic o C.

MatLab goza en la actualidad de un alto nivel de implantación en la resolución de problemas de ingeniería de control, álgebra lineal, proceso digital de imágenes y señales, etc.

MatLab dispone también en la actualidad de un amplio abanico de librerías especializadas denominadas Toolboxes, que extienden significativamente el número de funciones incorporadas en el programa principal. Estas Toolboxes cubren en la actualidad prácticamente casi todas las áreas principales en el mundo de la ingeniería y la simulación, destacando entre ellos la Toolbox de proceso de imágenes y señales, control robusto, estadística, análisis financiero, matemáticas simbólicas, redes neuronales, lógica difusa, identificación de sistemas, simulación de sistemas dinámicos, etc.

MatLab está disponible para una amplio número de plataformas: SUN, HP, Apple, PC, etc. y opera bajo sistemas operativos UNIX, Macintosh y Windows.

La elección de MatLab como entorno de desarrollo de este proyecto se debe a que MatLab integra análisis numérico, cálculo matricial, proceso de señal y visualización gráfica, todas ellas necesarias para poder desarrollar un sistema de reconocimiento de locutor como el que se desarrolla en el presente proyecto.

4.2.2 Image Processing Toolbox

La Image Processing Toolbox, es una de las Toolboxes que forman parte de MATLAB.

Esta Toolbox se utiliza para la adquisición de imágenes y su procesado. En este proyecto, se utiliza únicamente la primera funcionalidad (adquisición de imágenes) para poder cargar el logotipo del sistema de un fichero con formato gráfico (.tiff) y poder mostrarlo en la ventana de ayuda acerca de...

4.2.3 Signal Processing Toolbox

La Signal Processing Toolbox es otra de las Toolboxes de MATLAB.

El propósito de esta Toolbox es el de procesado de señales, en particular para este proyecto tienen especial interés las señales de voz.

Esta Toolbox, permite el análisis de señales para poder extraer los parámetros característicos de las mismas como pueden ser los coeficientes de predicción lineal o la Transformada Rápida de Fourier.

4.2.4 Neural Network Toolbox

La Neural Network Toolbox también forma parte del conjunto de Toolboxes de MATLAB.

Esta Toolbox permite crear varios tipos de Redes Neuronales Artificiales y realizar diversas operaciones sobre ellas, concretamente, para la realización de este proyecto se han empleado las Redes Neuronales Artificiales perceptrón multicapa, y se han realizado operaciones de entrenamiento y simulación.

4.2.5 HTM2CHM

HTM2CHM es un sencillo programa que nos permite generar archivos de ayuda en formato chm a partir de una serie de archivos en formato htm.

Para poder generar el archivo chm, los archivos htm origen tienen que estar organizados de forma jerárquica en un directorio, de la misma forma en la que se organizará el índice del archivo de ayuda chm.

Esta aplicación fue empleada para generar el archivo de ayuda del sistema SERAL que se desarrolla en este proyecto.

5. Aspectos relevantes del desarrollo del proyecto

En esta sección, se pretende justificar el porque de las opciones elegidas en las cuestiones más importantes planteadas durante el desarrollo del proyecto.

5.1 Aspectos relevantes del análisis

La decisión más importante a tomar durante la fase de análisis, es la selección del entorno de desarrollo con el que se realizara el sistema software.

Las posibles opciones que se presentan son dos, la primera es optar por un entorno de desarrollo de JAVA como es Eclipse o NetBeans y la segunda es optar por el entorno de desarrollo de MATLAB.

Finalmente, la decisión tomada, es la de elegir el entorno de desarrollo de MATLAB. Esta decisión se toma debido a las ventajas que ofrece MATLAB sobre la otra alternativa en los campos de procesamiento de señales y redes neuronales artificiales, que tienen gran importancia para la realización del proyecto.

Estas ventajas ofrecidas por MATLAB son las siguientes:

- Implementación de diversas operaciones para el procesamiento de señales.
- Implementación de diversas topologías de redes neuronales artificiales.
- Implementación de diversas operaciones sobre las redes neuronales artificiales.
- Disponibilidad de profesorado con experiencia en su uso.

5.2 Aspectos relevantes del diseño

En la fase de diseño se plantean diversas cuestiones importantes a las que hay que responder como son:

- El algoritmo utilizado para la detección de inicio y fin de palabras.
- Los parámetros elegidos para la construcción del patrón del locutor.
- La técnica utilizada para la construcción del patrón del locutor.
- La distribución de la información en directorios.

Respecto a la primera de las cuestiones, que es la selección de un algoritmo para la detección de inicio y fin de palabras, se puede optar por elegir uno de los muchos algoritmos existentes o por desarrollar un algoritmo propio.

La opción elegida para esta cuestión es la de implantar un algoritmo basado en el propuesto por Lamel en [Lamel, Rabiner, Rosenberg y Wilpon, 1991], que es un algoritmo basado en el nivel de energía.

La elección de esta opción se debe a las tres ventajas que ofrece sobre las otras opciones:

- Facilidad de implantación.
- Eficiencia de ejecución.
- Resultados adecuados para un nivel moderado de ruido.

El segundo problema que se plantea es el de elegir los parámetros que se van a utilizar para construcción de los patrones de los diversos locutores del sistema.

De los muchos parámetros que se pueden utilizar para este propósito, los elegidos son los coeficientes de predicción lineal (CPL), la Transformada Rápida de Fourier (TRF), la Tasa de Cruces por Cero (TCC), la Energía por Bandas de Frecuencia (EBF) y los parámetros delta.

Las razones por las que se eligen estos parámetros son las siguientes:

- Ofrecen la información necesaria la construcción de los patrones de los diversos locutores del sistema.
- Han mostrado ser eficientes en el modelado y representación de señales de voz.
- MATLAB ofrece gran facilidad para su cálculo.

La tercera cuestión planteada, y la más importante dentro de la fase de diseño, es la selección de la técnica utilizada para la construcción del patrón del locutor.

Como solución a este problema, se plantean tres posibles alternativas:

- Análisis Dinámico Temporal (ADT)
- Redes Neuronales Artificiales (RNA)
- Modelos Ocultos de Markov (MOM)

La decisión tomada es la de implantar las Redes Neuronales Artificiales, debido a la facilidad que ofrece MATLAB para trabajar con ellas y principalmente debido a los buenos resultados que ofrecen las Redes Neuronales Artificiales dentro del Reconocimiento Automático del Locutor (RAL). Además, se puede implementar con RNAs técnicas de RAL independientes del texto.

La elección de RNA implica decidir también las funciones de transferencia que se incluyen en las neuronas.

Como posibles alternativas para este problema se plantean las funciones de transferencia descritas en la figura 3.7, pero la elegida es la función sigmoide que es la que ofrece unos resultados mejores.

La elección de RNA también implica elegir la topología de la RNA que se va a implantar.

Las posibles soluciones a esta cuestión están descritas en el apartado 3.5.1 de esta memoria, siendo elegida como solución una RNA perceptrón multicapa con una capa oculta de 32 neuronas, una capa de entrada con tantas neuronas como parámetros de entrada y una de salida de 1 neurona. La elección de esta solución se debe al trabajo realizado en [Vivaracho, Cardeñoso y González, 2006].

La topología de la RNA implantada se puede ver en la siguiente figura:

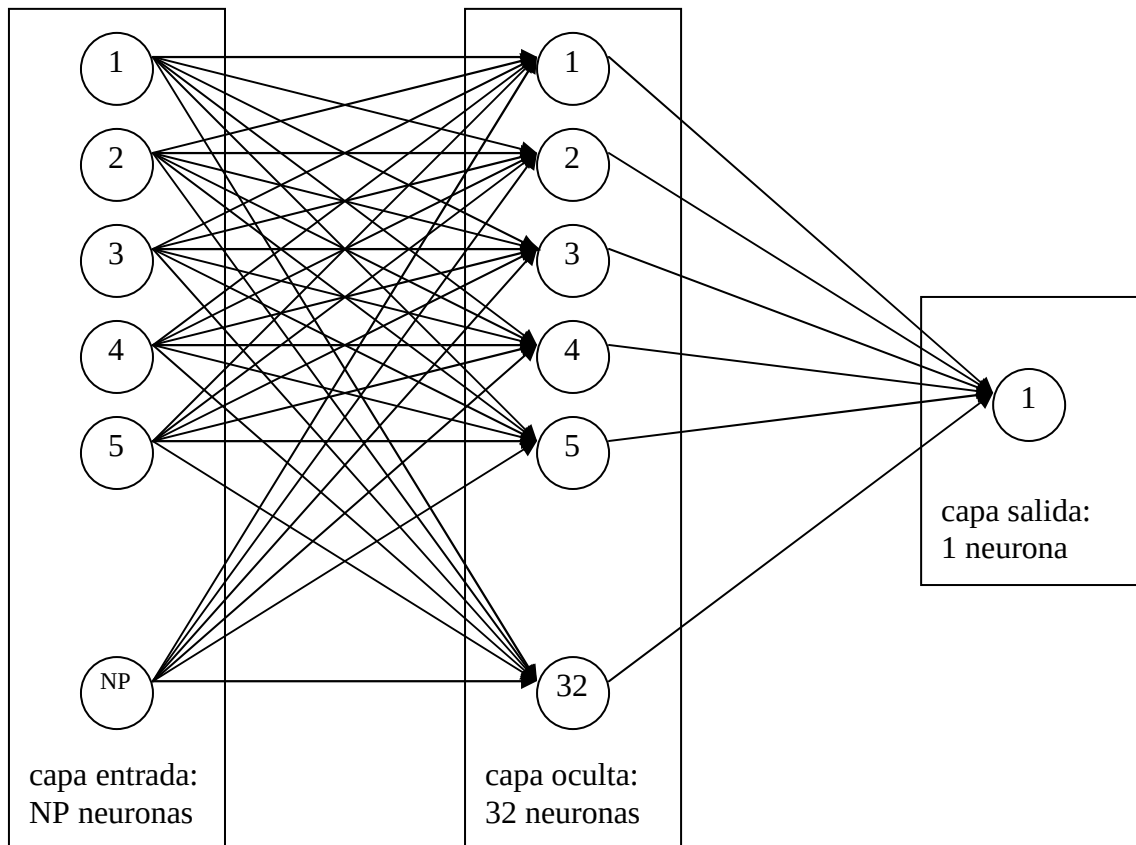


Figura 5.1 Topología de la RNA implantada

Como última cuestión a resolver dentro de la etapa de diseño se plantea la distribución de la información en directorios del sistema.

La solución tomada para este problema, es la de dividir el sistema en cuatro directorios, para agrupar la información por categorías de la siguiente manera:

- 1 directorio para los códigos fuente del sistema
- 1 directorio para los ficheros de audio del sistema
- 1 directorio para todos los datos que maneja el sistema (ya sean estadísticas, configuración, etc.).
- 1 directorio para los ficheros temporales del sistema.

5.3 Aspectos relevantes de la implementación

En la fase de implementación, también se plantea un problema a cerca de la distribución de directorios.

En este caso el problema es como distribuir los ficheros de audio dentro del directorio creado para tal efecto.

La solución tomada consiste en crear dentro del directorio de audio, tantos directorios como locutores tenga el sistema, (el nombre de estos directorios será el mismo que el del locutor correspondiente) y guardar los ficheros de audio de cada locutor dentro de su directorio.

Esta solución es elegida porque permite la selección automatizada de los ficheros de audio de cada locutor para poder realizar una prueba automática del sistema.

6. Trabajos relacionados

Dentro del apartado de trabajos se puede destacar el trabajo hecho por D. Carlos E. Vivaracho Pascual, D. Valentín Cardeñoso Payo y D. Cesar González Ferreras publicado en el artículo con título “Mejoras en un Sistema de Reconocimiento de Locutor Basado en RNA Mediante Entrenamiento con Normalización de Canal”.

Este trabajo resulta interesante para la realización del proyecto, en tanto en cuanto, se emplean Redes Neuronales Artificiales para la tarea de reconocimiento automático de locutor.

En este trabajo se plantean mejoras a que se pueden aplicar a una Red Neuronal Artificial para mejorar el rendimiento de esta en el reconocimiento automático de locutor.

De entre las mejoras propuestas en este trabajo se pueden mencionar:

- o Uso de una RNA con 32 neuronas en la capa oculta (arquitectura utilizada también en este proyecto).
- o Aplicación del logaritmo a la salida de la Red Neuronal Artificial.
- o Normalizar el canal a través del cual se adquieren las muestras de voz empleadas para el entrenamiento

Estas mejoras, junto con otras propuestas en el artículo, permiten mejorar la eficiencia del uso de las Redes Neuronales Artificiales en tareas de reconocimiento automático de locutor.

7. Conclusiones y líneas de trabajo futuras

7.1 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 utilizada para tareas de reconocimiento automático de locutor, a la vez que se ha conseguido un software que implementa una de ellas.

7.2 Líneas de trabajo futuras

SERAL no es un sistema cerrado, sino que está abierto a ampliaciones y mejoras, para conseguir unos resultados mejores.

De todas las ampliaciones que se pueden realizar sobre SERAL, cabe destacar el interés de las siguientes:

1. Incorporación de diferentes algoritmos para la detección de inicio y fin de palabras. Dentro de este punto, puede ser interesante el desarrollo de un nuevo algoritmo
2. Incorporación de más parámetros a parte de los coeficientes de predicción lineal, la transformada rápida de Fourier, la tasa de cruces por cero, la energía por bandas de frecuencia y los parámetros delta.
3. Incorporación de otras técnicas para la construcción del patrón de voz de los locutores.
Cabe destacar la importancia que, a parte de las Redes Neuronales Artificiales (RNA), tienen los Modelos Ocultos de Markov (MOM) dentro del Reconocimiento Automático del Locutor (RAL), aunque se pueden incorporar técnicas más sencillas como el Análisis Dinámico Temporal (ADT) que se aplica en situaciones de dependencia del texto.
4. Incorporación de nuevas topologías de Redes Neuronales Artificiales (RNA), como las descritas en el apartado 3.5.1 de esta memoria
5. Incorporación de de entrenamientos específicos para tramas etiquetadas de una cierta forma.

UNIVERSIDAD DE SALAMANCA

Escuela Politécnica Superior de Zamora

Ingeniería Técnica Informática de Gestión

Departamento de Informática y Automática, Área de Lenguajes y sistemas informáticos



Anexo I: Plan del proyecto software

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

Índice Anexo 1: Plan del proyecto software

Lista de tablas	46
Lista de figuras	47
A1.1 Planificación temporal del proyecto	49
A1.1.1 Tabla de precedencias	49
A1.1.2 Diseño del Grafo PERT	50
A1.1.3 Calculo de los tiempos	51
A1.1.4 Calculo de los caminos críticos	52
A1.2 Estudio de viabilidad del proyecto	53
A1.2.1 Determinación de gastos y beneficios	53
A1.2.2 Determinación de la rentabilidad	55
A1.2.3 Conclusiones	57
Acrónimos	58

Lista de tablas

Tabla A1.1. Tabla de precedencias de las tareas del proyecto.	49
Tabla A1.2. Resumen de los gastos e ingresos del proyecto	54

Lista de figuras

Figura A1.1. Grafo PERT del proyecto	50
Figura A1.2. Grafo PERT del proyecto con tiempos	51
Figura A1.3. Camino crítico del proyecto	52
Figura A1.4. Ingresos, gastos y beneficio acumulado del proyecto	55

A1.1 Planificación temporal del proyecto

El objetivo de este punto de la memoria del proyecto es el de determinar un calendario para la realización del mismo, y para lograrlo se utilizará la técnica de planificación temporal de los grafos PERT (Program Evaluation and Review Technique).

La elección de los grafos PERT como técnica para determinar el calendario del proyecto se debe a que dicha técnica permite ver las dependencias entre las distintas tareas que se planifican y además permite determinar el camino crítico (secuencia de tareas más largas, que determinan la duración total del proyecto).

Los pasos que implica dicha técnica son:

1. Realizar una tabla de precedencia
2. Dibujar el grafo correspondiente.
3. Calcular los tiempos más tempranos y más tardíos.
4. Calcular las holguras y caminos críticos.

A1.1.1 Tabla de precedencias

El contenido que aparecerá en la tabla de precedencias es:

1. Numeración de todas las tareas que están implicadas en la realización del proyecto.
2. Duración total de las tareas
3. Tareas que tienen que estar terminadas antes de poder empezar una nueva.

Esta información se muestra en la tabla siguiente:

Precedente	Tarea	Duración (días)	Descripción
-	A	2	Especificación de los requisitos iniciales
A	B	10	Estudio previo sobre la materia
A	C	4	Especificación de los requisitos finales
C	D	1	Análisis del hardware y software necesarios
D	E	5	Adquisición e instalación del hardware y software
B, C	F	15	Diseño del sistema
E, F	G	35	Programación del sistema
G	H	10	Realización de las pruebas
A	I	40	Realización de la documentación
I, H	J	5	Revisión de la documentación
J	K	4	Realización de la presentación

Tabla A1.1. Tabla de precedencias de las tareas del proyecto

A1.1.2 Diseño del Grafo PERT

Con la información contenida en la tabla de precedencias se diseña un grafo PERT, que muestre las relaciones existentes entre las distintas tareas que componen el proyecto, como el que se muestra en la figura.

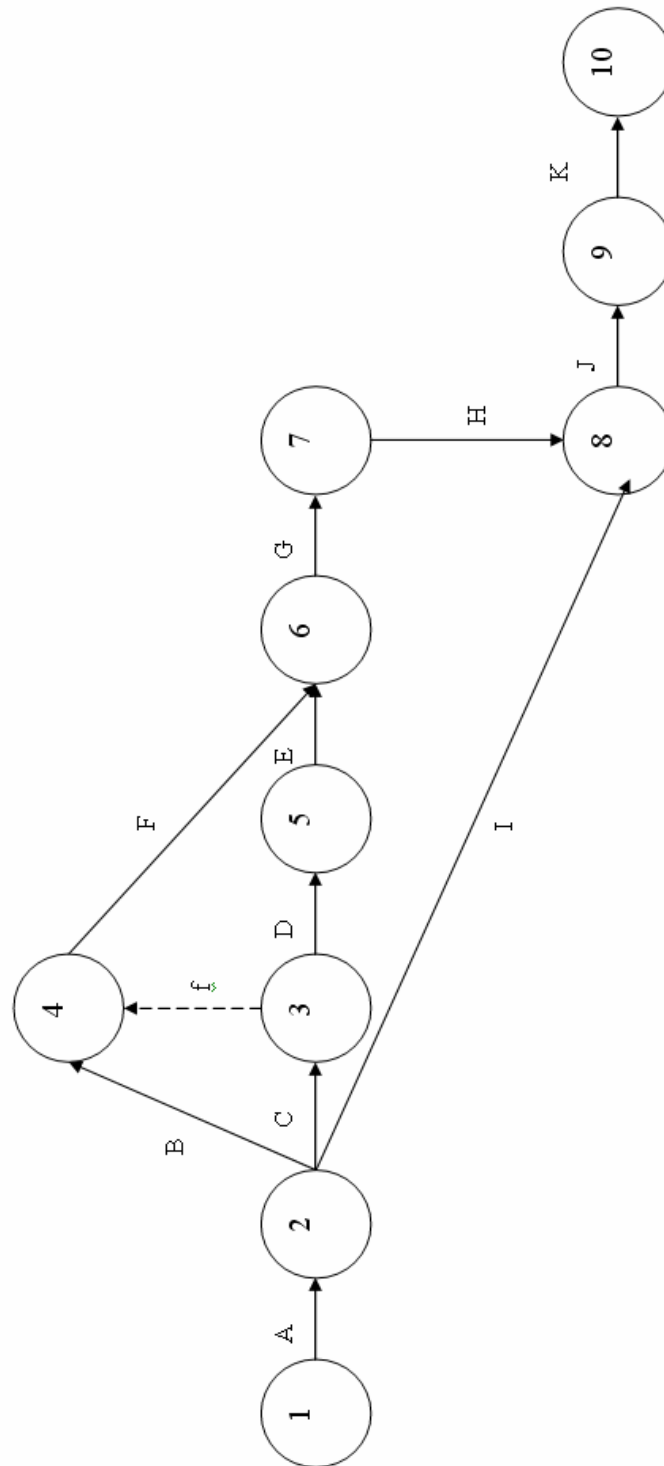


Figura A1.1. Grafo PERT del proyecto

A1.1.3 Calculo de los tiempos

A partir del grafo PERT de la figura A1.1 y de los tiempos de las tareas definidos en la tabla A1.1 se determinan los tiempos más tempranos y tardíos de las tareas. Esta información se muestra en la siguiente figura:

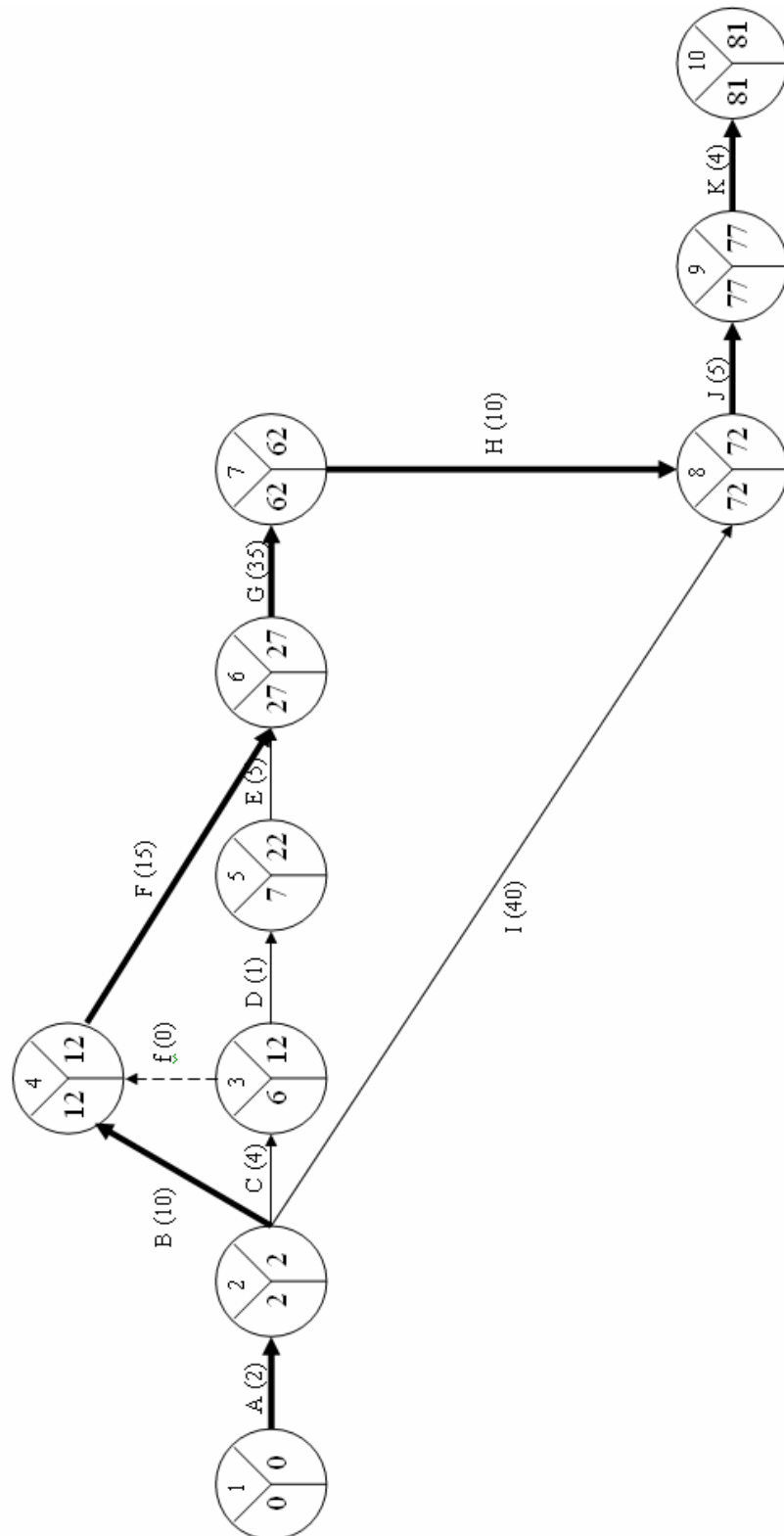


Figura A1.2. Grafo PERT del proyecto con tiempos

A1.1.4 Calculo de los caminos críticos

A partir de la información contenida en la figura A1.2 se determinan los caminos críticos (en el grafo de la figura A1.2 aparece marcado en negrita), formados por las tareas con holgura igual a cero.

Para calcular los caminos críticos del proyecto, se determinan las holguras de todas sus actividades:

$$H_A = 2 - 0 - 2 = 0 \text{ (Tarea crítica)}$$

$$H_B = 12 - 2 - 10 = 0 \text{ (Tarea crítica)}$$

$$H_C = 12 - 2 - 4 = 6$$

$$H_D = 22 - 6 - 1 = 15$$

$$H_E = 27 - 7 - 5 = 15$$

$$H_F = 27 - 12 - 15 = 0 \text{ (Tarea crítica)}$$

$$H_G = 62 - 27 - 35 = 0 \text{ (Tarea crítica)}$$

$$H_H = 72 - 62 - 10 = 0 \text{ (Tarea crítica)}$$

$$H_I = 72 - 2 - 40 = 30$$

$$H_J = 77 - 72 - 5 = 0 \text{ (Tarea crítica)}$$

$$H_K = 81 - 77 - 4 = 0 \text{ (Tarea crítica)}$$

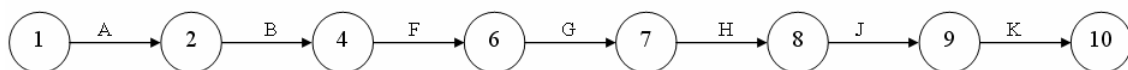


Figura A1.3. Camino crítico del proyecto

De esto se deduce que las tareas que determinan la duración del proyecto son:

- Especificación de los requisitos iniciales
- Estudio previo sobre la materia
- Diseño del sistema
- Programación del sistema
- Realización de las pruebas
- Revisión de la documentación
- Realización de la presentación

A1.2 Estudio de viabilidad del proyecto

El fin de este punto en la memoria del proyecto es el de determinar si el proyecto es viable o no, mediante la realización de un estudio económico que determine su rentabilidad.

A1.2.1 Determinación de gastos y beneficios

Para la realización del estudio económico, en primer lugar se determinarán los gastos y beneficios que conlleva la realización del proyecto.

En la siguiente lista se detallan todos los gastos y beneficios del proyecto:

Gastos:

1. Personal

De acuerdo con la planificación el proyecto será desarrollado en 81 días (tres meses) por dos personas (1500€ / mes)

Coste = 9000€

2. Formación del personal

La formación del personal necesaria para la realización del proyecto está valorada en 200€

Coste = 200€

3. Hardware

Para la realización del proyecto se necesitara la adquisición de un PC (1000€) y un micrófono (50€)

Coste = 1050€

4. Software

La licencia del programa necesario para desarrollar el sistema (MATLAB) vale 300€

Coste = 300€

NOTA: 300€ es el coste de la licencia para estudiante.

5. Supervisión y mantenimiento:

La supervisión y mantenimiento del sistema está valorado en 1000€

Coste = 1000€

NOTA: Los dos primeros años los cubre la garantía.

6. Material de fungible

El material empleado para el desarrollo del software está valorado en 200€

Coste = 200€

Beneficios:

1. Personal

Disminución del personal de seguridad valorado en 6000€

Beneficio = 6000€

NOTA: el control del sistema se llevará a cabo por una persona desde un ordenador central.

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

De forma resumida podemos ver los gastos y beneficios del proyecto en la tabla A1.2, donde también se incluye el beneficio de cada año y el beneficio total acumulado durante los 5 años de vida del sistema (tiempo en el que se amortizan los bienes informáticos).

	Año 1	Año 2	Año 3	Año 4	Año 5
Personal	6000€	6000€	6000€	6000€	6000€
Total ingresos	6000€	6000€	6000€	6000€	6000€
Personal	9000€	0€	0€	0€	0€
Formación del personal	200€	0€	0€	0€	0€
Hardware	1050€	0€	0€	0€	0€
Software	300€	0€	0€	0€	0€
Mantenimiento	0€	0€	1000€	1000€	1000€
Material de oficina	200€	0€	0€	0€	0€
Total gastos	10750€	0€	1000€	1000€	1000€
Beneficio neto anual	-4750€	6000€	5000€	5000€	5000€
Beneficio acumulado	-4750€	1250€	6250€	11250€	16250€

Tabla A1.2. Resumen de los gastos e ingresos del proyecto

A1.2.2 Determinación de la rentabilidad

Para determinar la rentabilidad que supone el proyecto se emplearán los criterios del PAY-BACK, VAN (Valor Actual Neto) y TIR (Tasa Interna de Rentabilidad).

PAY-BACK

Este criterio permite conocer el plazo de recuperación de la inversión mediante la fórmula:

$$PAY - BACK = \frac{A}{C}$$

donde: A = desembolso inicial

C = flujo de caja de cada año

El proyecto se considera rentable si la inversión se recupera.



$$PAY - BACK = 1 \text{ año} + \frac{4750}{5000} = 1 \text{ año} + 0.95$$

La inversión se recuperará en 1 año, 11 meses y 12 días.

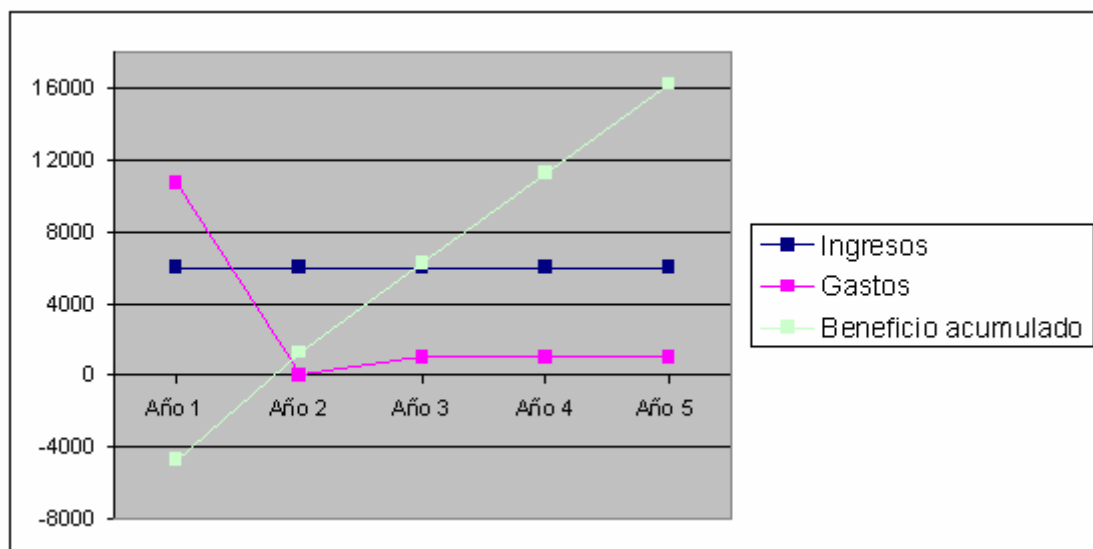


Figura A1.4. Ingresos, gastos y beneficio acumulado del proyecto

VAN

Este criterio permite calcular si un proyecto es rentable un mediante la fórmula:

$$VAN = -A + \frac{C_1}{(1+k_1)} + \frac{C_2}{(1+k_1)(1+k_2)} + \dots + \frac{C_n}{(1+k_1)(1+k_2) \dots (1+k_n)}$$

donde: A = desembolso inicial
C = flujo de caja de cada año
n = nº de años del proyecto
k = tasa de actualización

Si el VAN es mayor que 0 el proyecto es rentable.



Tasa de actualización $k = 0.0375$

$$VAN = -10750 + \frac{6000}{1.0375} + \frac{5000}{1.0375^2} + \frac{5000}{1.0375^3} + \frac{5000}{1.0375^4} + \frac{5000}{1.0375^5} = 12630.2$$

$VAN = 12630.2 \Rightarrow$ El proyecto es rentable.

TIR

Este criterio permite una vez determinado que el proyecto es rentable (mediante el criterio del VAN), calcular la rentabilidad de un proyecto mediante la fórmula:

$$0 = -A + \frac{C_1}{(1+k)} + \frac{C_2}{(1+k)^2} + \dots + \frac{C_n}{(1+k)^n}$$

donde: A = desembolso inicial
C = flujo de caja de cada año
n = nº de años del proyecto
k = rentabilidad



$$0 = -10750 + \frac{6000}{(1+k)} + \frac{5000}{(1+k)^2} + \frac{5000}{(1+k)^3} + \frac{5000}{(1+k)^4} + \frac{5000}{(1+k)^5} \Rightarrow k = 0.94985$$

A1.2.3 Conclusiones

Con los resultados obtenidos del estudio anterior, se puede concluir que el proyecto es rentable, y por lo tanto es factible su realización.

La viabilidad del proyecto esta garantizada.

Acrónimos

PERT	Program Evaluation and Review Technique
VAN	Valor Actual Neto
TIR	Tasa Interna de Rentabilidad

UNIVERSIDAD DE SALAMANCA

Escuela Politécnica Superior de Zamora

Ingeniería Técnica Informática de Gestión

Departamento de Informática y Automática, Área de Lenguajes y sistemas informáticos



Anexo II: Especificación de requisitos del software

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

Índice Anexo 2: Especificación de requisitos del software

Lista de tablas	64
Lista de figuras	65
Lista de cambios	67
A2.1 Introducción	68
A2.2 Lista de usuarios participantes	68
A2.3 Catálogo de requisitos del sistema	69
A2.4 Especificación de los requisitos.	70
A2.4.1 Requisitos funcionales	70
A2.4.2 Requisitos no funcionales	76
A2.4.3 Requisitos de información	78
A2.5 Interfaces de usuario	80
A2.5.1 Menú Archivo.	81
A2.5.2 Menú Operaciones	82
A2.5.3 Menú Estadística.	83
A2.5.4 Menú Ayuda.	83
A2.6 Pruebas	84
Acrónimos	85

Lista de tablas

Tabla A2.1. Lista de cambios	67
Tabla A2.2. RF-1: Entrenamiento del sistema	70
Tabla A2.3. RF-2: Verificación de locutor.	71
Tabla A2.4. RF-3: Identificación de locutor.	71
Tabla A2.5. RF-4: Eliminación de locutor	72
Tabla A2.6. RF-5: Ver estadísticas.	72
Tabla A2.7. RF-6: Borrar estadísticas	73
Tabla A2.8. RF-7: Grabar voz	73
Tabla A2.9. RF-8: Reproducir voz.	74
Tabla A2.10. RF-9: Visualizar señal de voz.	74
Tabla A2.11. RF-10: Cambiar la configuración.	75
Tabla A2.12. RNF-1: Bajo consumo de recursos	76
Tabla A2.13. RNF-2: Interfaz de usuario basada en ventanas.	76
Tabla A2.14. RNF-3: Existencia de menús	76
Tabla A2.15. RNF-4: Sistema operativo	77
Tabla A2.16. RNF-5: Documentación	77
Tabla A2.17. RI-1: Fichero de estadísticas	78
Tabla A2.18. RI-2: Fichero de configuración	78
Tabla A2.19. RI-3: Fichero de patrones de voz	78
Tabla A2.20. RI-4: Fichero de etiquetas.	79

Lista de figuras

Figura A2.1. Apariencia de la ventana inicial de la aplicación	81
Figura A2.2. Menú “Archivo”	81
Figura A2.3. Menú “Operaciones”	82
Figura A2.4. Menú “Estadísticas”	83
Figura A2.5. Menú “Ayuda”	83

Lista de cambios

Número	Fecha	Descripción	Autores
1	14-12-2006	Especificación de las primeras tareas que realizará el sistema como son las de entrenamiento, reconocimiento e identificación de locutor.	Hernando Silva Varela y Manuel Montero Villar
2	20-12-2006	Determinación de los métodos utilizados para realizar las tareas de reconocimiento e identificación	Hernando Silva Varela y Manuel Montero Villar
3	10-1-2007	Determinación de las características de la interfaz de usuario del sistema.	Hernando Silva Varela y Manuel Montero Villar
4	11-1-2007	Incorporación de las tareas de grabación y reproducción de señales de voz a las funciones del sistema	Hernando Silva Varela y Manuel Montero Villar
5	16-1-2007	Introducción de un osciloscopio para visualizar la señal de voz con la que se está trabajando como parte de la interfaz de usuario.	Hernando Silva Varela y Manuel Montero Villar
6	18-1-2007	Introducción de un fichero de estadísticas para controlar el funcionamiento del sistema.	Hernando Silva Varela y Manuel Montero Villar
7	23-1-2007	Incorporación de un fichero de configuración, para que esta se pueda cambiar de una ejecución a otra.	Hernando Silva Varela y Manuel Montero Villar
8	1-2-2007	Selección de MatLab como entorno de desarrollo.	Jaime Calvo Gallego, Hernando Silva Varela y Manuel Montero Villar
9	20-2-2007	Introducción de la función de zoom.	Hernando Silva Varela y Manuel Montero Villar
10	18-3-2007	Incorporación de botones en la parte superior de la aplicación para acelerar el acceso a determinadas funciones.	Hernando Silva Varela y Manuel Montero Villar
11	10-4-2007	División del osciloscopio en partes	Hernando Silva Varela y Manuel Montero Villar
12	5-6-2007	Incorporación del etiquetado	Hernando Silva Varela y Manuel Montero Villar

Tabla A2.1. Lista de cambios

A2.1 Introducción

El cumplimiento de los requisitos impuestos por el cliente es una parte esencial y muy a tener en cuenta a la hora de desarrollar un sistema.

El cumplimiento o no de estos requisitos determinará el grado de satisfacción del cliente con el sistema que ha sido desarrollado.

En este anexo se detallarán todos los requisitos impuestos sobre el sistema que se desarrolla en este proyecto fin de carrera con los que se pretende cumplir los objetivos marcados en el punto 2 de esta memoria, para ello se estructurará dicho anexo en los siguientes puntos:

- Lista de usuarios participantes.
- Catálogo de requisitos del sistema.
- Especificación de requisitos.
- Interfaces de usuario.
- Pruebas.

A2.2 Lista de usuarios participantes

Las personas involucradas en el análisis de requisitos son:

1. D. Hernando Silva Varela, tutor de este proyecto, y profesor del departamento de informática y automática de la universidad de Salamanca.
2. D. Jaime Calvo Gallego, profesor del departamento de informática y automática de la universidad de Salamanca.
3. D. Manuel Montero Villar, autor de este proyecto.

A2.3 Catálogo de requisitos del sistema

En este punto se expondrá una lista con los requisitos impuestos sobre el sistema, así como una descripción de los mismos.

- El sistema tendrá una opción que permita realizar un entrenamiento del mismo, es decir, esta opción permitirá crear los patrones de referencia de los distintos usuarios que el sistema tenga que reconocer.
- El sistema permitirá dos modos de funcionamiento, por un lado permitirá hacer una verificación de locutor, es decir, el sistema pedirá una muestra de voz y una identificación y las validará, y por otro lado se permitirá una identificación de locutor, en el que a partir de una muestra de voz, el sistema reconocerá al locutor.
- Para realizar las tareas de verificación e identificación de locutor se emplearán la técnica de las Redes Neuronales Artificiales (RNA).
- El sistema también permitirá visualizar las estadísticas de su utilización, esto consistirá en proporcionar la información del número de veces que la verificación o identificación del locutor se han realizado con éxito y el número de veces que el sistema ha rechazado al locutor.
- Otra funcionalidad requerida al sistema es la de permitir operaciones de grabación y reproducción de señales de voz.
- El sistema mostrará en un osciloscopio la señal de voz con la que se está trabajando, permitiendo realizar un zoom sobre él.
- Los patrones de voz de los distintos usuarios serán almacenados todos juntos en un único fichero.
- Existirá un fichero de configuración de donde el sistema cargará su configuración inicial, así como un fichero con las estadísticas de uso del sistema.
- Se podrá crear un fichero con etiquetas para cada fichero de voz.

A2.4 Especificación de los requisitos

El objetivo de este punto es el de especificar de manera detallada tanto los requisitos funcionales como los no funcionales y los requisitos de información que debe de cumplir el sistema desarrollado en este proyecto.

A2.4.1 Requisitos funcionales

Este apartado contiene una lista de tablas donde se describen cada uno de los requisitos funcionales que debe satisfacer el sistema.

RF-1	Entrenamiento del sistema	
Versión	1.0	
Autores	Hernando Silva Varela Manuel Montero Villar	
Objetivos	Construir los patrones de voz de los usuarios que tienen que ser reconocidos por el sistema	
Descripción	Se tiene que proporcionar al sistema las muestras de voz de cada usuario necesarias para que este pueda construir los patrones de voz asociados a cada usuario para un correcto funcionamiento	
Precondición	El software tiene que haber arrancado correctamente	
Secuencia normal	Paso	Acción
	1	El usuario introduce las muestras de voz necesarias
	2	El sistema procesa las muestras de voz y extrae de cada una los parámetros característicos
	3	A partir de los parámetros de las muestras se construye el patrón de voz
	4	El patrón es almacenado en un fichero con los patrones de los demás usuarios
Frecuencia	Cada vez que se introduzca un nuevo usuario en el sistema	
Importancia	Crítica	
Estado	Validado	
Comentarios	El cumplimiento de este requisito es vital para el correcto funcionamiento del sistema.	

Tabla A2.2. RF-1: Entrenamiento del sistema

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

RF-2	Verificación de locutor	
Versión	1.0	
Autores	Hernando Silva Varela Manuel Montero Villar	
Objetivos	Constatar que un usuario es quién dice ser	
Descripción	El sistema recoge una muestra de voz del locutor así como su identidad y comprueba que la identidad dada coincide con la del propietario de la muestra de voz	
Precondición	El locutor tiene que tener su patrón de voz guardado en el sistema	
Secuencia normal	Paso	Acción
	1	El usuario introduce su muestra de voz y su identificación
	2	El sistema construye un patrón de voz a partir de la muestra de voz obtenida del usuario
	3	El sistema extrae el patrón de voz correspondiente a la identificación dada
	4	El sistema compara ambos patrones y acepta o rechaza al usuario
Frecuencia	Cada vez que se verifique a un locutor	
Importancia	Alta	
Estado	Validado	
Comentarios		

Tabla A2.3. RF-2: Verificación de locutor

RF-3	Identificación de locutor	
Versión	1.0	
Autores	Hernando Silva Varela Manuel Montero Villar	
Objetivos	Determinar quién es el usuario que habla al sistema	
Descripción	El sistema recoge una muestra de voz del locutor y comprueba si es uno de los usuarios reconocidos por el sistema	
Precondición	El locutor tiene que tener su patrón de voz guardado en el sistema	
Secuencia normal	Paso	Acción
	1	El usuario introduce su muestra de voz
	2	El sistema construye un patrón de voz a partir de la muestra de voz obtenida del usuario
	3	El sistema compara este patrón con todos los almacenados en el sistema
	4	Si existe coincidencia con alguno de los patrones almacenados, el usuario es aceptado, de lo contrario, el usuario es rechazado
Frecuencia	Cada vez que se identifique a un locutor	
Importancia	Alta	
Estado	Validado	
Comentarios		

Tabla A2.4. RF-3: Identificación de locutor

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

RF-4	Eliminación de locutor	
Versión	1.0	
Autores	Hernando Silva Varela Manuel Montero Villar	
Objetivos	Eliminar del sistema un usuario que deja de serlo	
Descripción	El sistema recoge el nombre del usuario y si es uno de los que el sistema tiene reconocido lo elimina.	
Precondición	El usuario tiene que ser válido en el sistema	
Secuencia normal	Paso	Acción
	1	Se introduce el nombre del usuario
	2	El sistema compara el nombre con el de todos los usuarios reconocidos
	3	Si se encuentra el usuario introducido se elimina
Frecuencia	Cada vez que se elimine a un locutor	
Importancia	Alta	
Estado	Validado	
Comentarios		

Tabla A2.5. RF-4: Eliminación de locutor

RF-5	Ver estadísticas	
Versión	1.0	
Autores	Hernando Silva Varela Manuel Montero Villar	
Objetivos	Observar las estadísticas de utilización del sistema	
Descripción	El sistema proporciona información acerca del número de veces que la verificación o identificación del locutor se han realizado con éxito y el número de veces que el sistema ha rechazado al locutor.	
Precondición	El sistema ha tenido que se utilizado al menos una vez	
Secuencia normal	Paso	Acción
	1	El sistema abre el fichero de estadísticas
	2	Se muestra el resultado al usuario
Frecuencia	Cada vez que se pidan las estadísticas	
Importancia	Media	
Estado	Validado	
Comentarios		

Tabla A2.6. RF-5: Ver estadísticas

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

RF-6	Borrar estadísticas	
Versión	1.0	
Autores	Hernando Silva Varela Manuel Montero Villar	
Objetivos	Borrar las estadísticas de utilización del sistema	
Descripción	El sistema borra la información acerca del número de veces que la verificación o identificación del locutor se han realizado con éxito y el número de veces que el sistema ha rechazado al locutor.	
Precondición	El sistema ha tenido que se utilizado al menos una vez	
Secuencia normal	Paso	Acción
	1	El sistema resetea el contenido del fichero de estadísticas
Frecuencia	Cada vez que se borren las estadísticas	
Importancia	Media	
Estado	Validado	
Comentarios		

Tabla A2.7. RF-6: Borrar estadísticas

RF-7	Grabar voz	
Versión	1.0	
Autores	Hernando Silva Varela Manuel Montero Villar	
Objetivos	Grabar un fichero de audio con la voz de un usuario	
Descripción	El sistema permite a un usuario grabar su voz en un fichero de audio	
Precondición	Ninguna	
Secuencia normal	Paso	Acción
	1	El usuario introduce una muestra de voz en el sistema
	2	El usuario elige el fichero donde se guardará su voz
	3	El sistema graba el fichero con la voz del usuario
Frecuencia	Cada vez que grabe un fichero de voz.	
Importancia	Media	
Estado	Validado	
Comentarios		

Tabla A2.8. RF-7: Grabar voz

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

RF-8	Reproducir voz	
Versión	1.0	
Autores	Hernando Silva Varela Manuel Montero Villar	
Objetivos	Reproducir una señal de voz	
Descripción	El sistema permite a un usuario reproducir la señal de voz con la que esta trabajando a través de unos altavoces	
Precondición	Ninguna	
Secuencia normal	Paso	Acción
	1	El usuario introduce el nombre de un fichero de voz
	2	El usuario selecciona la opción de reproducir
	3	El sistema reproduce la voz a través de unos altavoces
Frecuencia	Cada vez que se reproduzca una señal de voz	
Importancia	Media	
Estado	Validado	
Comentarios		

Tabla A2.9. RF-8: Reproducir voz

RF-9	Visualizar una señal de voz	
Versión	1.0	
Autores	Hernando Silva Varela Manuel Montero Villar	
Objetivos	Observar una señal de voz	
Descripción	El sistema permite a un usuario ver la señal de voz con la que esta trabajando.	
Precondición	Ninguna	
Secuencia normal	Paso	Acción
	1	El sistema reproduce la señal de voz con la que está trabajando en un osciloscopio en la pantalla automáticamente.
Frecuencia	Siempre que el sistema trabaje con una señal de voz	
Importancia	Media	
Estado	Validado	
Comentarios		

Tabla A2.10. RF-9: Visualizar señal de voz

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

RF-10	Cambiar la configuración	
Versión	1.0	
Autores	Hernando Silva Varela Manuel Montero Villar	
Objetivos	Cambiar la configuración de trabajo del sistema	
Descripción	El sistema debe permitir al usuario cambiar alguno de los parámetros de su configuración que están almacenados en un fichero de configuración.	
Precondición	Ninguna	
Secuencia normal	Paso	Acción
	1	El sistema abre el fichero de configuración
	2	El usuario introduce y guarda la nueva configuración
Frecuencia	Cada vez que se cambie la configuración del sistema	
Importancia	Media	
Estado	Validado	
Comentarios		

Tabla A2.11. RF-10: Cambiar la configuración

A2.4.2 Requisitos no funcionales

Este apartado contiene una lista de tablas donde se describen cada uno de los requisitos no funcionales que debe satisfacer el sistema.

RNF-1	Bajo consumo de recursos
Versión	1.0
Autores	Hernando Silva Varela Manuel Montero Villar
Objetivos	Obtener un alto rendimiento de la máquina donde este funcionando la aplicación desarrollada
Descripción	La aplicación debe de funcionar con un bajo consumo de CPU, memoria y disco.
Importancia	Alta
Estado	Validado
Comentarios	

Tabla A2.12. RNF-1: Bajo consumo de recursos

RNF-2	Interfaz de usuario basada en ventanas
Versión	1.0
Autores	Hernando Silva Varela Manuel Montero Villar
Objetivos	Mayor facilidad en el uso de la aplicación
Descripción	La interfaz de la aplicación está basada en un sistema de ventanas.
Importancia	Alta
Estado	Validado
Comentarios	

Tabla A2.13. RNF-2: Interfaz de usuario basada en ventanas

RNF-3	Existencia de menús
Versión	1.0
Autores	Hernando Silva Varela Manuel Montero Villar
Objetivos	Mayor facilidad en el uso de la aplicación
Descripción	La aplicación debe proporcionar un menú para acceder a todas sus operaciones
Importancia	Alta
Estado	Validado
Comentarios	

Tabla A2.14. RNF-3: Existencia de menús

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

RNF-4	Sistema operativo
Versión	1.0
Autores	Hernando Silva Varela Manuel Montero Villar
Objetivos	Determinar el entorno en el que funcionará el sistema
Descripción	La aplicación debe de funcionar bajo un sistema WINDOWS.
Importancia	Alta
Estado	Validado
Comentarios	

Tabla A2.15. RNF-4: Sistema operativo

RNF-5	Documentación
Versión	1.0
Autores	Hernando Silva Varela Manuel Montero Villar
Objetivos	Mayor facilidad en el uso de la aplicación
Descripción	Conjuntamente con el sistema será entregada toda la documentación generada durante el desarrollo del mismo.
Importancia	Alta
Estado	Validado
Comentarios	

Tabla A2.16. RNF-5: Documentación

A2.4.3 Requisitos de información

Este apartado contiene una lista de tablas donde se describen cada uno de los requisitos de información que debe satisfacer el sistema.

RI-1	Fichero de estadísticas
Versión	1.0
Autores	Hernando Silva Varela Manuel Montero Villar
Objetivos	Almacenamiento de las estadísticas del sistema
Descripción	Todas las estadísticas de utilización del sistema se almacenarán en un único fichero, que contendrá datos acerca del número de veces en el que se identificó el locutor y el número de veces en el que no se identificó.
Importancia	Media
Estado	Validado
Comentarios	

Tabla A2.17. RI-1: Fichero de estadísticas

RI-2	Fichero de configuración
Versión	1.0
Autores	Hernando Silva Varela Manuel Montero Villar
Objetivos	Almacenamiento de la configuración del sistema
Descripción	Toda la configuración del sistema será almacenada en un único fichero que contendrá información acerca de cómo se procesará la señal de voz para obtener el patrón de voz.
Importancia	Alta
Estado	Validado
Comentarios	

Tabla A2.18. RI-2: Fichero de configuración

RI-3	Fichero de patrones de voz
Versión	1.0
Autores	Hernando Silva Varela Manuel Montero Villar
Objetivos	Almacenamiento de los patrones de voz de los usuarios
Descripción	Los patrones de voz de todos los usuarios reconocidos por el sistema serán almacenados en un único fichero.
Importancia	Alta
Estado	Validado
Comentarios	Este fichero no solo contiene los patrones de voz de los distintos usuarios del sistema, sino que también tiene una identificación de cada uno de ellos.

Tabla A2.19. RI-3: Fichero de patrones de voz

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

RI-4	Fichero de etiquetas
Versión	1.0
Autores	Hernando Silva Varela Manuel Montero Villar
Objetivos	Almacenamiento de las etiquetas de voz
Descripción	El sistema permitirá etiquetar las señales de voz, dichas etiquetas serán guardadas en ficheros con el mismo nombre que el fichero de voz y extensión .eti.
Importancia	Media
Estado	Validado
Comentarios	

Tabla A2.20. RI-4: Fichero de etiquetas

A2.5 Interfaces de usuario

La interfaz de usuario es clave para lograr que la aplicación que se está desarrollando sea fácil de usar e intuitiva, que son aspectos necesarios para que pueda ser utilizada por un usuario que no tiene grandes conocimientos en el tema sobre el que versa la aplicación.

Es por tanto la interfaz de usuario es uno de los aspectos más importantes dentro de la especificación de requisitos para conseguir la aceptación por parte del cliente, y por ello se le dedica un punto específico dentro de la especificación de requisitos.

En este apartado se exponen las especificaciones que debe cumplir la interfaz de usuario de la aplicación que se desarrolla en este proyecto.

La parte principal de la interfaz de usuario estará compuesta por una ventana, que contenga el osciloscopio (dividido en cuatro regiones) que permita ver la señal de voz con la que se está trabajando.

En la región superior del osciloscopio se muestra el espectrograma de la señal de voz, en las dos regiones intermedias se muestra la propia señal de voz, y en la región inferior, se muestran la tasa de cruces por cero de la señal de voz (línea azul), y la energía de la señal de voz (línea roja).

Además de esto, esta ventana tiene que permitir realizar todas las operaciones que se han descrito en el apartado A2.3 (Catálogo de requisitos del sistema) a través de unos menús situados en la parte superior de la misma.

Los menús que aparecerán tendrán los siguientes títulos:

- o Archivo.
- o Operaciones.
- o Estadísticas.
- o Ayuda.

Estos menús estarán situados inmediatamente debajo de la barra de título, que contendrá el nombre de la aplicación.

Debajo de los menús, existirán unos botones que dan un acceso rápido a alguna de las operaciones de los menús.

Cuando se esté trabajando con un archivo de audio, esta ventana mostrará información relativa al archivo, y a los parámetros empleados para el procesamiento de la señal. Esta información será presentada debajo del osciloscopio donde se visualiza la señal de voz.

La apariencia de la ventana será de forma aproximada la que se muestra en la figura siguiente:

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

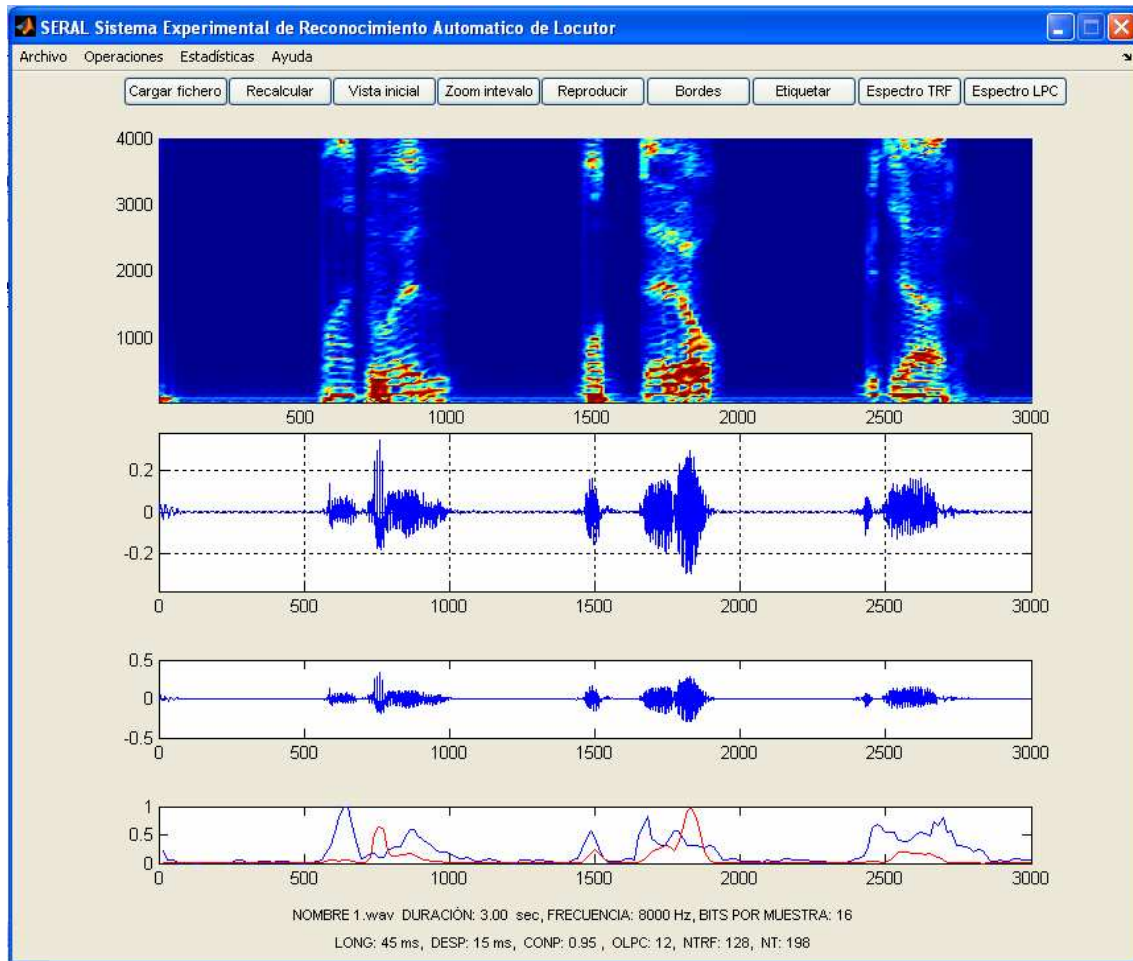


Figura A2.1. Apariencia de la ventana inicial de la aplicación

Una vez descrita la apariencia visual de la ventana inicial del sistema, se describirá el contenido de cada uno de los menús.

A2.5.1 Menú “Archivo”

El menú “Archivo” estará compuesto de dos opciones:

1. **Configurar:** Esta opción permitirá al usuario cambiar las opciones de configuración de la aplicación, que están almacenadas en un fichero y que se cargarán de manera automática cuando se inicia la aplicación.
2. **Abrir:** Esta opción da la posibilidad de reproducir un fichero de voz que se encuentre almacenado en el sistema.
3. **Salir:** Esta opción permite al usuario concluir la utilización de la aplicación cerrando la misma.

El aspecto visual del menú archivo será la siguiente:



Figura A2.2. Menú “Archivo”

A2.5.2 Menú “Operaciones”

El menú “Operaciones” tendrá cuatro opciones:

1. **Entrenamiento:** Esta opción permitirá realizar el entrenamiento de un nuevo usuario del sistema, de tal forma que se generará su patrón de voz y se almacenará en un fichero junto con el resto de patrones de voz de los diferentes usuarios del sistema.
2. **Reconocimiento:** Esta opción desplegará un submenú que contendrá a su vez dos nuevas opciones:
 - 2.1. **Verificación:** Esta opción permite realizar una verificación de locutor. Para poder realizar esta operación, previamente el usuario debe de haber entrenado el sistema para que este pueda generar su patrón de voz.
 - 2.2. **Identificación:** Esta opción permite realizar una identificación de locutor. Para poder realizar esta operación, previamente el usuario debe de haber entrenado el sistema para que este pueda generar su patrón de voz.
3. **Eliminar Locutor:** Esta opción permite eliminar un locutor que se encuentre registrado en el sistema.
4. **Grabación:** Esta opción da la posibilidad de grabar un fichero de voz, que posteriormente se podrá reproducir.
5. **Reproducir:** Esta opción da la posibilidad de reproducir un intervalo de la señal de voz que se ve en la pantalla seleccionando dos puntos de la misma con el ratón.
6. **Zoom:** Esta opción desplegará un submenú que contendrá a su vez tres nuevas opciones:
 - 6.1. **Acercar:** Esta opción permite visualizar en pantalla un 25% de la señal de voz alrededor del punto central de la señal de voz que se esté visualizando en ese momento.
 - 6.2. **Alejar:** Esta opción permite visualizar en pantalla un 200% de la señal de voz alrededor del punto central de la señal de voz que se esté visualizando en ese momento.
 - 6.3. **Intervalo:** Esta opción permite visualizar en pantalla la señal de voz comprendida entre dos puntos seleccionados con el ratón.
 - 6.4. **Vista inicial:** Esta opción permite visualizar en pantalla la señal de voz completa.

El aspecto visual del menú “Operaciones” será el siguiente:



Figura A2.3. Menú “Operaciones”

A2.5.3 Menú “Estadísticas”

El menú “Estadísticas” dispondrá de dos opciones diferentes:

1. **Ver:** Esta opción permite ver las estadísticas de uso de la aplicación, que se encuentran guardadas en un fichero de estadística.
2. **Borrar:** Esta opción permitirá borrar el contenido del fichero de estadísticas, borrando así todas las estadísticas de utilización de la aplicación hasta la fecha actual.

La apariencia visual del menú “Estadísticas” será la siguiente:

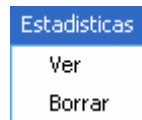


Figura A2.4. Menú “Estadísticas”

A2.5.4 Menú “Ayuda”

El menú “Ayuda” tendrá dos opciones:

1. **Ayuda del SERAL:** Esta opción permite al usuario ver la ayuda necesaria para hacer un correcto manejo de la aplicación, permitiendo al usuario resolver las dudas que tenga acerca del funcionamiento del sistema.
2. **Acerca de...:** Esta opción muestra la información concerniente a los desarrolladores de la aplicación.

El aspecto visual que tendrá el menú “Ayuda” será el siguiente:

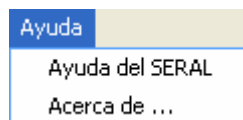


Figura A2.5. Menú “Ayuda”

A2.6 Pruebas

El objetivo de este punto es el de determinar las pruebas de aceptación que el sistema tiene que superar.

Para que el sistema sea aceptado debe satisfacer el siguiente conjunto de pruebas:

1. Prueba de cumplimiento de los requisitos

Esta prueba consiste en analizar de manera minuciosa el sistema para comprobar que cumple con todos los requisitos definidos en los apartados A2.3 y A2.4 del presente anexo.

2. Prueba de funcionamiento sin micrófono

Esta prueba consiste en comprobar que el sistema es capaz de realizar las tareas de verificación e identificación de locutor sin necesidad de un micrófono, basándose en ficheros de voz de los distintos usuarios que estén almacenados en el sistema.

3. Prueba de funcionamiento con distintas configuraciones

Esta prueba consiste en comprobar que el sistema es capaz de funcionar de manera adecuada modificando la configuración del fichero de configuración que el sistema tiene.

Acrónimos

SERAL	Sistema Experimental de Reconocimiento Automático de Locutor
RNA	Redes Neuronales Artificiales
RF	Requisito funcional
RNF	Requisito no funcional
RI	Requisito de información

UNIVERSIDAD DE SALAMANCA

Escuela Politécnica Superior de Zamora

Ingeniería Técnica Informática de Gestión

Departamento de Informática y Automática, Área de Lenguajes y sistemas informáticos



Anexo III: Especificación de Diseño

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

Índice Anexo 3: Especificación de diseño

Lista de tablas	91
Lista de figuras	92
A3.1 Introducción	94
A3.2 Ámbito del software.	94
A3.3 Diseño de datos	96
A3.4 Diseño arquitectónico	101
A3.5 Diseño de la interfaz	120
A3.6 Diseño procedimental	135
A3.7 Referencia cruzada a los requisitos	143
A3.8 Pruebas	146
A3.9 Entorno tecnológico del sistema	148
A3.10 Plan de desarrollo e implantación	149
Acrónimos	151

Lista de tablas

Tabla A3.1 controles de la ventana principal de SERAL	122
Tabla A3.2 controles del cuadro de diálogo de salida	123
Tabla A3.3 controles del cuadro de diálogo de abrir fichero	124
Tabla A3.4 controles del cuadro de diálogo de mostrar estadísticas.	125
Tabla A3.5 controles del cuadro de diálogo de borrar estadísticas	126
Tabla A3.6 controles del cuadro de diálogo de ayuda	127
Tabla A3.7 controles del cuadro de diálogo de reconocimiento.	128
Tabla A3.8 controles del cuadro de diálogo de mostrar resultados	129
Tabla A3.9 controles del cuadro de diálogo de nombrar locutor	130
Tabla A3.10 controles del cuadro de diálogo de inicio de grabación.	131
Tabla A3.11 controles del cuadro de diálogo de etiquetar	132
Tabla A3.12 controles del cuadro de diálogo de ver etiquetas	133
Tabla A3.13 controles del cuadro de diálogo de fichero de etiquetas.	134
Tabla A3.14 Referencia cruzada a los requisitos	144

Lista de figuras

Figura A3.1 estructura del fichero de configuración	96
Figura A3.2 estructura del fichero de estadísticas	97
Figura A3.3 estructura del fichero de patrones	97
Figura A3.4 estructura del fichero de voz	98
Figura A3.5 estructura del fichero de etiquetas.	99
Figura A3.6 estructura del vector endpoint	100
Figura A3.7 estructura de la matriz de parámetros	100
Figura A3.8 Diagrama de contexto	101
Figura A3.9 DFD 0 SERAL	103
Figura A3.10 DFD 1 Entrenar.	104
Figura A3.11 DFD 2 Procesar señal de voz	105
Figura A3.12 DFD 1.1 extraer parámetros	107
Figura A3.13 DFD 1.2 construir patrón de voz	108
Figura A3.14 DFD 1.3 almacenar patrón de voz	109
Figura A3.15 DFD 2.3 leer patrón de voz	110
Figura A3.16 DFD 2.4 comparar patrones de voz	111
Figura A3.17 DFD 3 eliminar locutor.	112
Figura A3.18 DFD 4 visualizar señal de voz	113
Figura A3.19 DFD 5 reproducir señal de voz	114
Figura A3.20 DFD 6 escribir fichero de voz	115
Figura A3.21 DFD 7 leer fichero de voz.	116
Figura A3.22 DFD 8 ver estadísticas	117
Figura A3.23 DFD 9 borrar estadísticas	118
Figura A3.24 DFD 10 actualizar configuración	119
Figura A3.25 Vista de la venta principal de SERAL	121
Figura A3.26 Menús de la ventana principal de SERAL	121
Figura A3.27 Vista del cuadro de diálogo de salida.	123
Figura A3.28 Vista del cuadro de diálogo de abrir fichero	124
Figura A3.29 Vista del cuadro de diálogo de mostrar estadísticas	125
Figura A3.30 Vista del cuadro de diálogo de borrar estadísticas.	126
Figura A3.31 Vista del cuadro de diálogo de ayuda	127
Figura A3.32 Vista del cuadro de diálogo de reconocimiento.	128
Figura A3.33 Vista del cuadro de diálogo de mostrar resultados.	129
Figura A3.34 Vista del cuadro de diálogo de nombrar locutor	130
Figura A3.35 Vista del cuadro de diálogo de inicio de grabación.	131
Figura A3.36 Vista del cuadro de diálogo de etiquetar	132
Figura A3.37 Vista del cuadro de diálogo de ver etiquetas	133
Figura A3.38 Vista del cuadro de diálogo de fichero de etiquetas.	134
Figura A3.39 Algoritmo ENDPOINT	136
Figura A3.40 Algoritmo ENDPOINT (continuación)	137
Figura A3.41 Algoritmo PARAMETROS.	138
Figura A3.42 Algoritmo ZOOMMAS	139
Figura A3.43 Algoritmo ZOOMMENOS	140
Figura A3.44 Algoritmo ZOOMINTERVALO.	141
Figura A3.45 Algoritmo REPRODUCEINTERVALO.	142
Figura A3.46 Árbol de directorios del sistema	149

A3.1 Introducción

El propósito de este anexo es el de determinar una solución que permita satisfacer los requisitos planteados en el anexo 2 (Especificación de requisitos del software).

El problema de determinar una solución no es sencillo, ya que dependiendo de la solución adoptada, la calidad final del sistema será mayor o menor, es por tanto objetivo de este anexo determinar una solución al problema que aporte la mayor calidad posible al sistema final.

En este anexo se detallará el entorno de funcionamiento del sistema, así como el diseño de todos los componentes del mismo (datos, interfaz, etc.), para ello se estructurará dicho anexo en los siguientes puntos:

- Ámbito del software
- Diseño de datos
- Diseño arquitectónico
- Diseño de la interfaz
- Diseño procedimental
- Referencia cruzada a los requisitos
- Pruebas
- Entorno tecnológico del sistema
- Plan de desarrollo e implementación

A3.2 Ámbito del software

Es importante, antes de empezar con el diseño del sistema, hacer una recapitulación del los objetivos que este tiene que satisfacer y de las restricciones que este debe de cumplir para un correcto funcionamiento.

OBJETIVOS

El principal objetivo de este sistema es el de reconocimiento automático de locutor, en sus modalidades de verificación automática de locutor e identificación automática de locutor, sin embargo estos no son los únicos objetivos que se plantean en el sistema.

Otros objetivos secundarios planteados son los siguientes:

- Grabación de ficheros de voz.
- Reproducción de ficheros de voz.
- Visualización de señales de voz.
- Cálculo de estadísticas en la utilización del sistema.

RESTRICCIONES

La principal restricción que se planteó para el desarrollo del sistema es el del sistema operativo donde este debe funcionar, siendo elegido el sistema operativo Microsoft Windows XP, no obstante como en el caso de los objetivos, existes restricciones secundarias como son:

- Creación de tres ficheros:
 1. Configuración.
 2. Estadísticas.
 3. Patrones de voz.
 4. Etiquetas.
- Mantener el fichero de configuración editable por el usuario
- Permitir borrar las estadísticas.

A3.3 Diseño de datos

Los datos manejados por el sistema desarrollado con el presente proyecto se pueden dividir en dos tipos bien diferenciados:

1. Ficheros de datos
2. Estructuras de datos

Este punto del anexo 3 está dedicado a la descripción de cada uno de ellos.

FICHEROS DE DATOS

El sistema maneja diferentes ficheros de datos, cada uno con una estructura diferente, en este apartado se determinará la estructura de todos los ficheros de datos utilizados por el sistema.

1. Fichero de configuración

El propósito de este fichero es guardar la configuración que utilizará el sistema para su correcto funcionamiento.

La configuración está compuesta por un conjunto de parámetros con su valor correspondiente.

Para saber cual es el objetivo de cada uno de los parámetros se incluye delante de el un comentario con una pequeña descripción

Por tanto la estructura del fichero de configuración es la que se establece en la siguiente figura:

```
% comentario descriptivo del parámetro1  
PARAMETRO1 = VALOR  
  
% comentario descriptivo del parámetro 2  
PARAMETRO2 = VALOR  
  
% comentario descriptivo del parámetro3  
PARAMETRO3 = VALOR  
  
% comentario descriptivo del parámetro4  
PARAMETRO4 = VALOR  
  
...
```

Figura A3.1 estructura del fichero de configuración

2. Fichero de estadísticas

El propósito de este fichero es guardar las estadísticas de utilización del sistema y además debe de poder ser borrado desde el propio sistema.

Las estadísticas están compuestas por el recuento del número de veces que se aceptó o rechazó un locutor y por el número de veces que se identificó o no se identificó un locutor.

Por tanto la estructura del fichero de estadísticas es la que se establece en la siguiente figura:

Nº de veces aceptado = VALOR Nº de veces rechazado = VALOR Nº de veces identificado = VALOR Nº de veces no identificado = VALOR
--

Figura A3.2 estructura del fichero de estadísticas

3. Fichero de patrones

El propósito de este fichero es el de almacenar los patrones de voz de los distintos usuarios.

Los patrones serán redes neuronales artificiales (RNA), y estarán todos guardados en una estructura de nombre net, que tendrá tantos campos como usuarios tenga el sistema.

Los nombres de los campos de la estructura net coincidirán con el nombre de los usuarios.

De esta forma, la estructura del fichero de patrones es la que se muestra en la siguiente figura:

net {	Red Neuronal Artificial del Locutor1
	Red Neuronal Artificial del Locutor2
	.
	.
}	Red Neuronal Artificial del LocutorN

Figura A3.3 estructura del fichero de patrones

4. Fichero de voz

Los ficheros de voz almacenan señales de voz de los distintos locutores del sistema.

El formato seleccionado para los ficheros de voz es el formato WAV y por tanto los ficheros de voz tendrán la siguiente estructura:

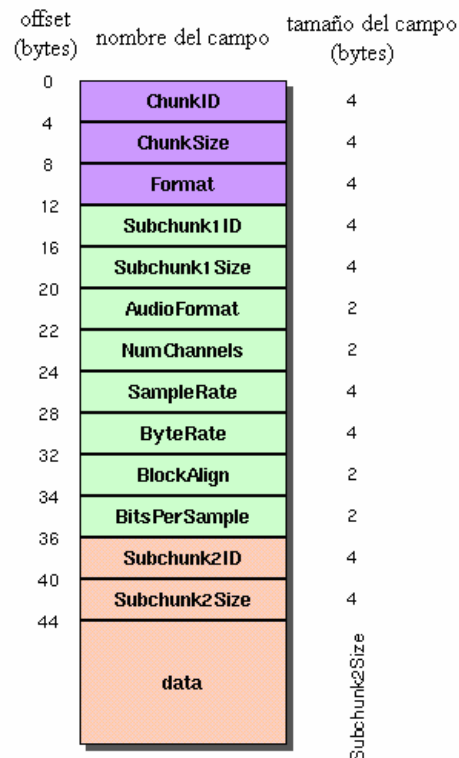


Figura A3.4 estructura del fichero de voz

donde:

ChunkID: Contiene las letras "RIFF" en formato ASCII (0x52494646).

ChunkSize: $4 + (8 + \text{SubChunk1Size}) + (8 + \text{SubChunk2Size})$.

Format: Contiene las letras "WAVE" (0x57415645).

Subchunk1ID: Contiene las letras "fmt" (0x666d7420).

Subchunk1Size: 16.

AudioFormat: 1 (sin compresión) otros valores indican el formato de compresión.

NumChannels: Mono = 1, Stereo = 2, etc.

SampleRate: Frecuencia de muestreo en Hz 8000, 44100, etc.

ByteRate: $\text{SampleRate} * \text{NumChannels} * \text{BitsPerSample} / 8$

BlockAlign: $\text{NumChannels} * \text{BitsPerSample} / 8$

BitsPerSample: 8 bits = 8, 16 bits = 16, etc.

Subchunk2ID: Contiene las letras "data" (0x64617461).

Subchunk2Size: $\text{NumSamples} * \text{NumChannels} * \text{BitsPerSample} / 8$

Data: Datos de audio.

5. Fichero de etiquetas

El propósito de este fichero es guardar las etiquetas que el usuario indica para cada fichero de voz.

Las etiquetas están compuestas por el punto inicio y final donde cada etiqueta comienza y acaba así como por la propia etiqueta.

Por tanto la estructura del fichero de etiquetas es la que se establece en la siguiente figura:

Inicio etiqueta1	Fin etiqueta1	etiqueta1
Inicio etiqueta2	Fin etiqueta2	etiqueta2
Inicio etiqueta3	Fin etiqueta3	etiqueta3
...	...	...

Figura A3.5 estructura del fichero de etiquetas

6. Fichero de ayuda

El fichero de ayuda contiene la ayuda de utilización de programa en el formato .chm para ser abierto con un navegador web.

ESTRUCTURAS DE DATOS

El sistema maneja diferentes estructuras de datos en función del objetivo que se pretenda satisfacer en cada momento.

A continuación se incluye una descripción de las estructuras de datos empleadas en el sistema:

1. vectores endpoint

Estos son los vectores (dos vectores) que devuelve la función endpoint descrita en el anexo 4 en el punto A4.3.

El contenido de estos vectores son el número de las tramas donde comienzan y terminan todas las palabras de una señal de voz.

La estructura de estos vectores es la siguiente:

$$\begin{aligned} \text{Inicio} &= [\text{trama de inicio } 1^{\circ} \text{ palabra, trama de inicio } 2^{\circ} \text{ palabra, ...}] \\ \text{Fin} &= [\text{trama de fin } 1^{\circ} \text{ palabra, trama de fin } 2^{\circ} \text{ palabra, ...}] \end{aligned}$$

Figura A3.6 estructura del vector endpoint

Donde las tramas son números enteros.

1. Matriz de parámetros

Esta matriz es la que devuelve la función parametros descrita en el anexo 4 en el punto A4.3.

El contenido de esta matriz son los parámetros de todas las tramas en las que se divide una señal de voz.

La estructura de esta matriz es:

$$\begin{bmatrix} \text{Parametro 1 de trama 1} & \text{Parametro 1 de trama 2} & . & . & \text{Parametro 1 de trama } n \\ \text{Parametro 2 de trama 1} & \text{Parametro 2 de trama 2} & . & . & \text{Parametro 2 de trama } n \\ . & . & . & . & . \\ . & . & . & . & . \\ \text{Parametro } n \text{ de trama 1} & \text{Parametro } n \text{ de trama 2} & . & . & \text{Parametro } n \text{ de trama } n \end{bmatrix}$$

Figura A3.7 estructura de la matriz de parámetros

Donde los parámetros son números.

A3.4 Diseño arquitectónico

Uno de los principales problemas que se plantean a la hora de diseñar el sistema a desarrollar es el de dividir el problema para poder abordarlo de mejor forma.

La resolución de este problema no es trivial y en ocasiones es difícil de conseguir.

Para tratar de resolver este problema en el sistema que se desarrolla en el presente proyecto se ha recurrido a la técnica de DFD's (Diagramas de Flujo de Datos), que permite dividir un sistema grande en otros más sencillos e ir tratándolos por niveles.

Inicialmente se determinan las interfaces del sistema con el exterior, el sistema SERAL debería recibir la siguiente información por parte de sus usuarios:

- Señal de voz
- Nombre de ficheros
- Id de locutor

De la misma forma el sistema debe dar cierta información a sus usuarios como es:

- Reproducción de voz
- Visualización de señal de voz
- Resultado del reconocimiento

Por tanto se plantea el siguiente diagrama de contexto para el sistema SERAL:

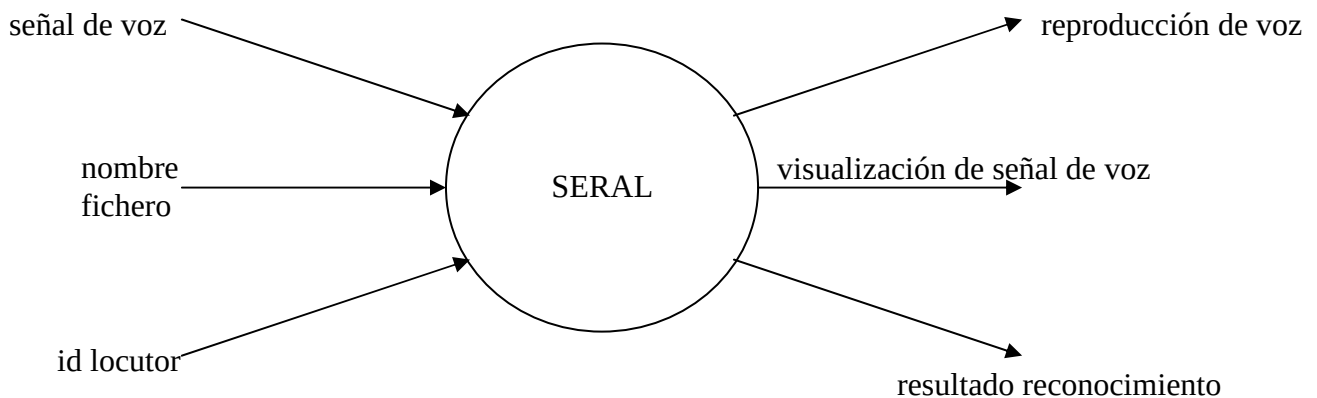


Figura A3.8 Diagrama de contexto

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

Una vez determinada la interfaz del sistema con el exterior, se descompone el sistema en varios subsistemas para satisfacer los requisitos funcionales recogidos en el anexo 2 dentro del apartado A2.4.1.

Para conseguir este objetivo, se divide el sistema en 10 subsistemas, encargado cada uno de ellos de satisfacer alguno de los anteriormente mencionados requisitos funcionales.

Los 10 subsistemas creados son los siguientes:

1. entrenar
2. procesar señal de voz
3. eliminar locutor
4. visualizar señal de voz
5. reproducir señal de voz
6. escribir fichero de voz
7. leer fichero de voz
8. ver estadísticas
9. borrar estadísticas
10. actualizar configuración

Sin embargo estos subsistemas no son independientes unos de otros, sino que se encuentran interrelacionados e intercambian información.

Parte de este intercambio de información se realiza a través de unos ficheros que fueron indicados en el anexo 2 (Especificación de requisitos del software), y que fueron descritos en el apartado A3.3 de este mismo anexo como son los siguientes:

- el fichero de configuración
- el fichero de patrones de voz
- el fichero de estadísticas
- los ficheros de audio

De acuerdo con esto la estructura del sistema puede quedar dispuesta conforme al siguiente DFD:

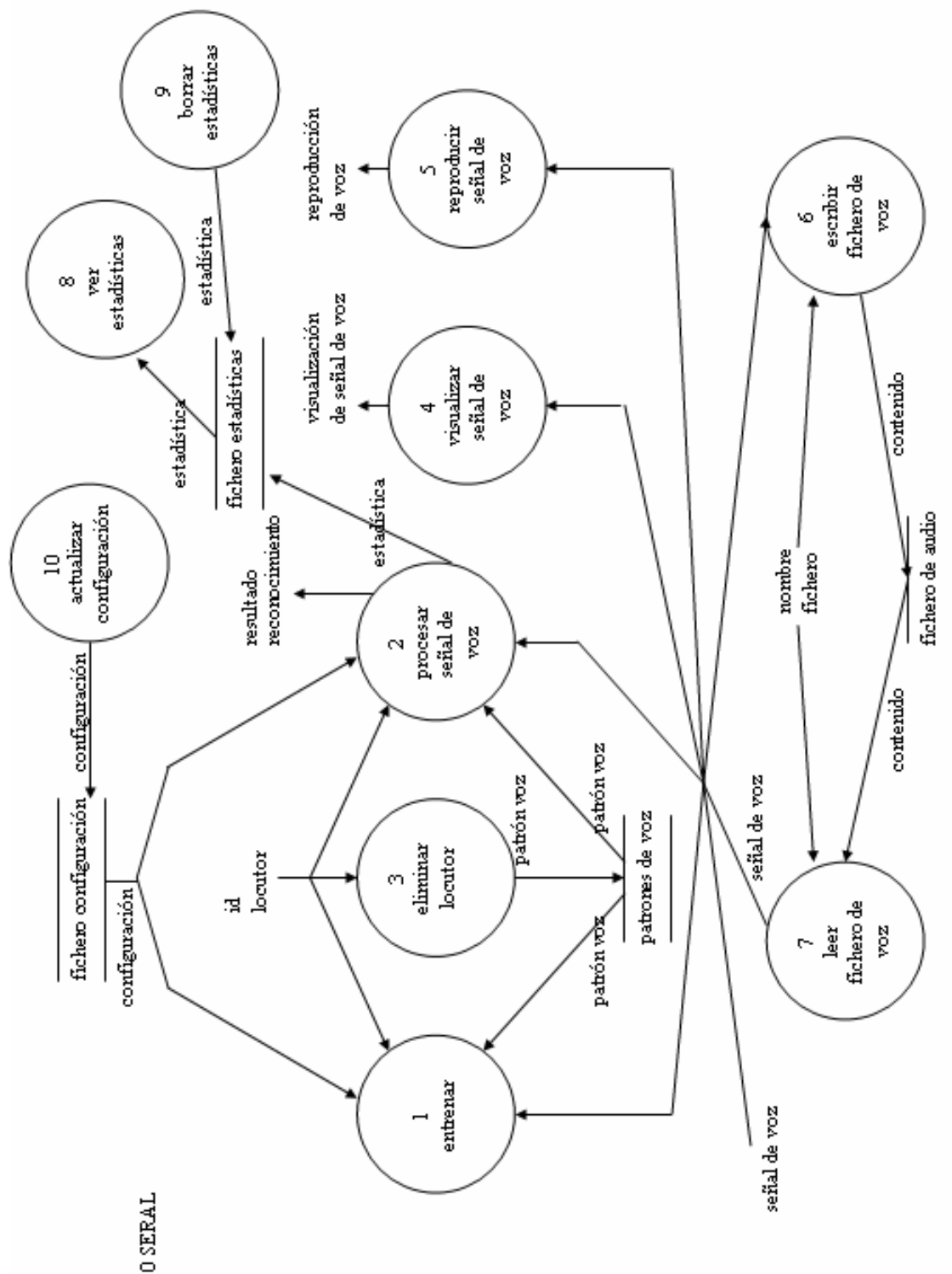


Figura A3.9 DFD 0 SERAL

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

Con la descomposición del sistema que muestra la figura anterior, este queda dividido en subsistemas más sencillos, sin embargo alguno de ellos no es todavía lo suficientemente sencillo, como es el caso de los siguientes subsistemas:

1. entrenar
2. procesar señal de voz

El subsistema 1 todavía podría descomponerse en otros más sencillos como son:

- 1.1 extraer parámetros
- 1.2 construir patrón de voz
- 1.3 almacenar patrón de voz

Estos subsistemas en los que se divide el subsistema 1, a su vez no son independientes unos de otros, sino que como ocurría con los subsistemas de nivel 1, también se encuentran interrelacionados.

Estos subsistemas en los que subdivide el subsistema 1 acceden a los mismos ficheros y flujos de datos que este, a la vez que añaden otros nuevos como son:

- parámetros
- patrón de voz

Estos flujos de datos nuevos son añadidos para la comunicación de los nuevos subsistemas como se puede ver en el siguiente DFD.

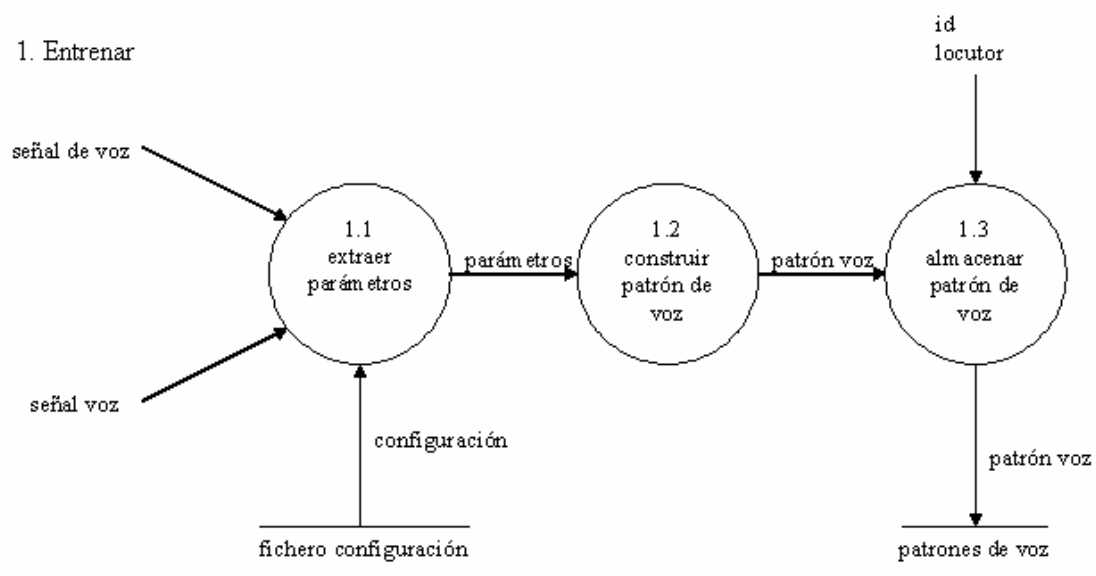


Figura A3.10 DFD 1 Entrenar

De igual manera el subsistema 2 también se descompone en otros más simples como son:

- 2.1 extraer parámetros
- 2.2 construir patrón de voz
- 2.3 leer patrón de voz
- 2.4 comparar patrones de voz

Estos subsistemas al igual que ocurría con el subsistema 1, no son independientes unos de otros, sino que están interrelacionados.

Los nuevos subsistemas también acceden a los mismos ficheros y flujos de datos que el subsistema 2, a la vez que añaden otros nuevos como son:

- parámetros
- patrón de voz

Estos flujos de datos nuevos son añadidos para la comunicación de los nuevos subsistemas como se puede ver en el siguiente DFD.

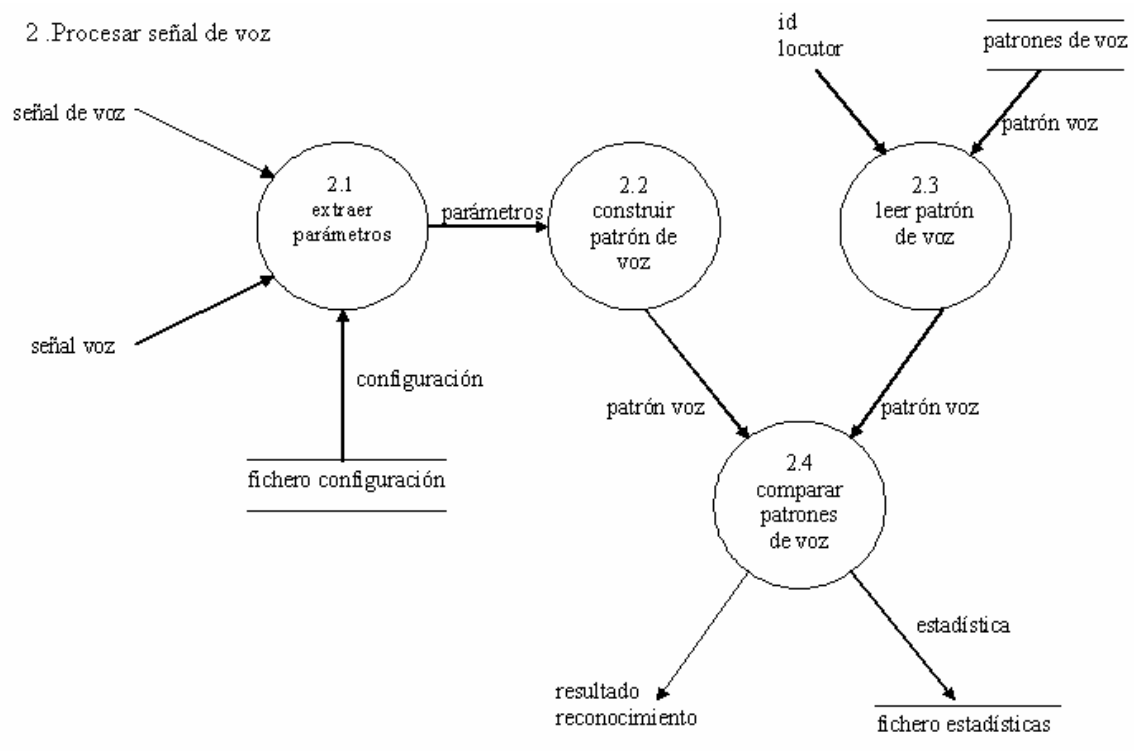


Figura A3.11 DFD 2 Procesar señal de voz

Con la descomposición de los subsistemas 1 y 2 que muestran las dos figuras anteriores, el sistema queda dividido en subsistemas suficientemente sencillos.

Como se ha indicado anteriormente los subsistemas definidos no son independientes y se tienen que comunicar por medio de flujos de datos, sin embargo estos no son tipos primitivos como puede ser el tipo de dato ENTERO y por tanto hay que especificarlos más detalladamente.

Con el propósito de definir los diferentes flujos de datos que existen en el sistema se desarrolla el siguiente diccionario de datos, donde los tipos complejos son descompuestos en otros más simples.

DICCIONARIO DE DATOS DEL SISTEMA

señal de voz = {número}

número = [0-9]

nombre fichero = {letra | número}

letra = [a-z | A-Z | signo puntuación]

signo puntuación = [. | - | _]

id locutor = {letra}

reproducción de voz = {número}

visualización de señal de voz = señal

señal = {número}

resultado reconocimiento = [aceptado | rechazado | identificado | no identificado]

patrón voz = {número}

configuración = {número}

estadística = n° veces aceptado + n° veces rechazado + n° veces identificado + n° veces no identificado

n° veces aceptado = {número}

n° veces rechazado = {número}

n° veces identificado = {número}

n° veces no identificado = {número}

contenido = {número}

parámetros = {número}

Otro de los aspectos que aspectos que hay que aclarar de los diagramas de flujo de datos, es la función de cada subsistema, ya que aunque esta pueda ser intuitiva por medio del nombre de los mismos, puede no quedar lo suficientemente claro el objetivo de los distintos subsistemas.

Con objeto de determinar brevemente la función de cada subsistema se incluye en este punto una pequeña descripción de los mismos, que en el caso de los más importantes será ampliada en el punto A3.6 de este mismo anexo.

DESCRIPCIÓN DEL PROCESO 1.1

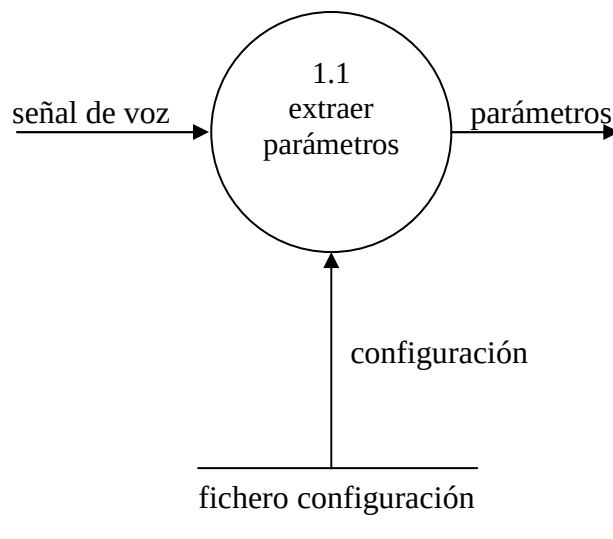


Figura A3.12 DFD 1.1 extraer parámetros

FUNCIONAMIENTO DEL PROCESO 1.1 EXTRAER PARÁMETROS

1. Se recibe la señal de voz
2. Se lee la configuración del fichero de configuración
3. Se divide la señal de voz en tramas
4. Se calculan los parámetros de cada trama
5. Se insertan estos parámetros en una matriz de parámetros
6. Se devuelve la matriz de parámetros

DATOS DEL DICCIONARIO EMPLEADOS

señal de voz = {número}
número = [0-9]
configuración = {número}
parámetros = {número}

NOTA: es subsistema 1.1 es equivalente al 2.1 y por tanto este último no será definido en el presente documento.

DESCRIPCIÓN DEL PROCESO 1.2

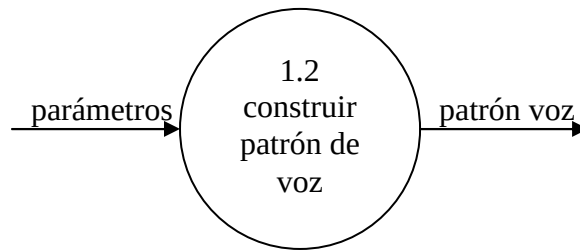


Figura A3.13 DFD 1.2 construir patrón de voz

FUNCIONAMIENTO DEL PROCESO 1.2 CONSTRUIR PATRÓN DE VOZ

DATOS DEL DICCIONARIO EMPLEADOS

NOTA: es subsistema 1.2 es equivalente al 2.2 y por tanto este último no será definido en el presente documento.

DESCRIPCIÓN DEL PROCESO 1.3

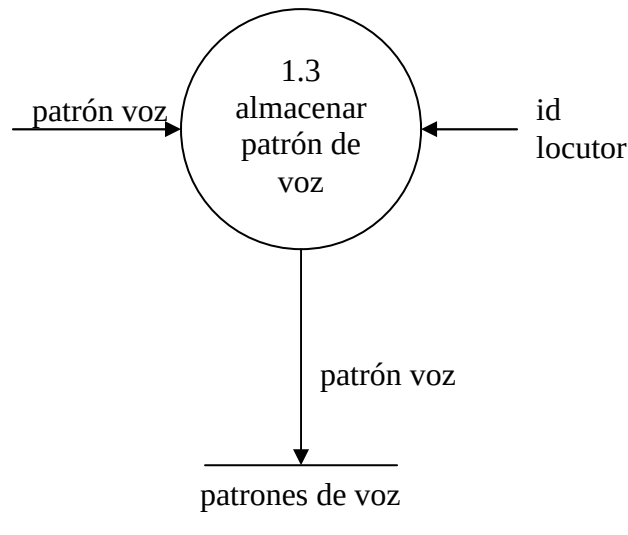


Figura A3.14 DFD 1.3 almacenar patrón de voz

FUNCIONAMIENTO DEL PROCESO 1.3 ALMACENAR PATRÓN DE VOZ

1. Recibir patrón de voz
2. Recibir id del locutor
3. Si el locutor existe
 - a. Descartar el patrón de voz
4. Si el locutor no existe
 - a. Almacenar el id del locutor en el fichero de patrones de voz.
 - b. Almacenar el patrón de voz en el fichero de patrones de voz.

DATOS DEL DICCIONARIO EMPLEADOS

id locutor = {letra}
letra = [a-z | A-Z | signo puntuación]
signo puntuación = [. | - | _]
patrón voz = {número}
número = [0-9]

DESCRIPCIÓN DEL PROCESO 2.3

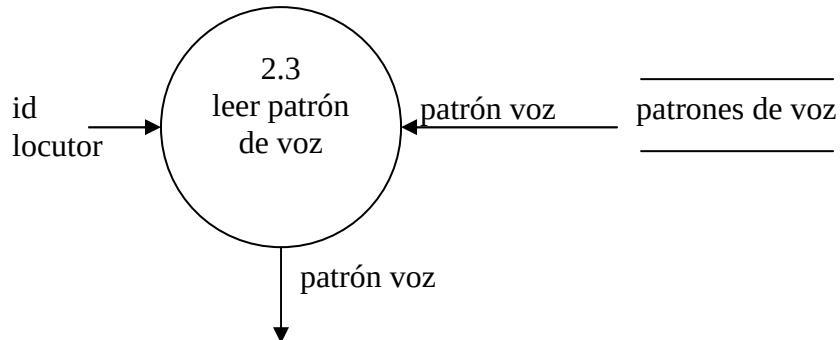


Figura A3.15 DFD 2.3 leer patrón de voz

FUNCIONAMIENTO DEL PROCESO 2.3 LEER PATRÓN DE VOZ

1. Recibir id del locutor
2. Si el locutor existe
 - a. Leer patrón de voz
3. Si el locutor no existe
 - a. Descartar petición

DATOS DEL DICCIONARIO EMPLEADOS

id locutor = {letra}
letra = [a-z | A-Z | signo puntuación]
signo puntuación = [. | - | _]
patrón voz = {número}
número = [0-9]

DESCRIPCIÓN DEL PROCESO 2.4

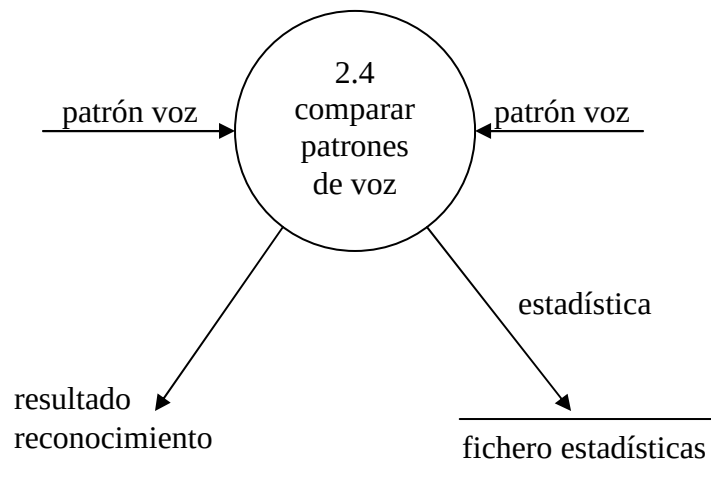


Figura A3.16 DFD 2.4 comparar patrones de voz

FUNCIONAMIENTO DEL PROCESO 2.4 COMPARAR PATRONES DE VOZ

1. Recibir patrón de voz del usuario a reconocer
2. Recibir patrón de voz del fichero
3. Si los patrones coinciden
 - a. Indicar resultado positivo
 - b. Actualizar estadísticas
4. Si los patrones no coinciden
 - a. Indicar resultado negativo
 - b. Actualizar estadísticas

DATOS DEL DICCIONARIO EMPLEADOS

patrón voz = {número}

número = [0-9]

resultado reconocimiento = [aceptado | rechazado | identificado | no identificado]

estadística = n° veces aceptado + n° veces rechazado + n° veces identificado + n° veces no identificado

n° veces aceptado = {número}

n° veces rechazado = {número}

n° veces identificado = {número}

n° veces no identificado = {número}

DESCRIPCIÓN DEL PROCESO 3

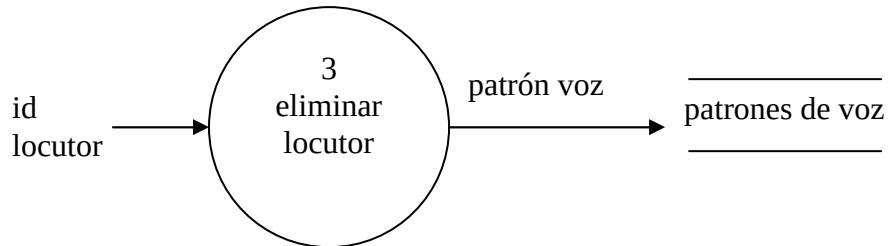


Figura A3.17 DFD 3 eliminar locutor

FUNCIONAMIENTO DEL PROCESO 3 ELIMINAR LOCUTOR

1. Recibir el id del locutor a eliminar
2. Buscar su patrón de voz en el fichero de patrones
3. Si se encuentra el patrón
 - a. Eliminar el patrón del fichero
4. Si no se encuentra el patrón
 - a. No realizar ninguna acción

DATOS DEL DICCIONARIO EMPLEADOS

id locutor = {letra}
letra = [a-z | A-Z | signo puntuación]
signo puntuación = [. | - | _]
patrón voz = {número}
número = [0-9]

DESCRIPCIÓN DEL PROCESO 4

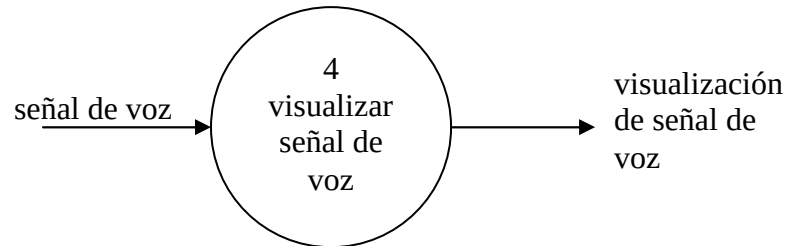


Figura A3.18 DFD 4 visualizar señal de voz

FUNCIONAMIENTO DEL PROCESO 4 VISUALIZAR SEÑAL DE VOZ

1. Recibir señal de voz
2. Dibujar la señal en el osciloscopio

DATOS DEL DICCIONARIO EMPLEADOS

señal de voz = {número}

número = [0-9]

visualización de señal de voz = señal

señal = {número}

DESCRIPCIÓN DEL PROCESO 5

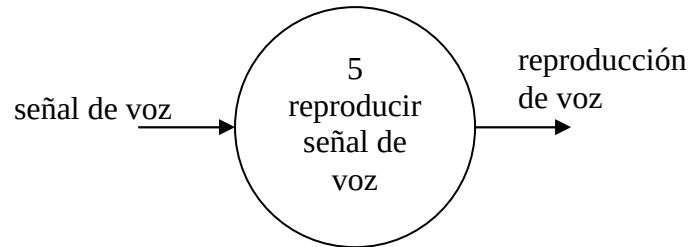


Figura A3.19 DFD 5 reproducir señal de voz

FUNCIONAMIENTO DEL PROCESO 5 REPRODUCIR SEÑAL DE VOZ

1. Recibir señal de voz
2. Enviar señal de voz a los altavoces

DATOS DEL DICCIONARIO EMPLEADOS

señal de voz = {número}

número = [0-9]

reproducción de voz = {número}

DESCRIPCIÓN DEL PROCESO 6

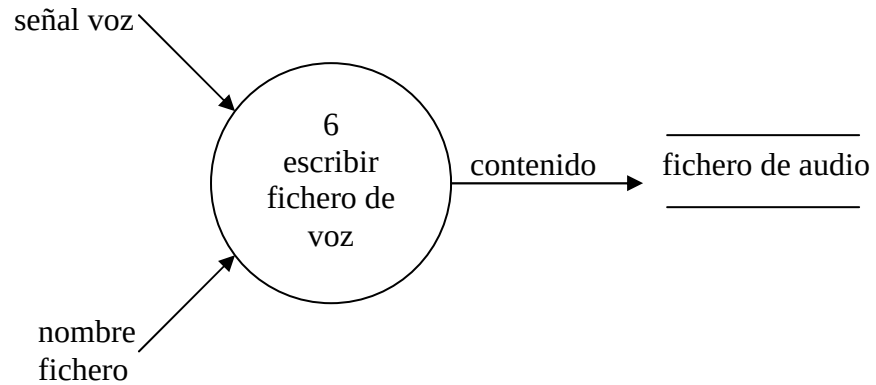


Figura A3.20 DFD 6 escribir fichero de voz

FUNCIONAMIENTO DEL PROCESO 6 ESCRIBIR FICHERO DE VOZ

1. Recibir señal de voz
2. Recibir nombre de fichero
3. Almacenar el contenido de la señal de voz en el fichero

DATOS DEL DICCIONARIO EMPLEADOS

señal de voz = {número}

número = [0-9]

nombre fichero = {letra | número}

letra = [a-z | A-Z | signo puntuación]

signo puntuación = [. | - | _]

contenido = {número}

DESCRIPCIÓN DEL PROCESO 7

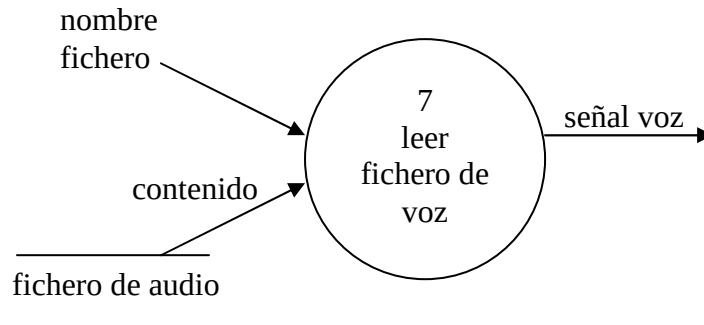


Figura A3.21 DFD 7 leer fichero de voz

FUNCIONAMIENTO DEL PROCESO 7 LEER FICHERO DE VOZ

1. Recibir nombre de fichero
2. Leer el contenido del fichero
3. Devolver la señal de voz leída

DATOS DEL DICCIONARIO EMPLEADOS

señal de voz = {número}

número = [0-9]

nombre fichero = {letra | número}

letra = [a-z | A-Z | signo puntuación]

signo puntuación = [. | - | _]

contenido = {número}

DESCRIPCIÓN DEL PROCESO 8

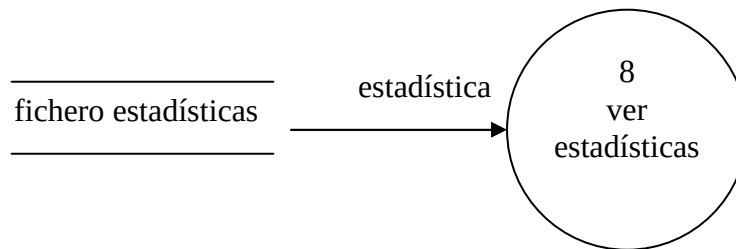


Figura A3.22 DFD 8 ver estadísticas

FUNCIONAMIENTO DEL PROCESO 8 VER ESTADÍSTICAS

1. Leer el contenido del fichero de estadísticas
2. Mostrar el contenido al usuario

DATOS DEL DICCIONARIO EMPLEADOS

estadística = nº veces aceptado + nº veces rechazado + nº veces identificado + nº veces no identificado

nº veces aceptado = {número}

nº veces rechazado = {número}

nº veces identificado = {número}

nº veces no identificado = {número}

DESCRIPCIÓN DEL PROCESO 9

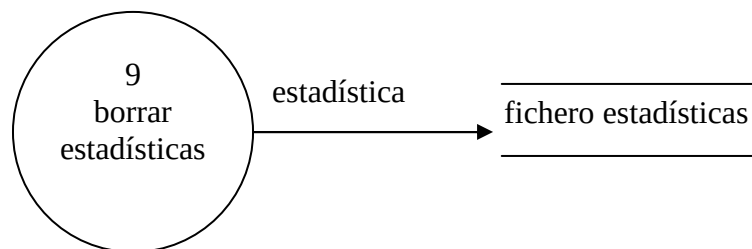


Figura A3.23 DFD 9 borrar estadísticas

FUNCIONAMIENTO DEL PROCESO 9 BORRAR ESTADÍSTICAS

1. Poner a 0 todas las estadísticas del fichero

DATOS DEL DICCIONARIO EMPLEADOS

estadística = nº veces aceptado + nº veces rechazado + nº veces identificado + nº veces no identificado

nº veces aceptado = {número}

nº veces rechazado = {número}

nº veces identificado = {número}

nº veces no identificado = {número}

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor
DESCRIPCIÓN DEL PROCESO 10

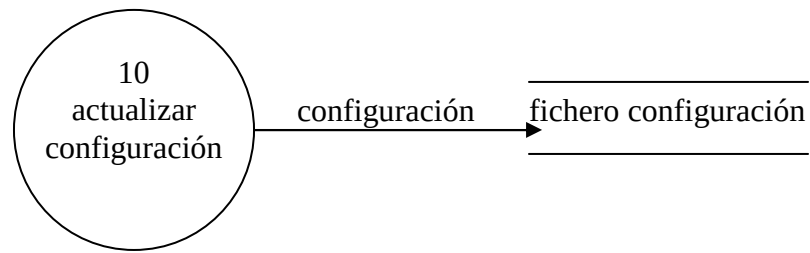


Figura A3.24 DFD 10 actualizar configuración

FUNCIONAMIENTO DEL PROCESO 10 ACTUALIZAR CONFIGURACIÓN

1. Abrir el fichero de configuración
2. Esperar que el usuario cambie la configuración
3. Grabar la nueva configuración en el fichero

DATOS DEL DICCIONARIO EMPLEADOS

configuración = {número}

A3.5 Diseño de la interfaz

La interfaz de usuario es una de las partes más importantes a tener en cuenta a la hora del diseño de un sistema interactivo.

Del diseño realizado para la interfaz dependerán aspectos tan importantes como la facilidad de uso del sistema, que es clave para la satisfacción del usuario.

La interfaz de usuario del sistema que se desarrolla con es presente proyecto, quedó inicialmente esbozada en el anexo 2 (Especificación de requisitos del software) dentro del apartado A2.5 (Interfaces de usuario), sin embargo la descripción más detallada de la interfaz se realizará en este punto.

Como se indica en el punto A2.5, la parte principal de la interfaz de usuario está compuesta por una ventana principal, sin embargo, esta no es la única que conforma la interfaz, sino que existen otras como:

- Cuadro de dialogo de salida
- Cuadro de dialogo de abrir fichero
- Cuadro de dialogo de mostrar estadísticas
- Cuadro de dialogo de borrar estadísticas
- Cuadro de dialogo de ayuda
- Cuadro de dialogo de reconocimiento
- Cuadro de diálogo de mostrar resultados
- Cuadro de diálogo de nombrar locutor
- Cuadro de diálogo de inicio de grabación
- Cuadro de diálogo de etiquetar
- Cuadro de diálogo de ver etiquetas
- Cuadro de diálogo de fichero de etiquetas

También hay que tener en cuenta que cada uno de los cuadros de dialogo mencionados tienen sus propios controles.

El objetivo de este punto es realizar un análisis detallado de todas las ventanas y controles que conforman la interfaz del sistema.

VENTANA PRINCIPAL

La ventana principal es la que se muestra en la siguiente figura:

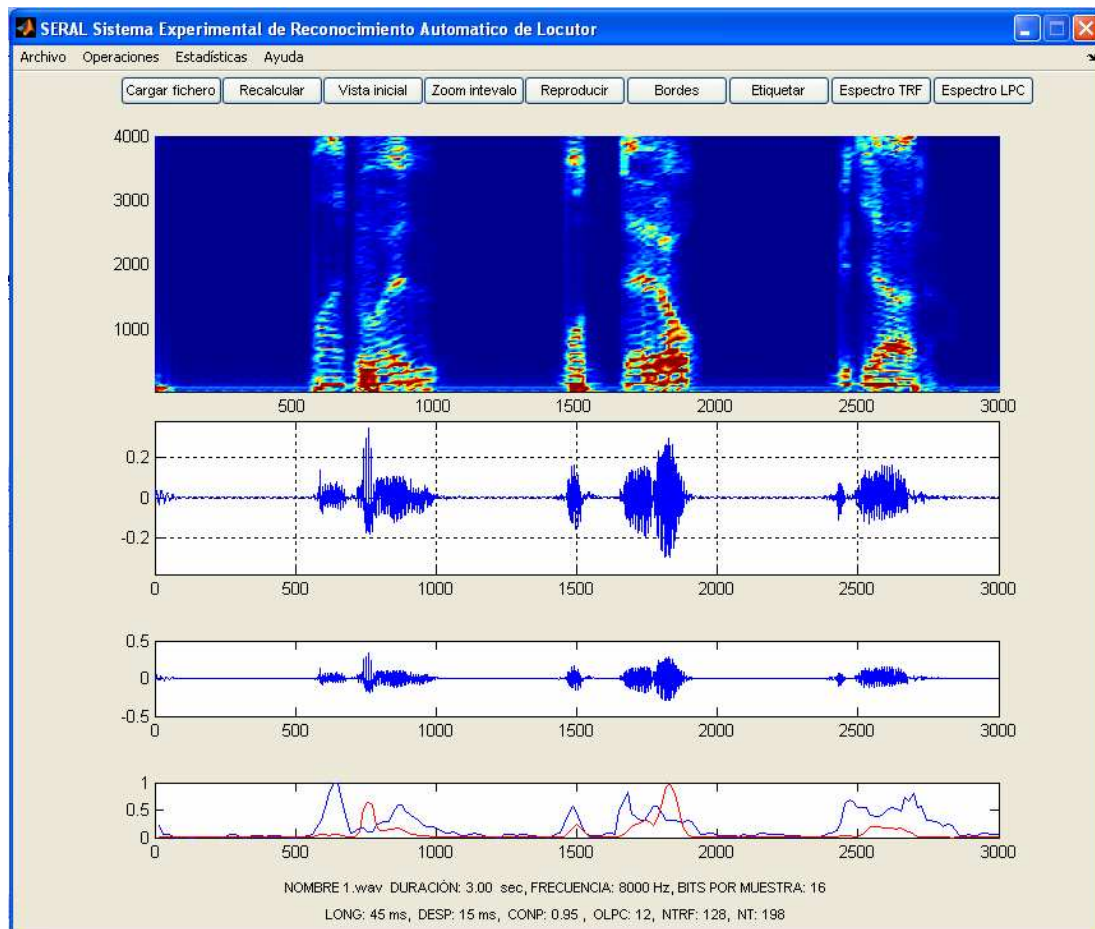


Figura A3.25 Vista de la venta principal de SERAL

Esta, está compuesta por dos etiquetas en la parte inferior, que muestran información sobre el fichero de voz reproducido y sobre los parámetros empleados para el procesado de la señal, un osciloscopio, que muestra la señal de voz con la que se trabaja, unos botones en la parte superior que dan acceso directo a ciertas operaciones de la aplicación y por unos menús, que permiten acceder a todas las operaciones de la aplicación.

El contenido de los menús es el que se puede ver en la siguiente figura:

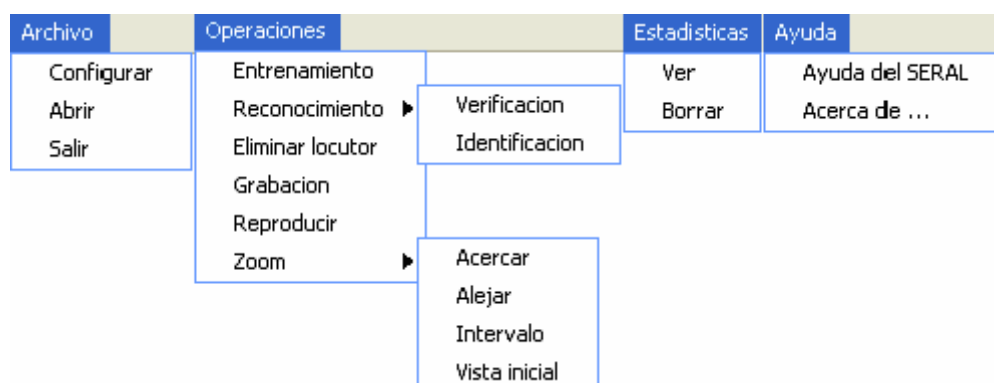


Figura A3.26 Menús de la ventana principal de SERAL

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

En la siguiente tabla se puede ver las acciones que conlleva activar alguno de los controles que tiene esta ventana.

Botón cargar fichero	Lanza el cuadro de diálogo del la figura 3.27 Reproduce el fichero seleccionado Muestra la señal de voz del fichero en la pantalla
Botón Vista inicial	Muestra en pantalla la señal completa
Botón Recalcular	Recalcula las gráficas de la señal que se están viendo en pantalla
Botón Zoom intervalo	Permite seleccionar dos puntos del osciloscopio Muestra en pantalla el contenido entre los dos puntos
Botón Reproducir	Permite seleccionar dos puntos del osciloscopio Reproduce la señal de voz contenida entre los dos puntos
Botón Bordes	Muestra en pantalla el inicio y fin de cada palabra de la señal de voz.
Botón Etiquetar	Pide dos puntos de la señal de voz. Pide la etiqueta a asignar a la señal de voz entre esos dos puntos.
Botón Espectro TRF	Muestra el espectrograma calculado con la TRF
Botón Espectro LPC	Muestra el espectrograma calculado con los coeficientes LPC
Menú Archivo	Despliega el menú Archivo
Opción Configurar	Lanza el editor notepad con el fichero de configuración seral.cfg
Opción Abrir	Igual que el botón Cargar fichero
Opción Salir	Lanza el cuadro de diálogo del la figura 3.26 Decide si hay que salir de la aplicación
Menú Operaciones	Despliega el menú Operaciones
Opción Entrenamiento	Pide el nombre del locutor con el cuadro de diálogo de la figura 3.33 Pide ficheros de voz con el cuadro de diálogo de la figura 3.27 Crea y entrena la red neuronal artificial para el locutor
Menú Reconocimiento	Despliega el menú Reconocimiento
Opción Verificación	Lanza el cuadro de diálogo del la figura 3.31 Realiza la verificación del locutor según el modo elegido
Opción Identificación	Lanza el cuadro de diálogo del la figura 3.31 Realiza la identificación del locutor según el modo elegido
Opción Eliminar locutor	Pide el nombre del locutor con el cuadro de diálogo de la figura 3.33 Elimina ese locutor si existe
Opción Grabación	Lanza el cuadro de diálogo del la figura 3.31 Graba una señal de voz por micrófono Guarda la señal de voz en el fichero seleccionado
Opción Reproducir	Igual que el botón Reproducir
Menú Zoom	Despliega el menú Zoom
Opción Acercar	Muestra en pantalla el 25% alrededor del punto central
Opción Alejar	Muestra en pantalla el 200% alrededor del punto central
Opción Intervalo	Igual que el botón Zoom intervalo
Opción Vista inicial	Igual que el botón Vista inicial
Menú Estadísticas	Despliega el menú Estadísticas
Opción Ver	Lanza el cuadro de diálogo del la figura 3.28
Opción Borrar	Lanza el cuadro de diálogo del la figura 3.29 Decide si hay que borrar las estadísticas
Menú Ayuda	Despliega el menú Ayuda
Opción Ayuda del SERAL	Abre el fichero de ayuda del SERAL
Opción Acerca de ...	Lanza el cuadro de diálogo del la figura 3.30

Tabla A3.1 controles de la ventana principal de SERAL

CUADRO DE DIALOGO DE SALIDA

El cuadro de dialogo de salida es el que se muestra en la siguiente figura:

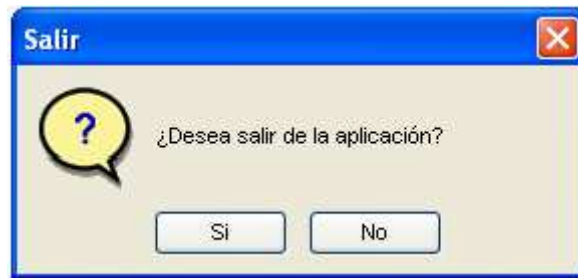


Figura A3.27 Vista del cuadro de diálogo de salida

Esta compuesto por un logotipo, una etiqueta con la pregunta “¿Desea salir de la aplicación?” y dos botones que ofrecen dos respuestas (“Si” y “No”).

En la siguiente tabla se puede ver las acciones que conlleva activar alguno de los controles que tiene este cuadro de diálogo.

Botón Si	Cierra el cuadro de diálogo Devuelve la cadena “Si”
Botón No	Cierra el cuadro de diálogo Devuelve la cadena “No”

Tabla A3.2 controles del cuadro de diálogo de salida

CUADRO DE DIALOGO DE ABRIR FICHERO

El cuadro de diálogo de abrir fichero es el que muestra la siguiente figura:

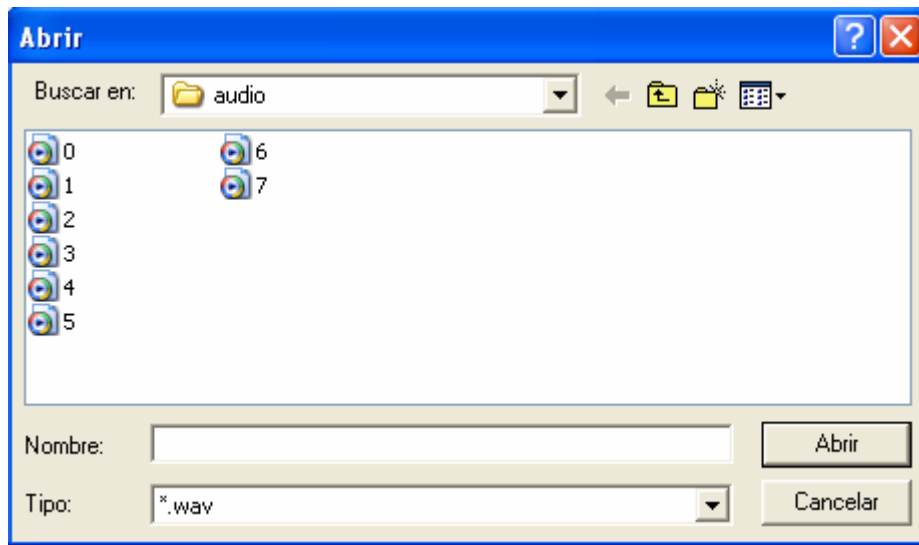


Figura A3.28 Vista del cuadro de diálogo de abrir fichero

Esta compuesto por un cuadro de texto donde introducir el nombre de un fichero o directorio, dos botones (Abrir y Canelar), una lista de ficheros y directorios.

En la siguiente tabla se puede ver las acciones que conlleva activar alguno de los controles que tiene este diálogo.

Botón Abrir	Si hay seleccionado un directorio Muestra el contenido del nuevo directorio Si hay seleccionado un fichero Devuelve la ruta completa del fichero
Botón Cancelar	Cierra el cuadro de dialogo
Lista de ficheros y directorios	Si hay seleccionado un directorio Muestra el contenido del nuevo directorio Si hay seleccionado un fichero Devuelve la ruta completa del fichero

Tabla A3.3 controles del cuadro de diálogo de abrir fichero

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor
CUADRO DE DIALOGO DE MOSTRAR ESTADÍSTICAS

El cuadro de dialogo de mostrar estadísticas es el que se muestra en la siguiente figura:

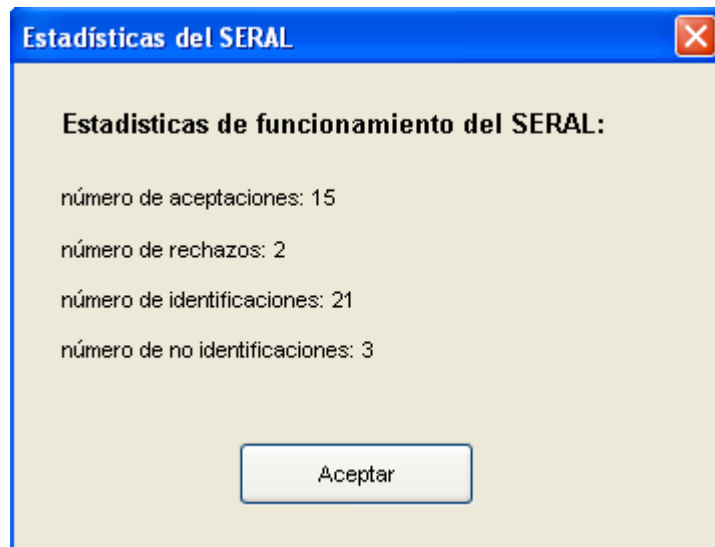


Figura A3.29 Vista del cuadro de diálogo de mostrar estadísticas

Esta compuesto por una etiqueta que muestra las estadísticas de utilización del sistema, y por un botón (Aceptar).

En la siguiente tabla se puede ver las acciones que conlleva activar alguno de los controles que tiene este cuadro de diálogo.

Botón Aceptar	Cierra el cuadro de diálogo
---------------	-----------------------------

Tabla A3.4 controles del cuadro de diálogo de mostrar estadísticas

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor
CUADRO DE DIALOGO DE BORRAR ESTADÍSTICAS

El cuadro de dialogo de borrar estadísticas se muestra en la siguiente figura:

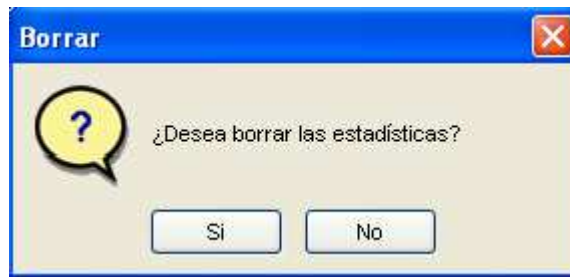


Figura A3.30 Vista del cuadro de diálogo de borrar estadísticas

Esta compuesto por un logotipo, una etiqueta con la pregunta “¿Desea borrar las estadísticas?” y dos botones que ofrecen dos respuestas (“Si” y “No”).

En la siguiente tabla se puede ver las acciones que conlleva activar alguno de los controles que tiene este cuadro de diálogo.

Botón Si	Cierra el cuadro de diálogo Devuelve la cadena “Si”
Botón No	Cierra el cuadro de diálogo Devuelve la cadena “No”

Tabla A3.5 controles del cuadro de diálogo de borrar estadísticas

CUADRO DE DIALOGO DE AYUDA

El cuadro de dialogo de ayuda se muestra en la siguiente figura:



Figura A3.31 Vista del cuadro de diálogo de ayuda

Esta compuesto por el logotipo de SERAL, una etiqueta con información acerca del sistema y un botón (Aceptar).

En la siguiente tabla se puede ver las acciones que conlleva activar alguno de los controles que tiene este cuadro de diálogo.

Botón Aceptar	Cierra el cuadro de diálogo
---------------	-----------------------------

Tabla A3.6 controles del cuadro de diálogo de ayuda

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor
CUADRO DE DIALOGO DE RECONOCIMIENTO

El cuadro de dialogo de reconocimiento es el que se muestra en la siguiente figura:

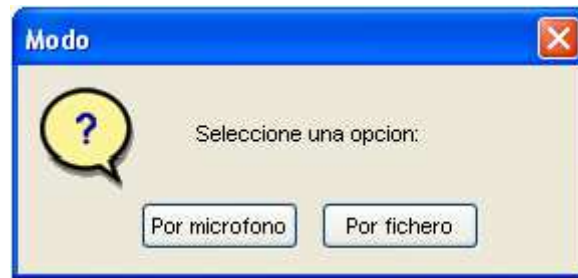


Figura A3.32 Vista del cuadro de diálogo de reconocimiento

Esta compuesto por un logotipo, una etiqueta que pregunta la modalidad con la que queremos hacer el reconocimiento (tanto la verificación como la identificación) y dos botones que ofrecen dos respuestas (“Por microfono” y “Por fichero”).

En la siguiente tabla se puede ver las acciones que conlleva activar alguno de los controles que tiene este cuadro de diálogo:

Botón Por microfono	Cierra el cuadro de diálogo Devuelve la cadena “Por microfono”
Botón Por fichero	Cierra el cuadro de diálogo Devuelve la cadena “Por fichero”

Tabla A3.7 controles del cuadro de diálogo de reconocimiento

CUADRO DE DIÁLOGO DE MOSTRAR RESULTADOS

El cuadro de diálogo empleado para mostrar resultados es el que se muestra en la siguiente figura:



Figura A3.33 Vista del cuadro de diálogo de mostrar resultados

Está compuesto por una etiqueta que muestra el resultado del reconocimiento y un botón para aceptar.

Los posibles resultados posibles son:

- Verificación exitosa: nombre (resultado): -Si la verificación del locutor nombre tiene éxito-
- Verificación no exitosa (resultado) -Si la verificación no tiene éxito-
- Identificación exitosa: nombre (resultado) -Si la identificación del locutor nombre tiene éxito-
- Identificación no exitosa -Si la identificación no tiene éxito-

En la siguiente tabla se puede ver las acciones que conlleva activar alguno de los controles que tiene este cuadro de diálogo:

Botón Aceptar	Cierra el cuadro de diálogo
---------------	-----------------------------

Tabla A3.8 controles del cuadro de diálogo de mostrar resultados

CUADRO DE DIÁLOGO DE NOMBRAR LOCUTOR

El cuadro de diálogo utilizado para introducir nombres de locutores es el que se muestra en la siguiente figura:

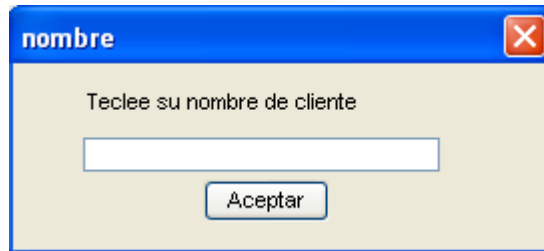


Figura A3.34 Vista del cuadro de diálogo de nombrar locutor

Esta compuesto por una etiqueta que pide que se introduzca el nombre de un locutor, un cuadro de texto, donde se introduce dicho nombre y un botón (Aceptar).

En la siguiente tabla se puede ver las acciones que conlleva activar alguno de los controles que tiene este cuadro de diálogo:

Botón Aceptar	Cierra el cuadro de diálogo Devuelve el nombre contenido en el cuadro de texto
---------------	---

Tabla A3.9 controles del cuadro de diálogo de nombrar locutor

CUADRO DE DIÁLOGO DE INCIO DE GRABACIÓN

El cuadro de diálogo utilizado para indicar el inicio de una grabación es el que se muestra en la siguiente figura:

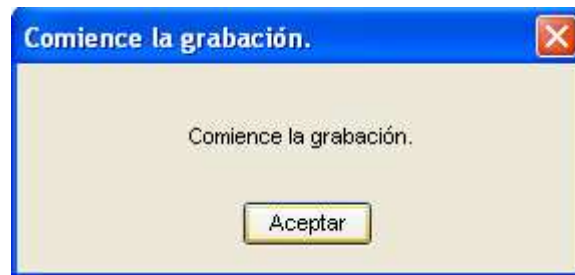


Figura A3.35 Vista del cuadro de diálogo de inicio de grabación

Está compuesto por una etiqueta que nos indica que podemos iniciar la grabación y un botón para aceptar.

En la siguiente tabla se puede ver las acciones que conlleva activar alguno de los controles que tiene este cuadro de diálogo:

Botón Aceptar	Cierra el cuadro de diálogo Se inicia la grabación
---------------	---

Tabla A3.10 controles del cuadro de diálogo de inicio de grabación

CUADRO DE DIÁLOGO DE ETIQUETAR

El cuadro de diálogo utilizado para recoger una etiqueta indicada por el usuario es el que se muestra en la siguiente figura:

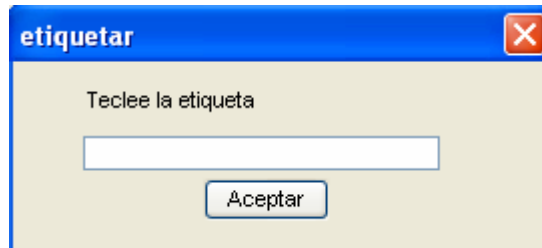


Figura A3.36 Vista del cuadro de diálogo de etiquetar

Está compuesto por una etiqueta informativa, un cuadro de texto donde se indica la etiqueta y un botón para aceptar.

En la siguiente tabla se puede ver las acciones que conlleva activar alguno de los controles que tiene este cuadro de diálogo:

Botón Aceptar	Cierra el cuadro de diálogo Devuelve la etiqueta contenida en el cuadro de texto
---------------	---

Tabla A3.11 controles del cuadro de diálogo de etiquetar

CUADRO DE DIÁLOGO DE VER ETIQUETAS

El cuadro de diálogo utilizado para decidir si se visualizan o no las etiquetas es el que se muestra en la siguiente figura:

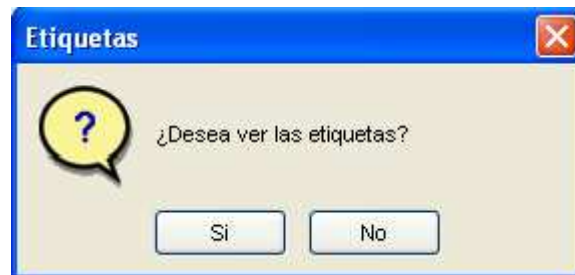


Figura A3.37 Vista del cuadro de diálogo de ver etiquetas

Está compuesto por un logotipo, una etiqueta con la pregunta “¿Desea ver las etiquetas?” y dos botones que ofrecen dos respuestas (“Si” y “No”).

En la siguiente tabla se puede ver las acciones que conlleva activar alguno de los controles que tiene este cuadro de diálogo.

Botón Si	Cierra el cuadro de diálogo Devuelve la cadena “Si”
Botón No	Cierra el cuadro de diálogo Devuelve la cadena “No”

Tabla A3.12 controles del cuadro de diálogo de ver etiquetas

CUADRO DE DIÁLOGO DE FICHERO DE ETIQUETAS

El cuadro de dialogo que permite decidir si se sobrescribe el fichero de etiquetas o si añaden etiquetas al existente el que se muestra en la siguiente figura:

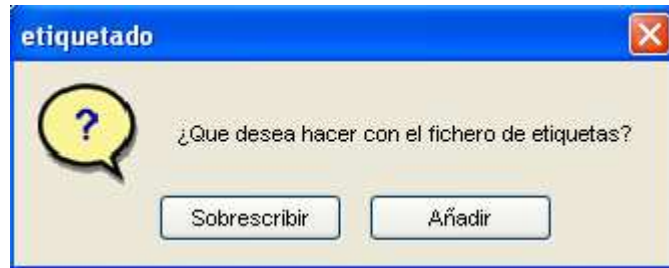


Figura A3.38 Vista del cuadro de diálogo de fichero de etiquetas

Está compuesto por un logotipo, una etiqueta que pregunta la operación que hay que realizar y dos botones que ofrecen dos respuestas (“Sobrescribir” y “Añadir”).

En la siguiente tabla se puede ver las acciones que conlleva activar alguno de los controles que tiene este cuadro de diálogo:

Botón Sobrescribir	Cierra el cuadro de diálogo Devuelve la cadena “Sobrescribir”
Botón Añadir	Cierra el cuadro de diálogo Devuelve la cadena “Añadir”

Tabla A3.13 controles del cuadro de diálogo de fichero de etiquetas

A3.6 Diseño procedimental

El diseño de los algoritmos que componen el sistema es uno de los problemas más importantes a los que hay que hacer frente en la fase de diseño del sistema, ya que en función de la calidad de estos diseños, la calidad y eficiencia final del sistema será mayor o menor.

El objetivo de este punto es el de recoger los detalles más importantes de los algoritmos que componen el sistema, ya que hay que tener en cuenta que el máximo grado de detalle se alcanzará con la implementación de estos algoritmos en un código debidamente comentado.

Para poder visualizar los mencionados detalles de los algoritmos se empleará la técnica de los diagramas de flujo.

Los algoritmos que se van a especificar son los siguientes:

- Algoritmo endpoint: La misión de este algoritmo es la de detectar el inicio y final de las palabras contenidas en una señal de voz, el algoritmo empleado en SERAL esta basado en el propuesto por Lamel en [Lamel, Rabiner, Rosenberg y Wilpon, 1991].
- Algoritmo parámetros: La misión de este algoritmo es la de calcular los parámetros de cada trama en las que se divide una señal de voz.
- Algoritmo zoommas: La misión de este algoritmo es la de mostrar una señal de voz con más nivel de detalle.
- Algoritmo zoommenos: La misión de este algoritmo es la de mostrar una señal de voz con menos nivel de detalle.
- Algoritmo zoomintervalo: La misión de este algoritmo es la de mostrar una región de una señal de voz definida por el usuario.
- Algoritmo reproduceintervalo: La misión de este algoritmo es la de reproducir una región de una señal de voz definida por el usuario.

Proyecto Fin de Carrera
 -SERAL- Sistema Experimental de Reconocimiento Automático de Locutor
ALGORITMO ENDPOINT

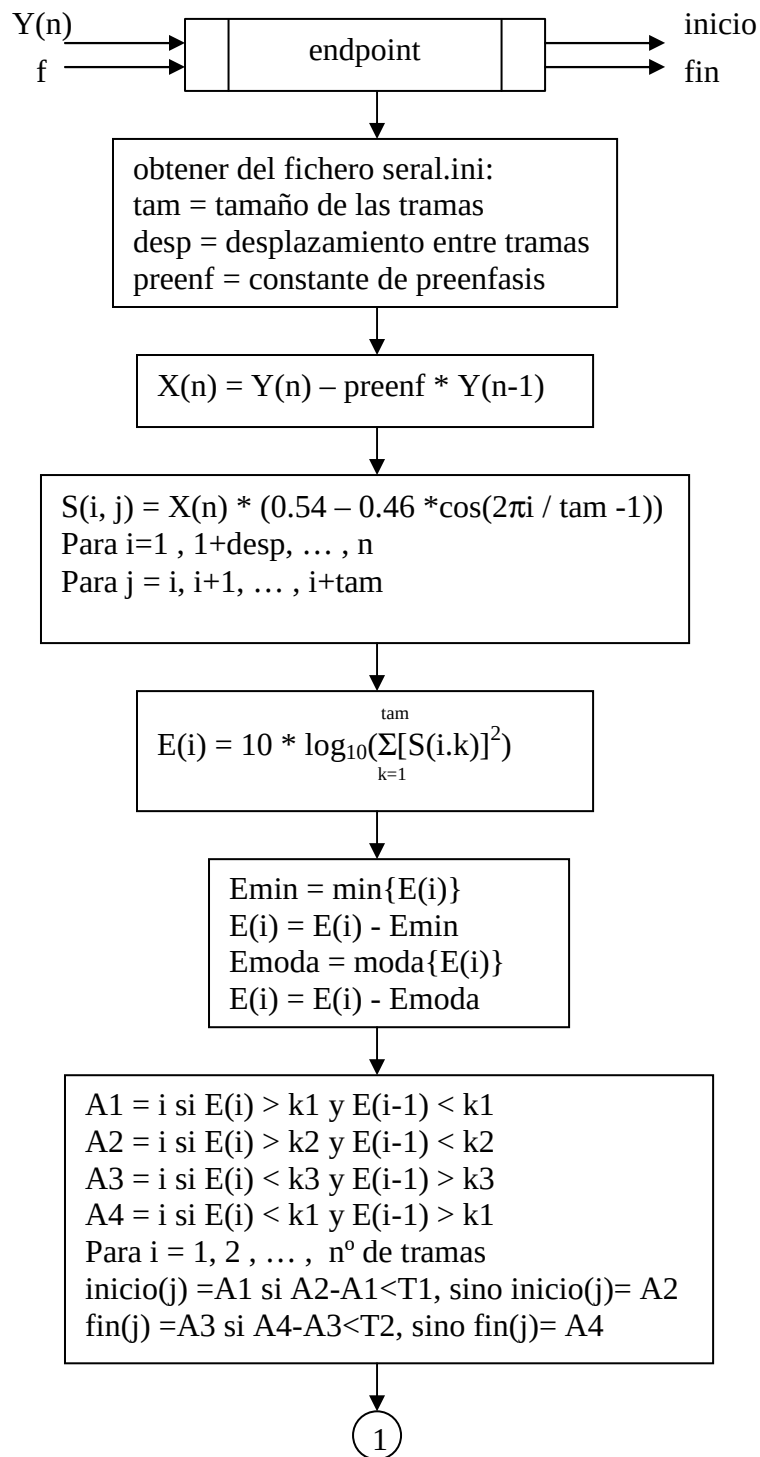


Figura A3.39 Algoritmo ENDPOINT

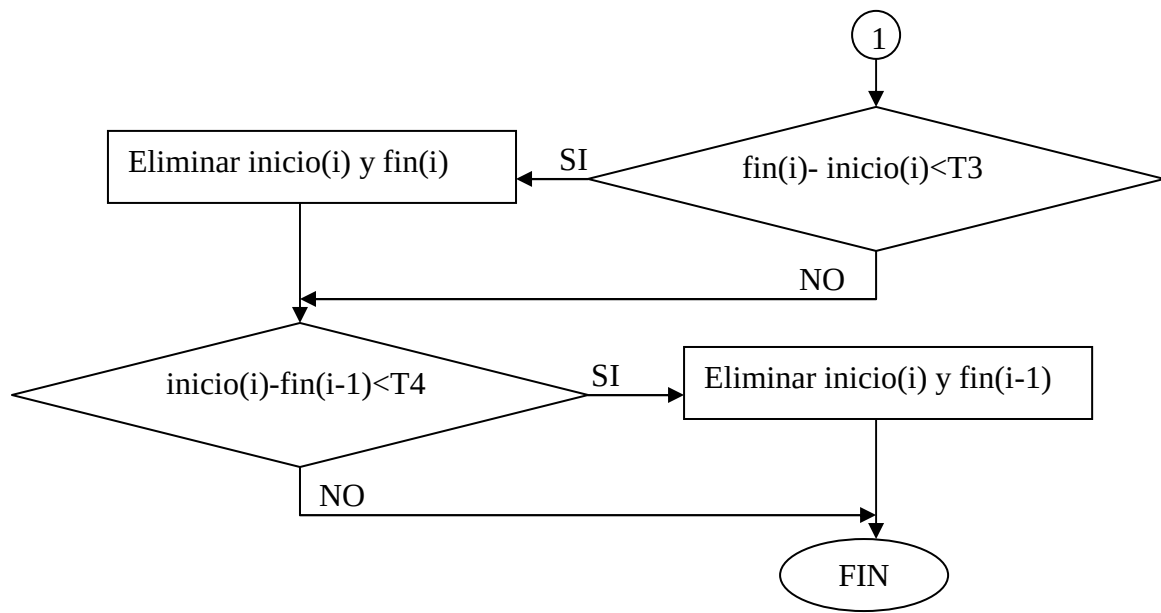


Figura A3.40 Algoritmo ENDPOINT (continuación)

De este algoritmo cabe destacar que la entrada $Y(n)$ es un vector de enteros que contiene la señal de voz.

La entrada f , es la frecuencia con la que se muestreo la señal de voz.

Las salidas inicio y fin son dos vectores de enteros que contiene el número de trama donde comienza y termina una palabra de la señal de voz.

ALGORITMO PARAMETROS

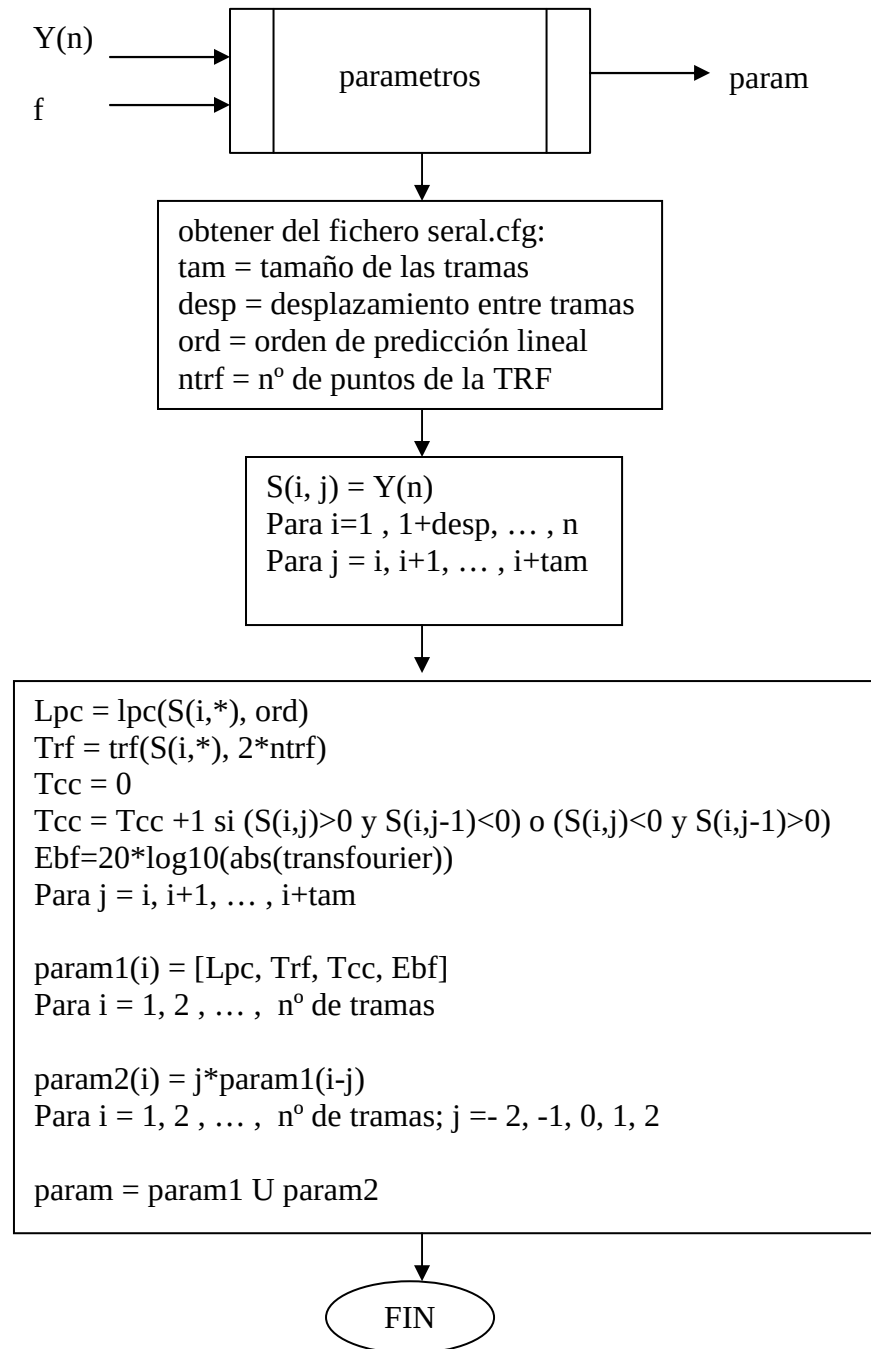


Figura A3.41 Algoritmo PARAMETROS

De este algoritmo cabe destacar que la entrada $Y(n)$ es un vector de enteros que contiene la señal de voz.

La entrada f , es la frecuencia con la que se muestreo la señal de voz.

La salida $param$ se una matriz que contiene los parámetros de todas las tramas en las que se ha dividido la señal de voz.

ALGORITMO ZOOMMAS

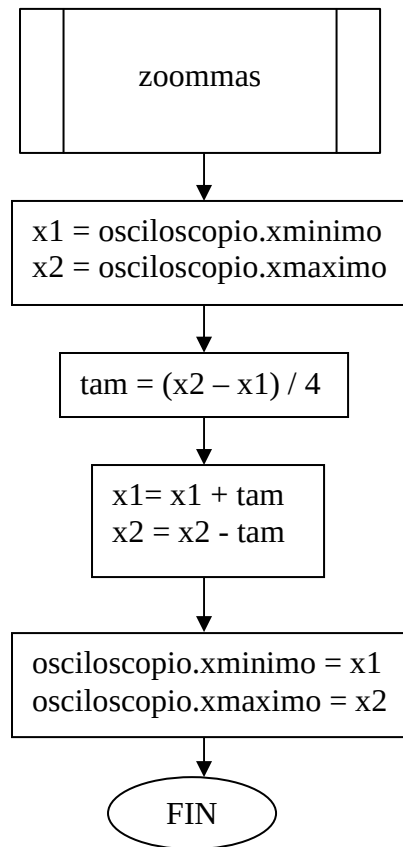


Figura A3.42 Algoritmo ZOOMMAS

De este algoritmo cabe destacar que `osciloscopio.xminimo` y `osciloscopio.xmaximo` hacen referencia a dos propiedades del objeto `osciloscopio` que contiene la visualización de la señal de voz.

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor
ALGORITMO ZOOMMENOS

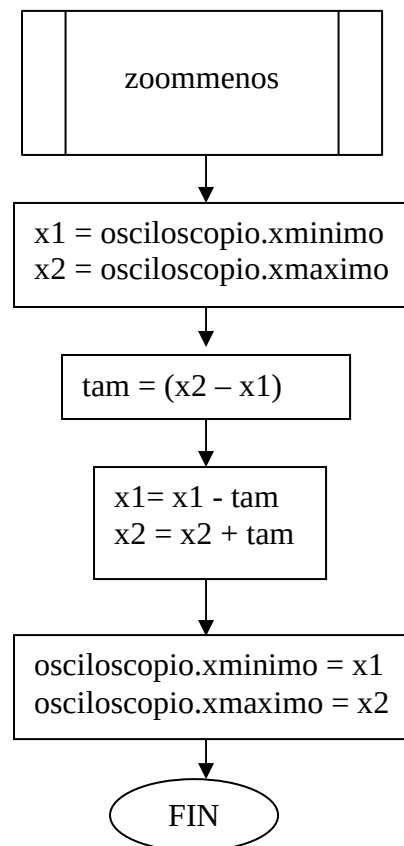


Figura A3.43 Algoritmo ZOOMMENOS

De este algoritmo cabe destacar que `osciloscopio.xminimo` y `osciloscopio.xmaximo` hacen referencia a dos propiedades del objeto `osciloscopio` que contiene la visualización de la señal de voz.

ALGORITMO ZOOMINTERVALO

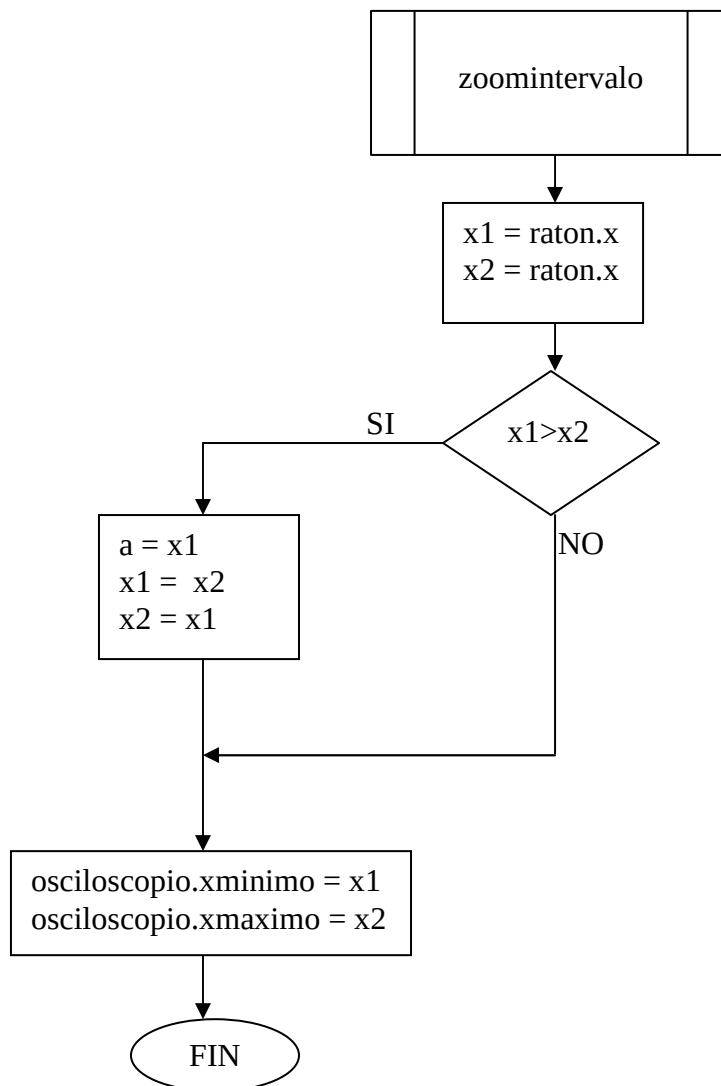


Figura A3.44 Algoritmo ZOOMINTERVALO

De este algoritmo cabe destacar que `raton.x` hace referencia a la coordenada x del ratón en el momento de hacer clic con el botón izquierdo y que `osciloscopio.xminimo` y `osciloscopio.xmaximo` hacen referencia a dos propiedades del objeto osciloscopio que contiene la visualización de la señal de voz.

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor
ALGORITMO REPRODUCEINTERVALO

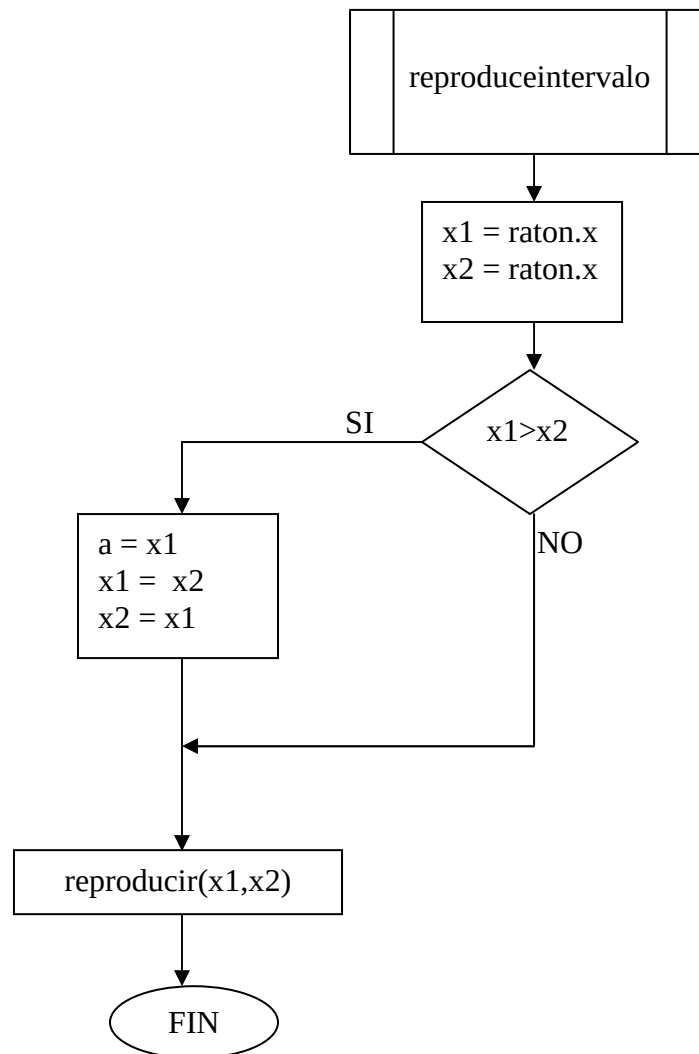


Figura A3.45 Algoritmo REPRODUCEINTERVALO

De este algoritmo cabe destacar que `raton.x` hace referencia a la coordenada x del ratón en el momento de hacer clic con el botón.

A3.7 Referencia cruzada a los requisitos

Uno de los objetivos planteados en este anexo es el realizar un diseño que satisfaga los requisitos planteados en el apartado A2.4 (Especificación de los requisitos) contenido en el anexo 2.

Para determinar el grado en el que satisfacen dichos objetivos se empleará una técnica matricial en la que por un lado se ponen los requisitos en las filas y los distintos módulos en los que se dividió el problema en las columnas.

En este caso los módulos que aparecen en las columnas se corresponden a cada uno de los procesos que aparecen en el DFD de nombre 0 SERAL (figura A3.8)

En la matriz se escribirá una X en la casilla que relaciona un requisito y módulo que lo satisface.

Antes de plantear la matriz, conviene recordar los requisitos funcionales del proyecto, que eran los siguientes:

1. Entrenamiento del sistema
2. Verificación de locutor
3. Identificación de locutor
4. Eliminación de locutor
5. Ver estadísticas
6. Borrar estadísticas
7. Grabar voz
8. Reproducir voz
9. Visualizar señal de voz
10. Cambiar la configuración

También es conveniente saber que los módulos en los que se divide el sistema son los siguientes:

1. Entrenar
2. Procesar señal de voz
3. Eliminar locutor
4. Visualizar señal de voz
5. Reproducir señal de voz
6. Escribir fichero de voz
7. Leer fichero de voz
8. Ver estadística
9. Borrar estadística
10. Actualizar configuración

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

módulo requisito	entrenar	procesar señal de voz	eliminar locutor	visualizar señal de voz	reproducir señal de voz	escribir fichero de voz	leer fichero de voz	ver estadística	borrar estadística	actualizar configuración
Entrenamiento del sistema	X									
Verificación del locutor		X								
Identificación del locutor		X								
Eliminación del locutor			X							
Ver estadísticas								X		
Borrar estadísticas									X	
Grabar voz						X				
Reproducir voz					X		X			
Visualizar señal de voz				X						
Cambiar la configuración										X

Tabla A3.14 Referencia cruzada a los requisitos

De lo que se observa en la tabla anterior, se deduce que todos los requisitos funcionales planteados inicialmente en el anexo 2, son satisfechos mediante el diseño especificado en este anexo y más concretamente en el punto A3.4 de este anexo, porque todas las filas (donde se representan los requisitos) tienen alguna X en alguna columna (donde se representan los módulos).

A pesar de que el diseño es correcto, cabe destacar la presencia de ciertos módulos de vital importancia, ya que son los únicos capaces de satisfacer alguno de los requisitos, estos módulos son los denominados “críticos”.

Estos módulos críticos, implican que a la hora de su codificación hay que tener un especial cuidado en su correcto funcionamiento.

La lista de módulos críticos es la siguiente:

1. Entrenar
2. Procesar señal de voz
3. Eliminar locutor
4. Visualizar señal de voz
5. Escribir fichero de voz
6. Ver estadística
7. Borrar estadística
8. Actualizar configuración

A3.8 Pruebas

Un aspecto importante a tener en cuenta a la hora de diseñar un sistema, es definir las pruebas individuales a las que se va a someter cada uno de los módulos desarrollados, así como las pruebas de integración de dichos módulos en el sistema.

En este punto se recogerán las pautas a seguir en cada una de las mencionadas pruebas.

Pruebas individuales

Sobre cada uno de los módulos del sistema se aplicarán las cuatro pruebas que se describen a continuación, y en el orden en el que están descritas.

En el caso en el que uno de los módulos no supere alguna de las pruebas, este se revisará para corregir del error, y tendrá que volver a superar de nuevo todas las pruebas.

1. Prueba de sintaxis

Esta prueba consiste en revisar el código del módulo en cuestión con el objetivo de encontrar algún error de sintaxis que se pudiera haber cometido durante la escritura del código.

2. Prueba de funcionamiento correcto

Esta prueba consiste en ejecutar el módulo de forma aislada, introduciendo datos que son correctos, y comprobar que en comportamiento del módulo es el que se esperaba.

3. Prueba de retorno de datos

Esta prueba podría considerarse como la última etapa de la anterior, aunque se considera independiente ya que no se aplica sobre todos los módulos, sino que solo se realiza sobre aquellos que retornan algún tipo de dato.

En esta prueba se verifica que los datos retornados por el módulo después de una ejecución correcta, tienen los valores y la estructura adecuados.

4. Prueba de funcionamiento incorrecto

Durante esta prueba se ejecuta el módulo de forma aislada, introduciendo datos que son incorrectos, y se comprueba el tratamiento que hace el módulo de dichos datos y que no retorna ningún valor en el caso que tuviera que retornar alguno.

Pruebas de integración

Una vez que los módulos han superado las pruebas individuales, se integran en el sistema y se somete de nuevo a estos módulos a otras pruebas de integración que se aplicarán en el orden en el que están descritas.

En el caso en el que uno de los módulos no supere alguna de las pruebas, este se revisará para corregir del error, y tendrá que volver a superar de nuevo todas las pruebas.

1. Prueba de funcionamiento correcto

Esta prueba consiste en ejecutar el módulo integrado en el sistema, introduciendo datos que son correctos, y comprobar que en comportamiento del módulo es el que se esperaba.

2. Prueba de retorno de datos

Al igual que ocurría con las pruebas individuales, se podría considerar como la última etapa de la anterior.

En esta prueba se verifica que los datos retornados por el módulo al sistema después de una ejecución correcta, tienen los valores y la estructura adecuados.

3. Prueba de funcionamiento incorrecto

Durante esta prueba se ejecuta el módulo integrado en el sistema, introduciendo datos que son incorrectos, y se comprueba el tratamiento que hace el módulo de dichos datos y que no retorna ningún valor en el caso que tuviera que retornar alguno.

A3.9 Entorno tecnológico del sistema

En este punto se describirán el entorno tanto físico como lógico necesario para un correcto funcionamiento del sistema, también serán tendidas en cuenta las restricciones que el sistema impone al entorno.

En el apartado de entorno físico se describirán aspectos como cantidad de memoria, dispositivos de Entrada/Salida etc.

Dentro del apartado de entorno lógico se describirán aspectos como Sistema Operativo, u otro software necesario.

ENTORNO FÍSICO

El equipo físico recomendado que necesita el sistema estará compuesto por la siguiente lista de elementos:

1. 512 MB de memoria RAM
2. 10 MB de disco duro
3. CPU de 3 GHz
4. Tarjeta de sonido de 16 bits
5. Micrófono

ENTORNO LÓGICO

El entorno lógico que necesita el sistema estará formado por la siguiente lista de elementos:

1. Sistema Operativo Microsoft Windows XP
2. MatLab 7.0
3. editor notepad
4. navegador web capaz de abrir ficheros con formato .chm

RESTRICCIONES

Una vez descrito el entorno tanto físico como lógico del sistema, hay que destacar las restricciones que existen sobre este para que el sistema pueda funcionar de una forma adecuada.

Estas restricciones afectan a los puntos 4 y 5 del entorno físico.

Las restricciones impuestas sobre la tarjeta de sonido son:

- soporte para Entrada/Salida de 8 y 16 bits
- soporte de una frecuencia de muestreo superior a 48 KHz

Las restricciones impuestas sobre el micrófono son:

- micrófono unidireccional
- Sensibilidad a frecuencias entre 20 Hz y 20 KHz

A3.10 Plan de desarrollo e implantación

Este sistema como cualquier otro requiere una implantación que requiere seguir una serie de pasos.

Para comprender mejor el proceso de implantación del sistema, conviene conocer como se distribuyen los datos que utiliza el sistema en diferentes directorios.

El árbol de directorios del sistema es el que se presenta en la siguiente figura:



Figura A3.46 Árbol de directorios del sistema

Dentro del directorio “audio”, se encontrarán todos los ficheros de voz que el sistema tenga almacenado de los distintos locutores.

Dentro del directorio “bin”, se encontrarán los ficheros fuentes necesarios para ejecutar el sistema.

Dentro del directorio “data” se encuentran los ficheros de datos del sistema como son:

- Fichero de estadísticas
- Fichero de patrones de voz
- Fichero de configuración
- Fichero con el logo del sistema
- Ficheros de etiquetas.

Dentro del directorio “temp” se encuentran los ficheros temporales que necesita el sistema, estos ficheros temporales se eliminarán al cerrar la aplicación.

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

Una vez descrito el árbol de directorios del sistema, se describirá a continuación el proceso de implantación del sistema.

El primer paso para la implantación del sistema es el de instalar todo el hardware y software necesario (descrito en el punto A3.9 de este anexo) para el funcionamiento del sistema.

El siguiente paso es el de crear el árbol de directorios anteriormente mencionado, con el contenido correspondiente.

El último paso necesario para poder utilizar el sistema es el de crear los patrones de voz, aunque esta es ya una tarea que se realiza desde el propio sistema.

Acrónimos

RNA	Redes Neuronales Artificiales
ASCII	American Standard Code for Information Interchange
DFD	Diagramas de Flujo de Datos
MB	MegaByte
RAM	Random Access Memory
CPU	Central Unit Processor

UNIVERSIDAD DE SALAMANCA

Escuela Politécnica Superior de Zamora

Ingeniería Técnica Informática de Gestión

Departamento de Informática y Automática, Área de Lenguajes y sistemas informáticos



Anexo IV: Documentación técnica de programación

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

Índice Anexo 4: Documentación técnica de programación

Lista de tablas	157
Lista de figuras	158
A4.1 Introducción.	160
A4.2 Documentación de las bibliotecas	161
A4.3 Código fuente.	165
A4.4 Manual del programador	184
A4.5 Pruebas unitarias	185
Acrónimos.	187

Lista de tablas

Tabla A4.1. función endpoint.	165
Tabla A4.2. función parametros.	166
Tabla A4.3. función reproducir	167
Tabla A4.4. función reproducirinterval	168
Tabla A4.5. función zoommas	169
Tabla A4.6. función zoommenos	170
Tabla A4.7. función zoominterval	171
Tabla A4.8. función vertodo.	172
Tabla A4.9. función grabar.	173
Tabla A4.10. función crearred	174
Tabla A4.11. función entrenar	175
Tabla A4.12. función simular.	176
Tabla A4.13. función bordes	177
Tabla A4.14. función parametros2.	178
Tabla A4.15. función parametros3.	179
Tabla A4.16. fichero seral	180
Tabla A4.17. fichero acercade.	180
Tabla A4.18. fichero dialogo.	181
Tabla A4.19. fichero reconocimiento	181
Tabla A4.20. fichero verestadística.	182
Tabla A4.21. fichero nombrarloctur	182
Tabla A4.22. fichero mostrarmensaje	182
Tabla A4.23. fichero etiquetar.	183
Tabla A4.24. fichero etiquetado	183

Lista de figuras

Figura A4.1 Cambio del directorio actual.	184
Figura A4.2 Ejecución del sistema	184
Figura A4.3 Descenso del error en las RNA.	185
Figura A4.4 Curva DET	186

A4.1 Introducción

El anexo 4 (Documentación Técnica de Programación), pretende ser un documento que recoja los aspectos más importantes del código fuente de la aplicación, así como de los elementos en los que se apoya dicho código fuente y los pasos necesarios para ejecutarlo.

Hay que destacar que para la realización del sistema se ha empleado el entorno de desarrollo MATLAB, que ha sido descrito en el punto 4.2 de esta memoria.

Este anexo está formado por cuatro apartados aparte del actual cuyos contenidos son los siguientes:

A4.2 Documentación de las bibliotecas

Contiene una descripción de las bibliotecas empleadas para el desarrollo del sistema final así como de las funciones de dichas bibliotecas que fueron utilizadas.

A4.3 Código fuente

Contiene una descripción del funcionamiento de cada uno de los módulos en los que se descompone el sistema, así como un pseudo-código de los mismos.

A4.4 Manual del programador

Contiene una descripción de las operaciones necesarias para conseguir el producto final y poder ejecutarlo.

A4.5 Pruebas unitarias

Contiene una descripción de las pruebas realizadas sobre el sistema.

A4.2 Documentación de las bibliotecas

Para la obtención del producto final, no se ha desarrollado ninguna biblioteca de funciones, sin embargo si se han utilizado tres bibliotecas (conocidas como Toolboxes), que trae incorporadas MATLAB (que fueron brevemente descritas en el apartado 4.2 de esta memoria).

Las bibliotecas utilizadas son las siguientes:

- Image Processing Toolbox
- Signal Processing Toolbox
- Neural Network Toolbox

El propósito que persigue este apartado es el de describir a modo de guía de referencia, la forma de utilizar las mencionadas bibliotecas, sin embargo, debido a la gran extensión que presenta cada una de las bibliotecas y a que para el desarrollo del sistema final sólo se han empleado una pequeña parte de las funciones que tiene cada una, sólo se describirán las partes de cada biblioteca que realmente han sido utilizadas.

Image Processing Toolbox

Esta biblioteca nos permite la adquisición de imágenes, así como el procesado de las mismas.

Las funciones empleadas de esta biblioteca son las siguientes:

Función `imread`

Esta función permite leer imágenes almacenadas en ficheros gráficos

`A = imread(filename,fmt)`

donde: A es la imagen devuelta por la función

filename es el nombre del fichero que contiene la imagen

fmt es el formato del fichero

Si el fichero no se encuentre en el directorio actual o en un directorio que forma parte del path de MATLAB, hay que especificar la ruta completa

Si no se encuentre ningún fichero con el nombre filename, se buscará un fichero cuyo nombre sea filename.fmt.

La lista de formatos gráficos que soporta la función es la siguiente:

- | | |
|---------|-----------------|
| 1 .bmp | 2 .cur |
| 3 .gif | 4 .hdf |
| 5 .ico | 6 .jpg o .jpeg |
| 7 .pbm | 8 .pcx |
| 9 .pgm | 10 .png |
| 11 .pnm | 12 .ppm |
| 13 .ras | 14 .tif o .tiff |
| 15 .xwd | |

Signal Processing Toolbox

Esta biblioteca permite el procesado de señales, en particular, en este proyecto tienen especial interés las señales de voz.

Las funciones empleadas de esta biblioteca son las siguientes:

Función abs

Esta función calcula la magnitud de un número complejo

$$Y = \text{abs}(X)$$

donde Y es la magnitud del número complejo.
X es el número complejo.

Si X no es un número complejo se supone $X = X + 0i$

Dado $X = a + bi$ la función es equivalente a: $Y = \sqrt{a^2 + b^2}$

Función fft

Esta función permite calcular la transformada rápida de Fourier de una señal

$$Y = \text{fft}(X, n)$$

donde Y es la transformada rápida de Fourier.
X es la señal.

n es el número de puntos con el que se calcula la transformada rápida de Fourier.

Si la longitud del vector X donde está almacenada la señal es menor que n, X es ampliado hasta una longitud n con 0.

Si la longitud del vector X donde está almacenada la señal es mayor que n, X es trucado hasta n.

Si X es una matriz, la longitud de las columnas es ajustada de la misma forma.

Función lpc

Esta función calcula los coeficientes de predicción lineal de una señal, mediante el algoritmo de autocorrelación.

$$[a, g] = \text{lpc}(x, p)$$

donde a son los coeficientes de predicción lineal
g es la ganancia
x es la señal
p es el orden de predicción lineal

Si p no se especifica, se considera un p igual a la longitud de x menos 1.

Si x es una matriz, se calcula a y g para cada una de sus columnas.

Neural Network Toolbox

Esta biblioteca esta diseñada para el trabajo con redes neuronales artificiales, permitiendo gran cantidad de operaciones sobre ellas.

Las funciones empleadas de esta biblioteca son las siguientes:

Función newff

Esta función crea una red que utiliza el algoritmo de retropropagación del error

`net = newff(PR,[S1 S2...SNl},{TF1 TF2...TFNl},BTF,BLF,PF)`

donde net es la red neuronal

PR es una matriz con el mínimo y máximo valor que pueden tomar las entradas

Si es el número de neuronas de la i-ésima capa

TFi es la función de transferencia de la i-ésima capa

BTF es la función de aprendizaje

BLF es la función de ajuste de los pesos de las conexiones entre neuronas

PF es la función de representación de la neurona

TFi por defecto es tansig, pero puede tomar otros valores como logsig, purelin, etc.

BTF por defecto es traingdx, pero puede tomar otros valores como trainbfg, trainrp, etc.

El valor por defecto de BLF es learnqdm, pero puede tomar otros valores como learncon, leqrn, etc.

El valor por defecto de PF es mse pero puede tomar otros valores como mae, sse, etc.

Función train

Esta función permite entrenar la red neuronal basandose en el algoritmo introducido en el parámetro BTF de la función newff.

`net = train(net,P,T)`

donde net (el valor devuelto) es la red ya entrenada

net (el parámetro) es la red a entrenar

P son los parámetros de entrenamiento

T son los objetivos a alcanzar con el entrenamiento

Para que el entrenamiento se efectúe correctamente hay que dar valores a las siguientes propiedades de net:

net.trainParam.epochs: Número de iteraciones del algoritmo (por defecto 10)

net.trainParam.goal: Error con el que se detiene el algoritmo (por defecto 0)

net.trainParam.lr: Tasa de aprendizaje (por defecto 0.01)

net.trainParam.time: tiempo máximo de entrenamiento (por defecto infinito)

Función sim

Esta función simula el comportamiento de una red neuronal artificial a partir de un conjunto de entradas.

$Y = \text{sim}(\text{net}, P)$

donde Y son las salidas que da la red después de simularla

net es la red que se simulará

P son las entradas con las que se simulará la red

A4.3 Código fuente

El código fuente desarrollado con el presente proyecto está organizado en diferentes ficheros, siguiendo la filosofía de MATLAB donde cada fichero se corresponde con una función.

A continuación se incluye una tabla con la información más relevante a cerca de cada una de las funciones que componen el código fuente.

FUNCIÓN ENDPOINT:

Nombre	endpoint
Sinopsis	La función endpoint, es una función que dada una señal de voz y su correspondiente frecuencia de muestreo devuelve el número de las tramas en las que empiezan y terminan las palabras que contiene la señal de voz
Ejemplo de llamada	[inicio,fin]=endpoint(señal,frecuencia);
Argumentos de entrada	señal = señal de voz frecuencia = frecuencia de muestreo de la señal
Valores de retorno	inicio = tramas en las que comienza una palabra fin = tramas en las que termina una palabra
Pseudocódigo	FUNCION endpoint RECEIVE señal = señal de voz frecuencia = frecuencia de muestreo de la señal DEVUELVE inicio = tramas en las que comienza una palabra fin = tramas en las que termina una palabra INICIO obtener el tamaño de las tramas del fichero de configuración. obtener la distancia entre tramas del fichero de configuración. obtener la constante de preenfasis del fichero de configuración. preenfatizar la señal. segmentar la señal en tramas y aplicar una ventana de Haming. obtener la energía de la señal. normalizar la energía de la señal. detectar los posibles puntos de inicio y fin según el nivel de energía. asociar los puntos de inicio y fin suficientemente cercanos. FIN
Autor	Manuel Montero Villar
Fecha	29 marzo 2007
Versión	2.0

Tabla A4.1. función endpoint

FUNCIÓN PARAMETROS:

Nombre	parametros
Sinopsis	La función parametros, es una función que dada una señal de voz y su correspondiente frecuencia de muestreo devuelve una matriz con los parámetros de cada una de las tramas en las que se dividió la señal de voz.
Ejemplo de llamada	p = parametros(señal,frecuencia);
Argumentos de entrada	señal = señal de voz frecuencia = frecuencia de muestreo de la señal
Valores de retorno	p = matriz con de parámetros de cada trama
Pseudocódigo	FUNCION parametros RECIBE señal = señal de voz frecuencia = frecuencia de muestreo de la señal DEVUELVE p = matriz con de parámetros de cada trama INICIO obtener el tamaño de las tramas del fichero de configuración. obtener la distancia entre tramas del fichero de configuración. obtener el orden de predicción lineal del fichero de configuración. obtener el número de puntos de la TRF del fichero de configuración. segmentar la señal en tramas calcular los LPC de cada trama calcular la TRF de cada trama calcular la tasa de cruces por cero de cada trama calcular la EBF añadir los LPC a la matriz. añadir la TRF a la matriz. añadir la tasa de cruces por cero a la matriz. añadir la EBF a la matriz calcular los parámetros delta para todos los elementos de la matriz FIN
Autor	Manuel Montero Villar
Fecha	1 abril 2007
Versión	3.0

Tabla A4.2. función parametros

FUNCIÓN REPRODUCIR:

Nombre	reproducir
Sinopsis	La función reproducir, es una función que dado el nombre de un fichero de voz, guarda su contenido en un fichero temporal, lo reproduce y dibuja la señal de voz que contiene en la pantalla.
Ejemplo de llamada	reproducir(nombre);
Argumentos de entrada	nombre = nombre de un fichero de voz
Valores de retorno	SIN RETORNO
Pseudocódigo	FUNCION reproducir RECIBE nombre = nombre de un fichero de voz DEVUELVE SIN RETORNO INICIO Leer el fichero. Guardar su contenido en un fichero temporal. Reproducir la señal de voz. Visualizar la señal de voz. FIN
Autor	Manuel Montero Villar
Fecha	23 marzo 2007
Versión	1.0

Tabla A4.3. función reproducir

FUNCIÓN REPRODUCIRINTERVAL:

Nombre	reproducirinterval
Sinopsis	La función reproducirinterval, es una función que dada dos puntos de una señal de voz seleccionados con el ratón reproduce la señal de voz comprendida entre ellos.
Ejemplo de llamada	reproducirinterval();
Argumentos de entrada	SIN ENTRADA
Valores de retorno	SIN RETORNO
Pseudocódigo	FUNCION reproducirinterval RECIBE SIN ENTRADA DEVUELVE SIN RETORNO INICIO obtener el punto origen de la reproducción. obtener el punto final de la reproducción. Si los puntos están desordenados ordenar los puntos. Fin Si los puntos están fuera de la señal introducir los puntos en la señal Fin reproducir el rango seleccionado FIN
Autor	Manuel Montero Villar
Fecha	11 abril 2007
Versión	3.0

Tabla A4.4. función reproducirinterval

FUNCIÓN ZOOMMAS:

Nombre	zoommas
Sinopsis	La función zoommas, es una función que busca el punto central de la señal de voz, muestra en pantalla el 25% de la señal alrededor de ese punto.
Ejemplo de llamada	zoommas();
Argumentos de entrada	SIN ENTRADA
Valores de retorno	SIN RETORNO
Pseudocódigo	FUNCION zoommas RECIBE SIN ENTRADA DEVUELVE SIN RETORNO INICIO obtener el punto central de la señal de voz. calcular los nuevos rangos de visualización a partir de los actuales. Si los rangos están fuera de la señal corregir los rangos Fin obtener la señal de voz hasta los nuevos rangos establecer los nuevos rangos en el osciloscopio superior dibujar la señal en el osciloscopio inferior, cambiando de color la zona de la señal que no se ve en el superior delimitar la zona de la señal que se ve en el osciloscopio superior con dos líneas FIN
Autor	Manuel Montero Villar
Fecha	11 abril 2007
Versión	3.0

Tabla A4.5. función zoommas

FUNCIÓN ZOOMMENOS:

Nombre	zoommenos
Sinopsis	La función zoommenos, es una función que busca el punto central de la señal de voz, muestra en pantalla el 200% de la señal alrededor de ese punto.
Ejemplo de llamada	zoommenos();
Argumentos de entrada	SIN ENTRADA
Valores de retorno	SIN RETORNO
Pseudocódigo	FUNCION zoommenos RECIBE SIN ENTRADA DEVUELVE SIN RETORNO INICIO obtener el punto central de la señal de voz. calcular los nuevos rangos de visualización a partir de los actuales. Si los rangos están fuera de la señal corregir los rangos Fin obtener la señal de voz hasta los nuevos rangos establecer los nuevos rangos en el osciloscopio superior dibujar la señal en el osciloscopio inferior, cambiando de color la zona de la señal que no se ve en el superior delimitar la zona de la señal que se ve en el osciloscopio superior con dos líneas FIN
Autor	Manuel Montero Villar
Fecha	11 abril 2007
Versión	3.0

Tabla A4.6. función zoommenos

FUNCIÓN ZOOMINTERVAL:

Nombre	zoominterval
Sinopsis	La función zoominterval, es una función que dados dos puntos de una señal de voz, muestra en pantalla la zona la señal comprendida entre los puntos seleccionados.
Ejemplo de llamada	zoominterval();
Argumentos de entrada	SIN ENTRADA
Valores de retorno	SIN RETORNO
Pseudocódigo	FUNCION zoominterval RECIBE SIN ENTRADA DEVUELVE SIN RETORNO INICIO obtener los puntos de la señal de voz. calcular los nuevos rangos de visualización a partir los puntos. Si los rangos están desordenados ordenar los rangos Fin Si los rangos están fuera de la señal corregir los rangos Fin obtener la señal de voz hasta los nuevos rangos establecer los nuevos rangos en el osciloscopio superior dibujar la señal en el osciloscopio inferior, cambiando de color la zona de la señal que no se ve en el superior delimitar la zona de la señal que se ve en el osciloscopio superior con dos líneas. FIN
Autor	Manuel Montero Villar
Fecha	11 abril 2007
Versión	3.0

Tabla A4.7. función zoominterval

FUNCIÓN VERTODO:

Nombre	vertodo
Sinopsis	La función vertodo, es una función muestra en pantalla la señal completa.
Ejemplo de llamada	vertodo();
Argumentos de entrada	SIN ENTRADA
Valores de retorno	SIN RETORNO
Pseudocódigo	FUNCION vertodo RECIBE SIN ENTRADA DEVUELVE SIN RETORNO INICIO obtener la señal de voz. obtener el tamaño de la señal de voz. calcular los nuevos rangos de visualización a partir del tamaño. establecer los nuevos rangos en el osciloscopio superior dibujar la señal en el osciloscopio inferior. FIN
Autor	Manuel Montero Villar
Fecha	18 abril 2007
Versión	1.0

Tabla A4.8. función vertodo

FUNCIÓN GRABAR:

Nombre	grabar
Sinopsis	La función grabar, es una función que dada la ruta de un fichero, graba una muestra de voz por micrófono con la frecuencia indicada en el fichero de configuración y la graba en el fichero indicado.
Ejemplo de llamada	grabar(fichero);
Argumentos de entrada	fichero = ruta de un fichero
Valores de retorno	SIN RETORNO
Pseudocódigo	FUNCION grabar RECIBE fichero = ruta de un fichero DEVUELVE SIN RETORNO INICIO obtener la frecuencia de grabación del fichero de configuración. grabar la señal de voz por el micrófono. grabar la señal de voz en el fichero. grabar la señal de voz en un fichero temporal por defecto. mostrar la señal por pantalla FIN
Autor	Manuel Montero Villar
Fecha	11 abril 2007
Versión	3.0

Tabla A4.9. función grabar

FUNCIÓN CREAMED:

Nombre	creamed
Sinopsis	La función creamed, es una función que dado el nombre de un locutor, crea una red neuronal artificial para él, si no existe, y devuelve un valor indicando si se creó la red neuronal artificial o no.
Ejemplo de llamada	ret=creamed(nombre);
Argumentos de entrada	nombre = nombre del locutor
Valores de retorno	ret = valor que indica si se creo una red o no
Pseudocódigo	FUNCION creamed RECIBE nombre = nombre del locutor DEVUELVE ret = valor que indica si se creo una red o no INICIO se definen los mínimos y máximos de las entradas. Si el locutor no existe se crea una nueva red se guarda la red en el fichero ret = 1 sino ret = 0 Fin FIN
Autor	Manuel Montero Villar
Fecha	7 mayo 2007
Versión	2.0

Tabla A4.10. función creamed

FUNCIÓN ENTRENAR:

Nombre	entrenar
Sinopsis	La función entrenar, es una función que dado el nombre de un locutor, una matriz de parámetros y otra de objetivos, entrena la red neuronal artificial del locutor, para que con esos parámetros se obtengan los objetivos correspondientes.
Ejemplo de llamada	entrenar(nombre, param, obj);
Argumentos de entrada	nombre = nombre del locutor param = matriz de parámetros obj = matriz de objetivos
Valores de retorno	SIN RETORNO
Pseudocódigo	FUNCION entrenar RECIBE nombre = nombre del locutor param = matriz de parámetros obj = matriz de objetivos DEVUELVE SIN RETORNO INICIO obtener el nº de épocas del entrenamiento del fichero de configuración obtener la tasa de error del entrenamiento del fichero de configuración obtener la tasa de aprendizaje del entrenamiento del fichero de configuración Si el locutor existe cargar su red del fichero de redes entrenar la red del locutor guardar la red entrenada Fin FIN
Autor	Manuel Montero Villar
Fecha	7 mayo 2007
Versión	7.0

Tabla A4.11. función entrenar

FUNCIÓN SIMULAR:

Nombre	simular
Sinopsis	La función simular, es una función que dado el nombre de un locutor y una matriz de parámetros, simula la red neuronal artificial del locutor, y devuelve el resultado que dio la red para esa matriz de parámetros.
Ejemplo de llamada	res=simular(nombre, param);
Argumentos de entrada	nombre = nombre del locutor param = matriz de parámetros
Valores de retorno	res = resultado de la simulación
Pseudocódigo	FUNCION simular RECIBE nombre = nombre del locutor param = matriz de parámetros DEVUELVE res = resultado de la simulación INICIO ajusta la matriz de parámetros al tamaño adecuado Si el locutor existe cargar su red del fichero de redes simular la red con la matriz de parámetros obtener el resultado de la simulación guardar el resultado de la simulación en res Fin FIN
Autor	Manuel Montero Villar
Fecha	7 mayo 2007
Versión	3.0

Tabla A4.12. función simular

FUNCIÓN BORDES:

Nombre	bordes
Sinopsis	La función bordes, es una función que muestra en la pantalla el inicio y fin de cada palabra que forma la señal de voz que se está visualizando en el osciloscopio
Ejemplo de llamada	bordes();
Argumentos de entrada	SIN ENTRADA
Valores de retorno	SIN RETORNO
Pseudocódigo	FUNCION simular RECIBE SIN ENTRADA DEVUELVE SIN RETORNO INICIO Obtiene la señal de voz que se está visualizando Obtiene los puntos de inicio y fin de las palabras Pinta en el osciloscopio unas barras delimitando las palabras FIN
Autor	Manuel Montero Villar
Fecha	17 mayo 2007
Versión	3.0

Tabla A4.13. función bordes

FUNCIÓN PARAMETROS2:

Nombre	parametros2
Sinopsis	La función parametros2, es una función que dada una señal de voz y su correspondiente frecuencia de muestreo devuelve una matriz con los parámetros LPC de cada una de las tramas en las que se dividió la señal de voz, necesarios para visualizar el espectrograma.
Ejemplo de llamada	p = parametros2(señal,frecuencia);
Argumentos de entrada	señal = señal de voz frecuencia = frecuencia de muestreo de la señal
Valores de retorno	p = matriz con de parámetros LPC de cada trama
Pseudocódigo	FUNCION parametros2 RECIBE señal = señal de voz frecuencia = frecuencia de muestreo de la señal DEVUELVE p = matriz con de parámetros LPC de cada trama INICIO obtener el tamaño de las tramas del fichero de configuración. obtener la distancia entre tramas del fichero de configuración. obtener el orden de predicción lineal del fichero de configuración. obtener el número de puntos de la TRF del fichero de configuración. segmentar la señal en tramas calcular los LPC de cada trama aproximar los LPC a la TRF. añadir los LPC a la matriz. FIN
Autor	Manuel Montero Villar
Fecha	15 junio 2007
Versión	3.0

Tabla A4.14. función parametros2

FUNCIÓN PARAMETROS3:

Nombre	parametros3
Sinopsis	La función parametros3, es una función que dada una señal de voz y su correspondiente frecuencia de muestreo devuelve una matriz con la TRF, la tasa de cruces por cero y la energía de cada una de las tramas en las que se dividió la señal de voz.
Ejemplo de llamada	[p, tcc, ener] = parametros3(señal,frecuencia);
Argumentos de entrada	señal = señal de voz frecuencia = frecuencia de muestreo de la señal
Valores de retorno	p = matriz con la TRF de cada trama tcc = tasa de cruces por cero ener = energía
Pseudocódigo	FUNCION parametros3 RECIBE señal = señal de voz frecuencia = frecuencia de muestreo de la señal DEVUELVE p = matriz con de parámetros de cada trama tcc = tasa de cruces por cero ener = energía INICIO obtener el tamaño de las tramas del fichero de configuración. obtener la distancia entre tramas del fichero de configuración. obtener el número de puntos de la TRF del fichero de configuración. segmentar la señal en tramas calcular la TRF de cada trama calcular la tasa de cruces por cero de cada trama calcular el nivel de energía de cada trama FIN
Autor	Manuel Montero Villar
Fecha	15 junio 2007
Versión	3.0

Tabla A4.15. función parametros3

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

A parte de los ficheros de código fuentes que contienen funciones utilizadas por la aplicación descritos anteriormente, existen otros ficheros de código fuente que contienen las GUI (Graphic User Interface) de la aplicación, que se describen a continuación:

Nombre	seral
Sinopsis	El fichero seral, contiene la interfaz principal de la aplicación, en él se describen las acciones a tomar cuando se pulsa un botón de la interfaz o un elemento del menú. Este fichero contiene también un osciloscopio donde se dibuja la señal de voz con la que se está trabajando.
Ejemplo de llamada	seral ();
Valores de retorno	SIN RETORNO
Organización	Departamento de Informática y Automática Universidad de Salamanca
Autor	Manuel Montero Villar
Fecha	23 marzo 2007
Versión	10.0

Tabla A4.16. fichero seral

Nombre	acercade
Sinopsis	El fichero acercade, contiene un cuadro de dialogo con información general de la aplicación que se presentará cuando se pulsa la opción acerca de ... del menú ayuda.
Ejemplo de llamada	acercade();
Valores de retorno	SIN RETORNO
Organización	Departamento de Informática y Automática Universidad de Salamanca
Autor	Manuel Montero Villar
Fecha	24 marzo 2007
Versión	2.0

Tabla A4.17. fichero acercade

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

Nombre	dialogo
Sinopsis	El fichero dialogo, contiene un cuadro de dialogo que presenta una pregunta de respuesta si o no según el título que se le pase de argumento. Es utilizado para preguntar si se quiere salir de la aplicación o si se quiere borrar el fichero de estadísticas de la aplicación.
Ejemplo de llamada	nombre = dialogo(titulo);
Valores de retorno	nombre = nombre del botón del cuadro de diálogo que se pulsó
Organización	Departamento de Informática y Automática Universidad de Salamanca
Autor	Manuel Montero Villar
Fecha	23 marzo 2007
Versión	1.0

Tabla A4.18. fichero dialogo

Nombre	reconocimiento
Sinopsis	El fichero reconocimiento, contiene un cuadro de dialogo que pregunta el modo en el que se realizará la operación de reconocimiento de locutor. Se utilizará tanto en la operación de verificación de locutor como en la de identificación de locutor.
Ejemplo de llamada	modo = reconocimiento()
Valores de retorno	modo = modo en el que se realizara el reconocimiento de locutor - por micrófono. - por fichero.
Organización	Departamento de Informática y Automática Universidad de Salamanca
Autor	Manuel Montero Villar
Fecha	25 marzo 2007
Versión	2.0

Tabla A4.19. fichero reconocimiento

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

Nombre	verestadistica
Sinopsis	El fichero verestadistica, contiene un cuadro de dialogo que muestra la información estadística de utilización de la aplicación. Se utilizará cuando se pulsa la opción ver del menú estadística.
Ejemplo de llamada	verestadistica();
Valores de retorno	SIN RETORNO
Organización	Departamento de Informática y Automática Universidad de Salamanca
Autor	Manuel Montero Villar
Fecha	27 marzo 2007
Versión	4.0

Tabla A4.20. fichero verestadistica

Nombre	nombrarlocutor
Sinopsis	El fichero nombrarlocutor, contiene un cuadro de dialogo que pide con un cuadro de texto el nombre de uno de los locutores del sistema. Es utilizado para obtener el nombre de un locutor para realizar un entrenamiento, un reconocimiento, o una eliminación de locutor.
Ejemplo de llamada	nombre = nombrarlocutor();
Valores de retorno	nombre = nombre del locutor.
Organización	Departamento de Informática y Automática Universidad de Salamanca
Autor	Manuel Montero Villar
Fecha	29 abril 2007
Versión	1.0

Tabla A4.21. fichero nombrarlocutor

Nombre	mostrarmensaje
Sinopsis	El fichero mostrarmensaje, contiene un cuadro de dialogo que permite mostrar al usuario un mensaje por parte del sistema. Es utilizado para presentar los resultados de las verificaciones y de las identificaciones, y para mostrar otros mensajes, como los de inicio de grabación.
Ejemplo de llamada	mostrarmensaje (mensaje);
Valores de retorno	SIN RETORNO
Organización	Departamento de Informática y Automática Universidad de Salamanca
Autor	Manuel Montero Villar
Fecha	1 mayo 2007
Versión	2.0

Tabla A4.22. fichero mostrarmensaje

Proyecto Fin de Carrera
-SERAL- Sistema Experimental de Reconocimiento Automático de Locutor

Nombre	etiquetar
Sinopsis	El fichero etiquetar, contiene un cuadro de dialogo que permite al usuario introducir una etiqueta en el sistema. Es utilizado para recoger las etiquetas que se quieren asociar a las señales de voz
Ejemplo de llamada	etiqueta = etiquetar ();
Valores de retorno	etiqueta = cadena de texto
Organización	Departamento de Informática y Automática Universidad de Salamanca
Autor	Manuel Montero Villar
Fecha	7 junio 2007
Versión	2.0

Tabla A4.23. fichero etiquetar

Nombre	etiquetado
Sinopsis	El fichero etiquetado, contiene un cuadro de dialogo que permite al usuario decidir si sobrescribe el fichero de etiquetas o añade una etiqueta al existente. Es utilizado al añadir etiquetas cuando las etiquetas existentes no se visualizan
Ejemplo de llamada	opción= etiquetado ();
Valores de retorno	opción = cadena de texto
Organización	Departamento de Informática y Automática Universidad de Salamanca
Autor	Manuel Montero Villar
Fecha	17 junio 2007
Versión	2.0

Tabla A4.24. fichero etiquetado

A4.4 Manual del programador

Al ser MATLAB un lenguaje interpretado, no hay que compilar el código fuente, para poder ejecutar el sistema, sino que los pasos a seguir son los siguientes:

1. Arrancar MATLAB y cambiar el directorio actual, poner como directorio actual aquel donde se encuentren los códigos fuente de la aplicación. Para ello hay que modificar la pestaña marcada en la figura A4.1

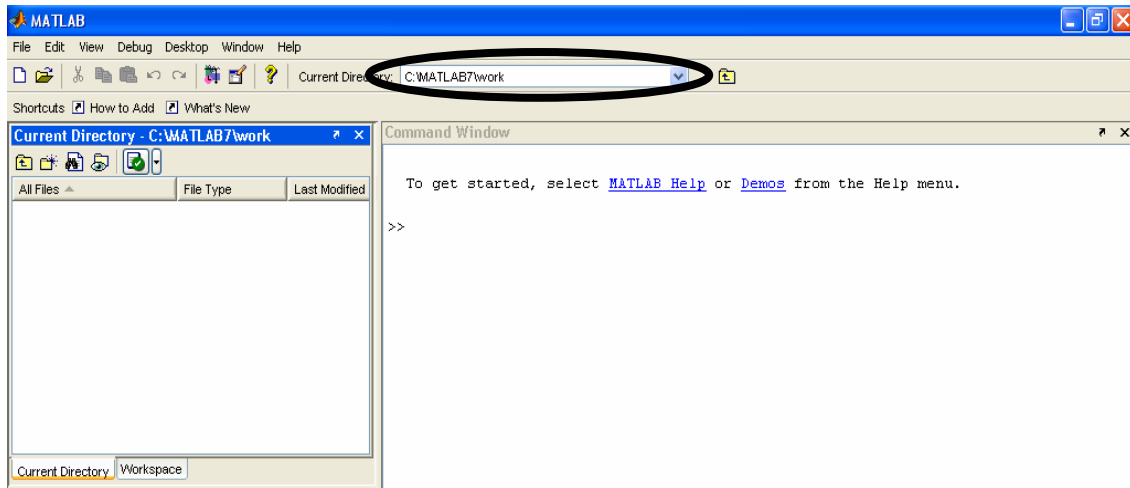


Figura A4.1 Cambio del directorio actual

2. Una vez cambiado el directorio actual, se ejecuta el fichero que contiene la ventana principal del sistema (en este caso seral.m) ejecutando el comando seral desde la línea de comandos de MATLAB.

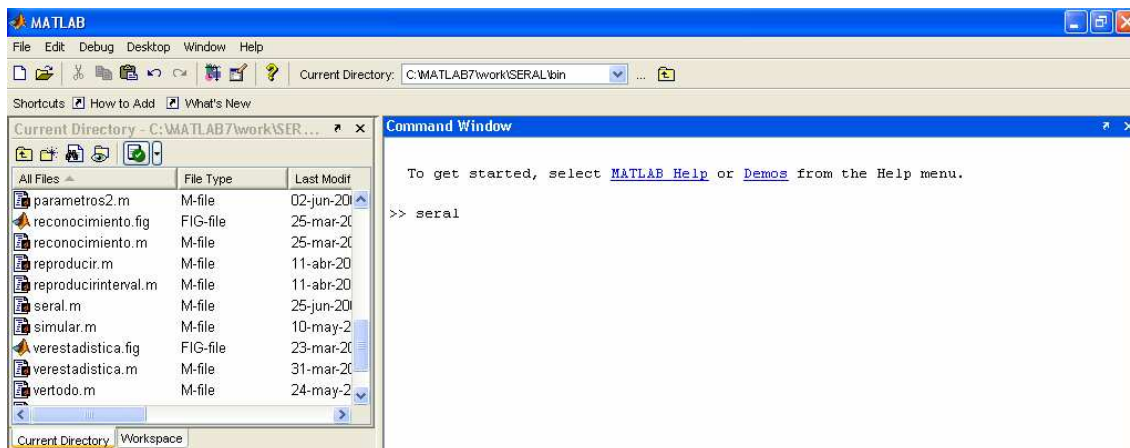


Figura A4.2 Ejecución del sistema

A4.5 Pruebas unitarias

En este punto del anexo cuatro, se detallarán los resultados obtenidos durante la realización de las pruebas del sistema.

Para realizar las pruebas del sistema se empleó un corpus formado por un conjunto de trece locutores diferentes, teniendo por cada locutor diez muestras de voz, que hacen un total de 130 muestras de voz diferentes.

ENTRENAMIENTO

Para la tarea de entrenamiento se emplearon la mitad de las muestras de voz que componen el corpus, es decir, cinco muestras de cada locutor.

En el entrenamiento se construye una Red Neuronal Artificial diferente por cada uno de los locutores.

Para la realización del entrenamiento, se utilizaron los siguientes parámetros:

- o Orden de predicción lineal: 12
- o Puntos en la Transformada Rápida de Fourier: 128
- o Número de etapas de entrenamiento: 22000
- o Error de aprendizaje: 0.001
- o Tasa de aprendizaje: 0.0001
- o Número de neuronas en la capa oculta: 256

El descenso del error existente en cada RNA, se produce de manera exponencial con cada etapa, como se muestra en la siguiente gráfica:

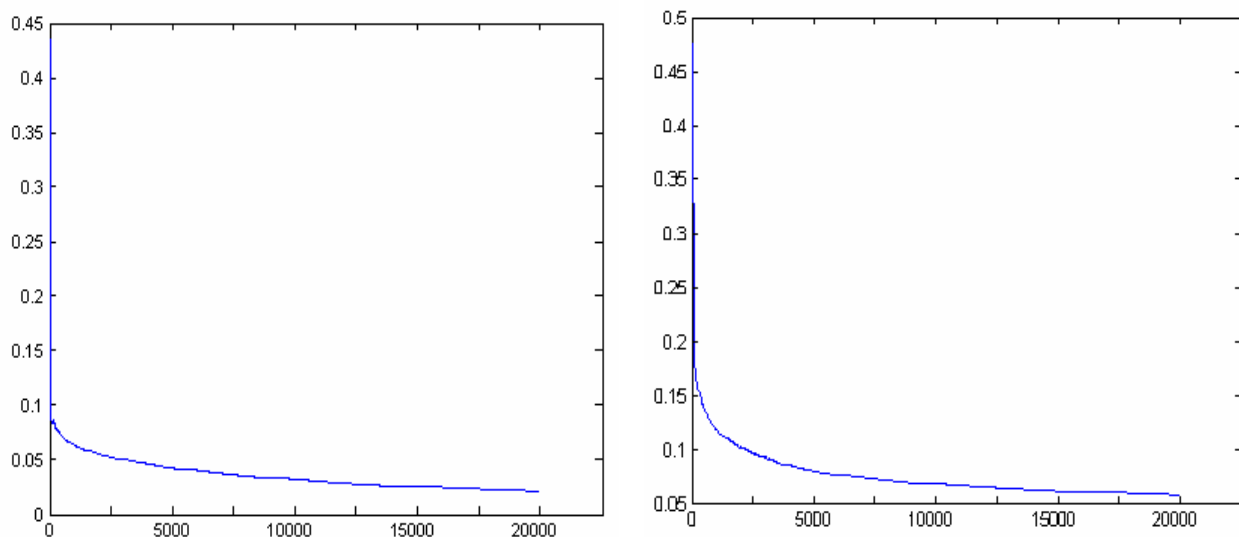


Figura A4.3 Descenso del error en las RNA

PRUEBAS

Para la realización de las pruebas, se emplearon las muestras de voz del corpus que no fueron utilizadas en el entrenamiento.

Para medir la eficiencia del sistema, se muestra la curva DET (Data Error Trade-off), que muestra la relación entre la TFA y la TFR:

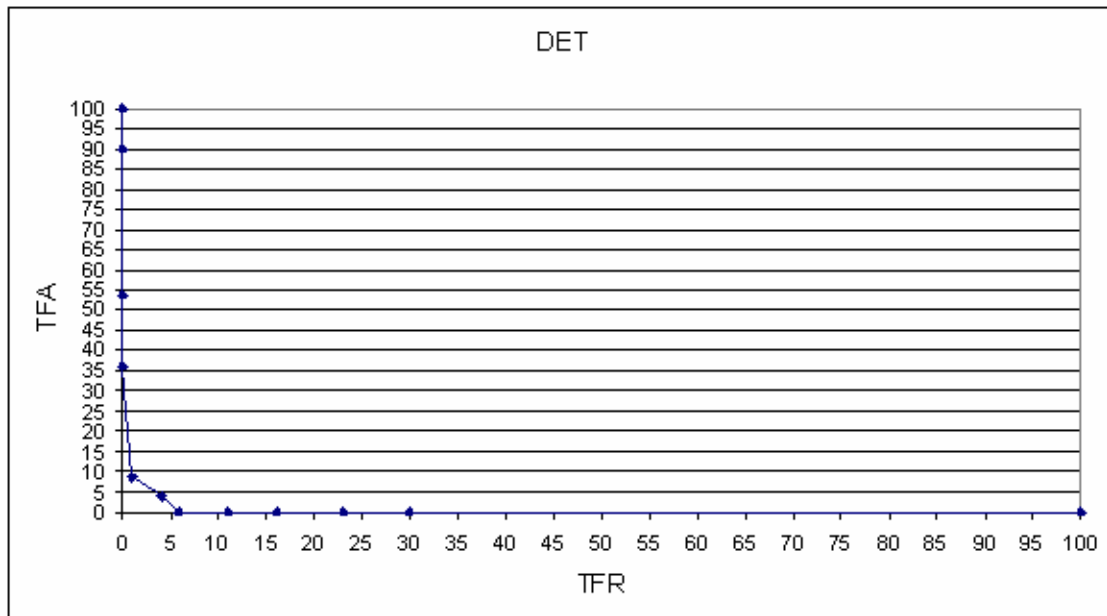


Figura A4.4 Curva DET

Acrónimos

MATLAB	MATrix LABoratory
LPC	Linear Prediction Coefficients
TRF	Transformada Rápida de Fourier
GUI	Graphic User Interface
DET	Data Error Trade-off

UNIVERSIDAD DE SALAMANCA

Escuela Politécnica Superior de Zamora

Ingeniería Técnica Informática de Gestión

Departamento de Informática y Automática, Área de Lenguajes y sistemas informáticos



Anexo V: Manuales de usuario

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

Índice Anexo 5: Manuales de usuario

Lista de tablas	193
Lista de figuras	194
A5.1 Documento de instalación y configuración	196
A5.2 Manual de usuario	199
A5.2.1 Configurar el sistema.	199
A5.2.2 Abrir un fichero de voz.	200
A5.2.3 Reproducir un intervalo	201
A5.2.4 Grabar un fichero de voz.	202
A5.2.5 Zoom.	203
A5.2.5.1 Acercar.	203
A5.2.5.2 Alejar.	204
A5.2.5.3 Intervalo.	205
A5.2.5.4 Vista inicial	206
A5.2.6 Ver estadísticas.	207
A5.2.7 Borrar estadísticas.	208
A5.2.8 Entrenamiento.	209
A5.2.9 Borrar locutor.	211
A5.2.10 Verificación.	212
A5.2.10.1 Por fichero.	212
A5.2.10.2 Por micrófono.	214
A5.2.11 Identificación.	216
A5.2.11.1 Por fichero.	216
A5.2.11.2 Por micrófono.	218
A5.2.12 Detección de inicio y fin de palabras.	220
A5.2.13 Etiquetado.	221
A5.2.14 Alternancia del espectrograma.	222
A5.2.15 Ayuda.	223
Acrónimos.	224

Lista de tablas

Lista de figuras

Figura A5.1 Menú Archivo -> Configurar.	199
Figura A5.2 Menú Archivo -> Abrir.	200
Figura A5.3 Cuadro de diálogo de abrir fichero	200
Figura A5.4 Menú Operaciones -> Reproducir.	201
Figura A5.5 Osciloscopio	201
Figura A5.6 Menú Operaciones -> Grabación	202
Figura A5.7 Cuadro de diálogo de grabar fichero.	202
Figura A5.8 Menú Operaciones -> Zoom -> Acercar.	203
Figura A5.9 Menú Operaciones -> Zoom -> Alejar.	204
Figura A5.10 Menú Operaciones -> Zoom -> Intervalo	205
Figura A5.11 Selección de un intervalo en el osciloscopio	205
Figura A5.12 Menú Operaciones -> Zoom -> Vista inicial.	206
Figura A5.13 Menú Estadísticas -> Ver.	207
Figura A5.14 Cuadro de diálogo de mostrar estadísticas	207
Figura A5.15 Menú Estadísticas -> Borrar.	208
Figura A5.16 Cuadro de diálogo de borrar estadísticas	208
Figura A5.17 Menú Operaciones -> Entrenamiento.	209
Figura A5.18 Cuadro de diálogo de introducción de nombre de locutor ..	209
Figura A5.19 Cuadros de diálogo de introducción de un fichero de voz del impostor (izquierda) y del locutor auténtico (derecha).	209
Figura A5.20 Menú Operaciones -> Eliminar locutor.	211
Figura A5.21 Cuadro de diálogo de introducción de nombre de locutor. .	211
Figura A5.22 Menú Operaciones -> Reconocimiento -> Verificación ...	212
Figura A5.23 Cuadro de diálogo de introducción de nombre de locutor. .	212
Figura A5.24 Cuadro de diálogo de selección de opción.	212
Figura A5.25 Cuadro de diálogo de selección de fichero de voz.	213
Figura A5.26 Menú Operaciones -> Reconocimiento -> Verificación ...	214
Figura A5.27 Cuadro de diálogo de introducción de nombre de locutor. .	214
Figura A5.28 Cuadro de diálogo de selección de opción.	214
Figura A5.29 Cuadro de diálogo de inicio de grabación	215
Figura A5.30 Menú Operaciones -> Reconocimiento -> Identificación ..	216
Figura A5.31 Cuadro de diálogo de selección de opción.	216
Figura A5.32 Cuadro de diálogo de selección de fichero de voz.	217
Figura A5.33 Menú Operaciones -> Reconocimiento -> Identificación ..	218
Figura A5.34 Cuadro de diálogo de selección de opción.	218.
Figura A5.35 Cuadro de diálogo de inicio de grabación	218
Figura A5.36 Detección de inicio y fin de palabras.	220
Figura A5.37 Selección de dos puntos de la señal de voz	221
Figura A5.38 Cuadro de diálogo de etiquetar.	221
Figura A5.39 Señal de voz etiquetada	221
Figura A5.40 Botones Espectro TRF y Espectro LPC.	222
Figura A5.41 (a)Espectrograma calculado con TRF (b)Espectrograma calculado con LPC	222
Figura A5.42 Menú Ayuda.	223
Figura A5.43 Cuadro de diálogo de Ayuda Acerca de...	223

A5.1 Documento de instalación y configuración

En este punto se especificará la manera de instalar el sistema en un ordenador, así como los requisitos de hardware y software que dicho ordenador debe cumplir y la configuración que hay que hacer del propio sistema.

La instalación del sistema en un ordenador, es realmente sencilla, ya que basta con copiar el directorio SERAL (contenido en el CD que se adjunta con el presente proyecto) en el directorio del ordenador donde se quiere instalar la aplicación.

Sin embargo el ordenador donde se instale la aplicación deberá cumplir una serie de requisitos para conseguir un correcto funcionamiento del sistema.

REQUISITOS HARDWARE

El equipo físico recomendado que necesita el sistema estará compuesto por la siguiente lista de elementos:

1. 512 MB de memoria RAM
2. 10 MB de disco duro
3. CPU de 3 GHz
4. Tarjeta de sonido con soporte para Entrada/Salida de 8 y 16 bits.
5. Micrófono unidireccional con sensibilidad a frecuencias entre 20 Hz y 20 KHz.

REQUISITOS SOFTWARE

El entorno lógico que necesita el sistema estará formado por la siguiente lista de elementos:

1. Sistema Operativo Microsoft Windows XP (versión 5.1.2600)
2. MatLab 7.0.0.19920 (R 14)
3. editor notepad (versión 5.1 o superior)
4. navegador web capaz de abrir ficheros con formato .chm:
 - Microsoft Internet Explorer (versión 6.0.2800 o superior)
 - Mozilla Firefox (versión 5.0 o superior)

Para lograr que el sistema funcione correctamente, a parte de cumplir con los requisitos anteriormente mencionados, hay que configurar el sistema de manera adecuada.

Para configurar el sistema hay que editar las propiedades almacenadas en el fichero seral.cfg (contenido en el directorio data del sistema). A continuación se indican las propiedades almacenadas en el fichero así como sus posibles valores y restricciones:

- **Propiedad FRE**

Esta propiedad determina la frecuencia en Hz con la que se muestrearán las señales de voz que se graben con el sistema.

- **Propiedad LONG**

Esta propiedad determina la longitud de las tramas en las que se divide una señal de voz en milisegundos.

- **Propiedad DESP**

Esta propiedad determina el desplazamiento de las tramas en milisegundos.

- **Propiedad CONP**

Esta propiedad determina el valor de la constante de preenfasis con la que se preenfatará la señal de voz

- **Propiedad OLPC**

Esta propiedad determina el orden de predicción lineal con el que se calcularán los coeficientes de predicción lineal.

- **Propiedad NTRF**

Esta propiedad determina el número de puntos con el que se calculara la transformada rápida de Fourier.

Esta propiedad presenta la restricción de que su valor tiene que ser una potencia de dos.

- **Propiedad NFVOZ**

Esta propiedad determina el número de muestras de voz del locutor son necesarias para crear un patrón de referencia.

- **Propiedad NLOC**

Esta propiedad determina el número de locutores registrados en el sistema.

- **Propiedad UGI**

Esta propiedad determina el umbral general de identificación en tanto por ciento.

Cuanto más bajo sea el valor de esta propiedad mayor será la tasa de Falsa No Identificación (FNI) y menor la tasa de Falsa Identificación (FI) y viceversa.

- **Propiedad UGV**

Esta propiedad determina el umbral general de verificación en tanto por ciento.

Cuanto más bajo sea el valor de esta propiedad mayor será la tasa de Falso Rechazo (FR) y menor la tasa de Falsa Aceptación (FA) y viceversa.

- **Propiedad NMCL**

Esta propiedad determina el número de muestras consecutivas que se toma de cada locutor en el entrenamiento.

- **Propiedad NEE**

Esta propiedad determina el número de etapas que se realizarán como máximo en el entrenamiento.

- **Propiedad EPE**

Esta propiedad determina la diferencia entre el valor esperado en el entrenamiento y el obtenido con la que se detiene del entrenamiento.

- **Propiedad TA**

Esta propiedad determina la tasa de aprendizaje del entrenamiento

- **Propiedad NNCO**

Esta propiedad determina el número de neuronas en la capa oculta de la RNA

A5.2 Manual de usuario

El contenido de este punto pretende ser un documento de ayuda que contenga toda la información necesaria para poder utilizar el sistema SERAL.

A5.2.1 Configurar el sistema.

La configuración del sistema SERAL esta guardada en el fichero seral.cfg que se encuentra situado en el directorio **data** dentro del árbol de directorios de la aplicación.

Al realizar el primer entrenamiento, se crea una copia del fichero seral.cfg llamada seral.ini y situada en el mismo directorio. Será con la configuración guardada en seral.ini con la que se hagan los entrenamientos, verificaciones e identificaciones.

La aplicación nos permite editar dicho fichero para poder modificar la configuración con la que trabaja el sistema, para poder editarlo, el sistema requiere que se tenga instalado en el ordenador el editor notepad.

Para poder editar el fichero seral.cfg, hay que seguir los siguientes pasos:

1. Seleccione la opción Configurar dentro del menú Archivo



Figura A5.1 Menú Archivo -> Configurar

2. Modifique los parámetros que desee del fichero de configuración.
3. Guarde la nueva configuración cuando termine de modificar los parámetros.

NOTA: El sistema esta entrenado para funcionar con la configuración por defecto, si modifica alguno de los parámetros, es posible que tenga que modificar más para que el sistema siga funcionando correctamente.

NOTA: La configuración con la que se entrena el sistema, tiene que ser la misma que con la que se realizan las pruebas.

NOTA: Al reconfigurar el sistema, conviene pulsar el botón recalcular, para regenerar las señales que estamos visualizando en pantalla con la nueva configuración.

A5.2.2 Abrir un fichero de voz.

SERAL permite abrir y reproducir un fichero de audio con extensión .wav y ver la señal de audio que contiene en la pantalla.

Para poder abrir un fichero de voz hay que seguir los siguientes pasos:

1. Seleccione la opción Abrir dentro del menú Archivo



Figura A5.2 Menú Archivo -> Abrir

2. Seleccione el fichero que quiere abrir, moviéndose por el árbol de directorios del sistema a través de la ventana:

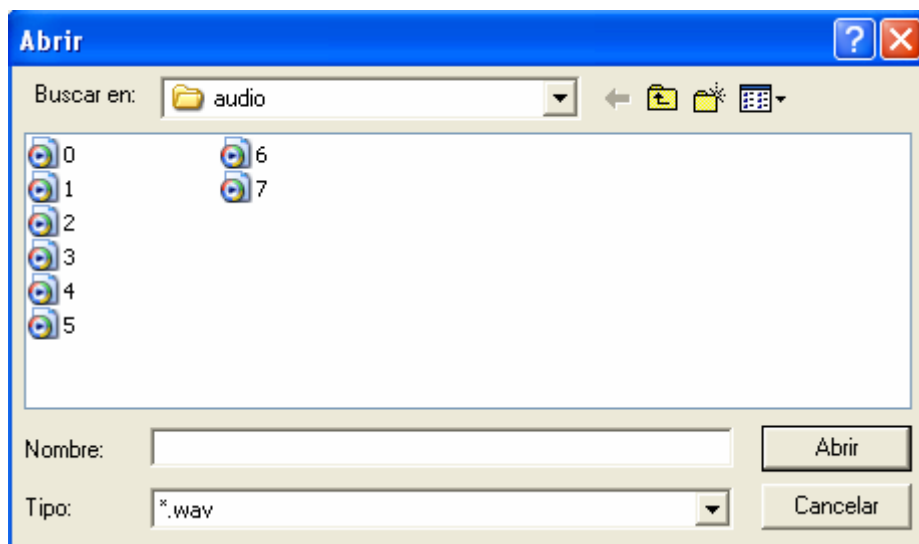


Figura A5.3 Cuadro de diálogo de abrir fichero

Para seleccionar un fichero, haga doble clic en su nombre o escriba este en el cuadro de texto y haga clic en Abrir, inicialmente se listan los ficheros de audio que se encuentran en el directorio **audio** dentro del árbol de directorios de la aplicación.

3. Si el fichero elegido tiene etiquetas, seleccione si quiere visualizarlas o no.

Esta misma operación se puede realizar a través del botón Cargar fichero situado en la parte superior de la aplicación en lugar de hacerlo a través de la opción Abrir del menú Archivo.

A5.2.3 Reproducir un intervalo.

SERAL permite reproducir un intervalo definido por el usuario de la señal de voz visualizada en el en la pantalla.

Para poder reproducir un intervalo de una señal de voz hay que seguir los siguientes pasos:

1. Seleccione la opción Reproducir dentro del menú Operaciones

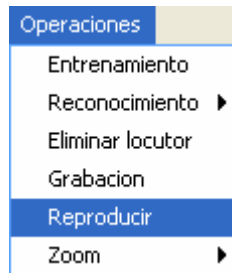


Figura A5.4 Menú Operaciones -> Reproducir

2. Seleccione con dos clic's la región de la señal de voz que quiere reproducir

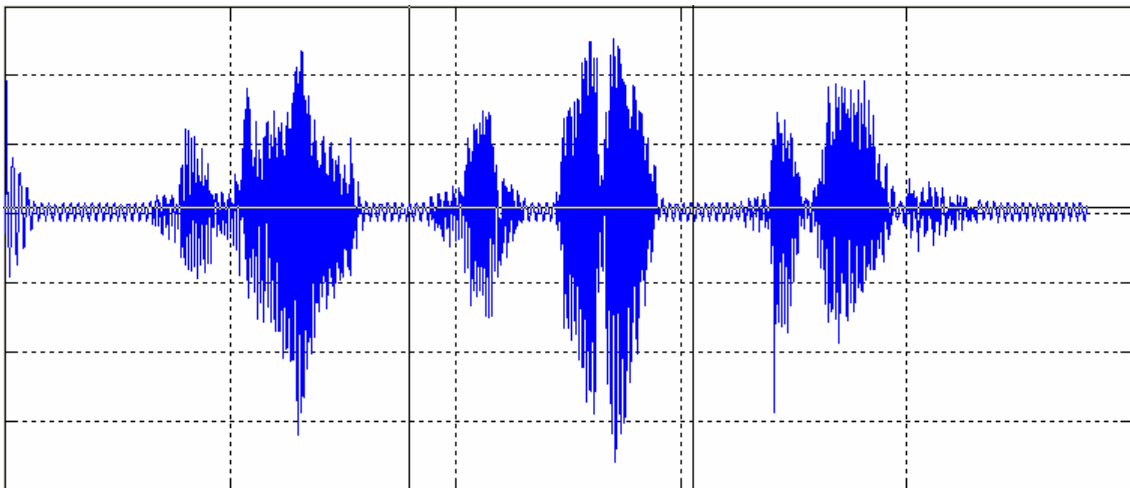


Figura A5.5 Osciloscopio

Esta misma operación se puede realizar a través del botón Reproducir situado en la parte superior de la aplicación en lugar de hacerlo a través de la opción Reproducir del menú Operaciones.

A5.2.4 Grabar un fichero de voz.

SERAL permite grabar un fichero de voz con una muestra de voz adquirida a través de un micrófono conectado al sistema en tiempo real.

Para poder realizar una grabación de voz hay que seguir los siguientes pasos:

1. Seleccione la opción Grabación dentro del menú Operaciones

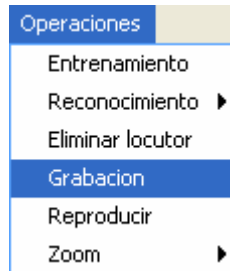


Figura A5.6 Menú Operaciones -> Grabación

2. Seleccione la ubicación y nombre del nuevo fichero a través de la ventana:

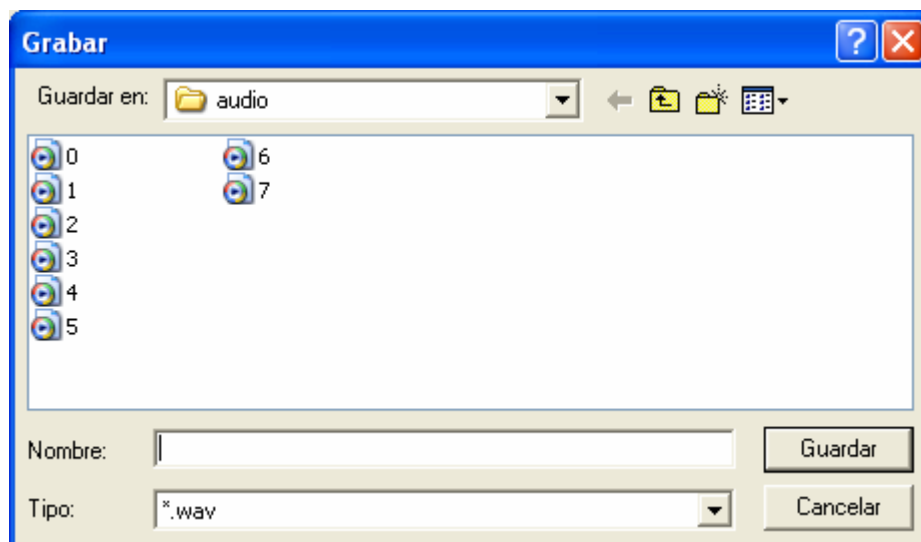


Figura A5.7 Cuadro de diálogo de grabar fichero

Si desea sobrescribir un fichero existente puede escribir su nombre en el cuadro de texto o seleccionarlo de la lista, si desea crear un nuevo fichero tiene que escribir su nombre en el cuadro de texto, inicialmente se encuentra en el directorio **audio** dentro del árbol de directorios de la aplicación.

3. Inmediatamente después de indicar el nombre del fichero se inicia la grabación de voz durante el tiempo establecido.

A5.2.5 Zoom

A5.2.5.1 Acercar

SERAL permite acercar una señal de voz visualizada en pantalla para poder verla con mayor nivel de detalle.

Para poder ver una zona de la señal con mayor nivel de detalle hay que seguir los siguientes pasos:

1. Seleccione la opción Acercar dentro del menú Operaciones $\Rightarrow$ Zoom

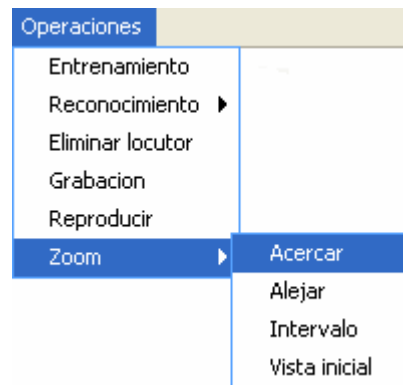


Figura A5.8 Menú Operaciones -> Zoom -> Acercar

2. Ahora en la pantalla se mostrará el 25% alrededor del punto central de la señal de voz.

A5.2.5.2 Alejar

SERAL también permite alejar una señal de voz visualizada en pantalla para poder ver la región que está a su alrededor.

Para poder alejar la señal hay que seguir los siguientes pasos:

1. Seleccione la opción Alejar dentro del menú Operaciones $\Rightarrow$ Zoom

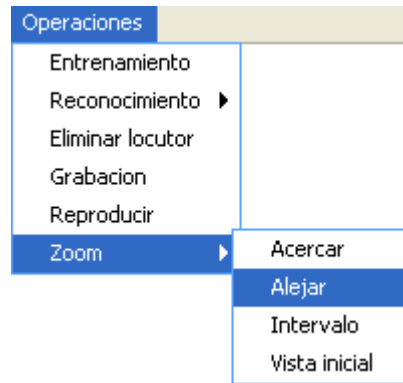


Figura A5.9 Menú Operaciones -> Zoom -> Alejar

2. Ahora en la pantalla se mostrará el 200% alrededor del punto central de la señal de voz.

A5.2.5.3 Intervalo

SERAL permite visualizar una zona concreta de una señal de voz seleccionada por el usuario.

Para poder realizar esta operación hay que seguir los siguientes pasos:

1. Seleccione la opción Intervalo dentro del menú Operaciones $\Rightarrow$ Zoom

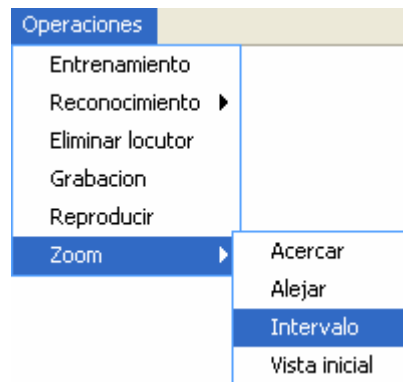


Figura A5.10 Menú Operaciones -> Zoom -> Intervalo

2. Seleccione con dos clic's la región de la señal de voz que quiere visualizar

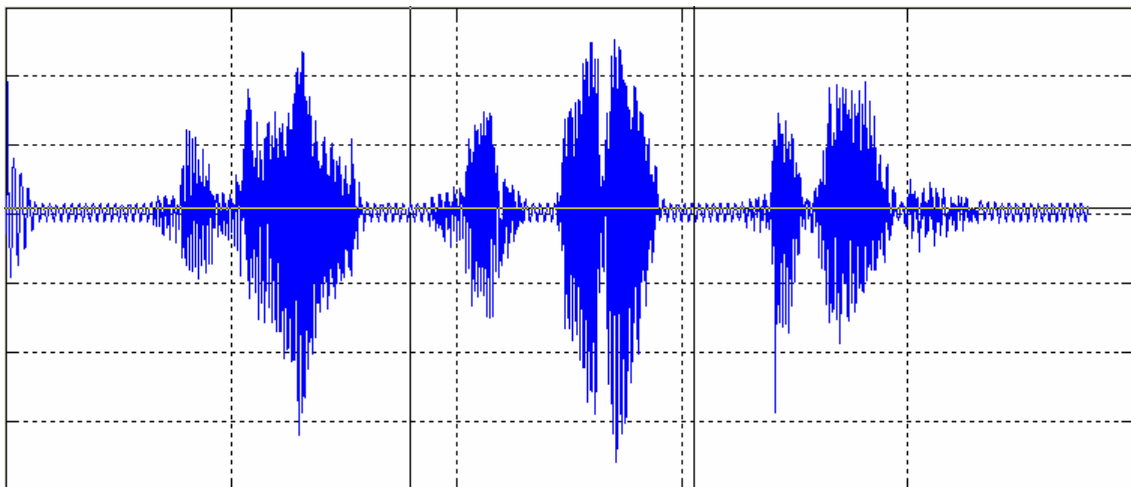


Figura A5.11 Selección de un intervalo en el osciloscopio

Esta misma operación se puede realizar a través del botón Zoom intervalo situado en la parte superior de la aplicación en lugar de hacerlo a través de la opción Intervalo del menú Operaciones $\Rightarrow$ Zoom.

A5.2.5.4 Vista inicial

SERAL permite visualizar la señal de voz completa cuando se está visualizando una región de la misma.

Para realizar esta operación hay que seguir los siguientes pasos:

1. Seleccione la opción Vista inicial dentro del menú Operaciones $\Rightarrow$ Zoom

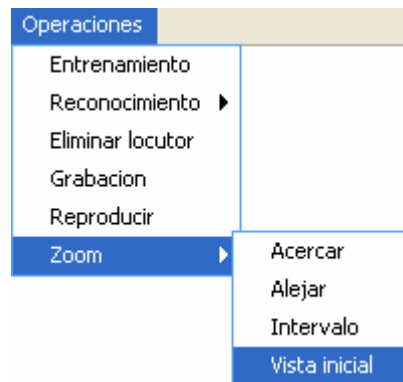


Figura A5.12 Menú Operaciones -> Zoom -> Vista inicial

2. Ahora en la pantalla se visualizará la señal de voz completa.

Esta misma operación se puede realizar a través del botón Vista inicial situado en la parte superior de la aplicación en lugar de hacerlo a través de la opción Vista inicial del menú Operaciones $\Rightarrow$ Zoom

A5.2.6 Ver estadísticas

SERAL guarda un fichero con estadísticas donde se almacena el número de veces que un locutor ha sido aceptado o rechazado y el número de veces que un locutor ha sido identificado o no identificado.

Este fichero se llama estadistica.est y está guardado en el directorio **data** dentro del árbol de directorios de la aplicación.

El sistema SERAL permite visualizar el contenido del fichero de estadísticas.

Para poder realizar esta operación hay que seguir los siguientes pasos:

1. Seleccione la opción Ver dentro del menú Estadísticas

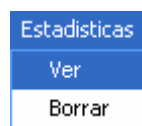


Figura A5.13 Menú Estadísticas -> Ver

Aparecerá en pantalla un cuadro de diálogo con la información estadística.

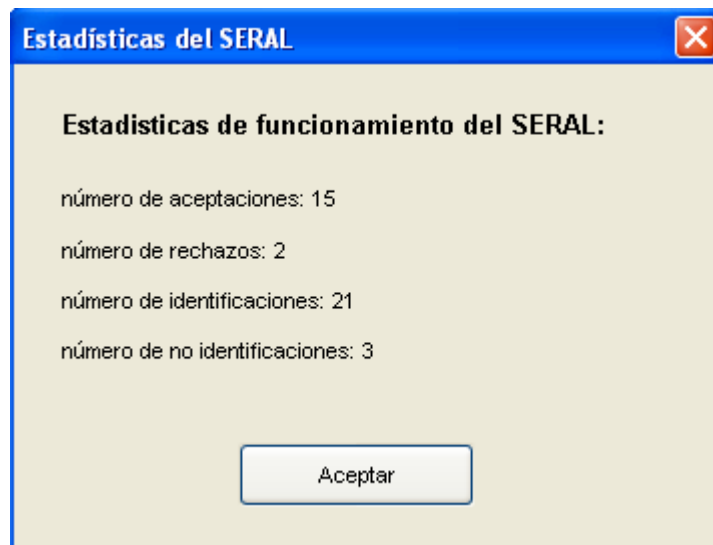


Figura A5.14 Cuadro de diálogo de mostrar estadísticas

A5.2.7 Borrar estadísticas

El sistema SERAL permite borrar las estadísticas del fichero estadistica.ini para que estas vuelvan a empezar desde cero.

Para borrar las estadísticas de SERAL hay que seguir los siguientes pasos:

1. Seleccione la opción Borrar dentro del menú Estadísticas

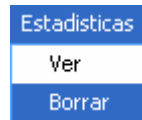


Figura A5.15 Menú Estadísticas -> Borrar

2. Aparecerá un cuadro de diálogo como el siguiente para confirmar la operación

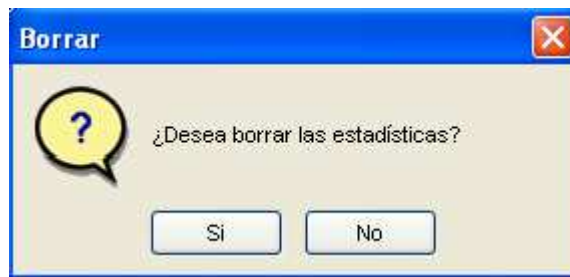


Figura A5.16 Cuadro de diálogo de borrar estadísticas

3. Seleccionando la opción Si las estadísticas se borrarán.

A5.2.8 Entrenamiento

Con esta opción el sistema SERAL, permite dar de alta nuevos locutores en el sistema y entrenar sus correspondientes Redes Neuronales Artificiales.

Para poder dar de alta un locutor en SERAL, hay que seguir los siguientes pasos:

1. Seleccione la opción Entrenamiento del menú Operaciones

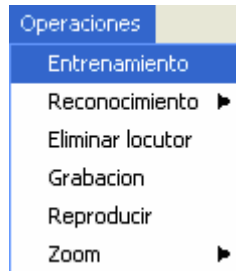


Figura A5.17 Menú Operaciones -> Entrenamiento

2. Introduzca el nombre del nuevo locutor en el siguiente cuadro de diálogo.

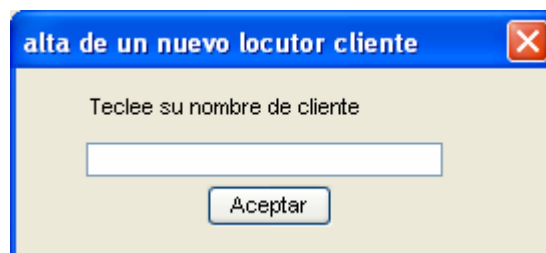


Figura A5.18 Cuadro de diálogo de introducción de nombre de locutor

3. Introduzca alternativamente, muestras grabadas del locutor y de impostores, con la ayuda de los siguientes cuadros de diálogo. El título del cuadro de diálogo, le indicará si tiene que introducir una muestra del locutor o del impostor.



Figura A5.19 Cuadros de diálogo de introducción de un fichero de voz del impostor (izquierda) y del locutor auténtico (derecha).

Se requerirán tantas muestras de voz como estén indicadas en el fichero de configuración.

NOTA: Si el nombre del locutor introducido ya existe en el sistema, no se creará ningún nuevo locutor.

A5.2.9 Borrar locutor

SERAL permite borrar locutores del sistema, borrando su Red Neuronal Artificial del fichero de patrones de locutores.

Para poder dar de baja un locutor hay que realizar las siguientes operaciones:

1. Seleccione la opción Eliminar locutor del menú Operaciones

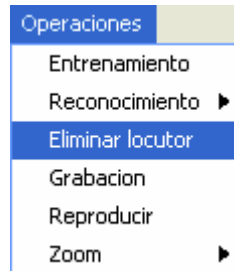


Figura A5.20 Menú Operaciones -> Eliminar locutor

2. Introduzca el nombre del locutor a eliminar en el siguiente cuadro de diálogo.

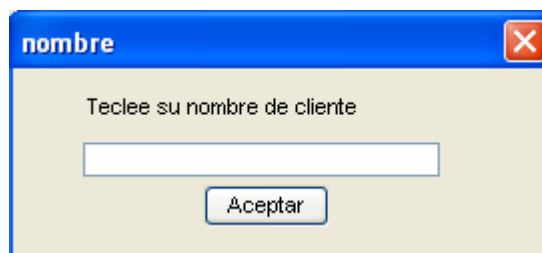


Figura A5.21 Cuadro de diálogo de introducción de nombre de locutor

NOTA: El nombre introducido, tiene que ser el de uno de los locutores existentes en el sistema, sino, no se realizará la eliminación de ningún locutor.

A5.2.10 Verificación

A5.2.10.1 Por fichero

La configuración con la que se entrena el sistema, tiene que ser la misma que con la que se realizan las pruebas, y se carga desde seral.ini (fichero que guarda la configuración del entrenamiento).

SERAL permite realizar la operación de verificación de locutor mediante un fichero de voz o mediante una señal de voz adquirida desde el micrófono en tiempo real.

Para realizar la verificación de locutor con un fichero de voz hay que seguir los siguientes pasos:

1. Seleccione la opción Verificación dentro del menú Operaciones ⇒ Reconocimiento.



Figura A5.22 Menú Operaciones -> Reconocimiento -> Verificación

2. Introduzca el nombre del locutor que quiere verificar en el siguiente cuadro de diálogo:

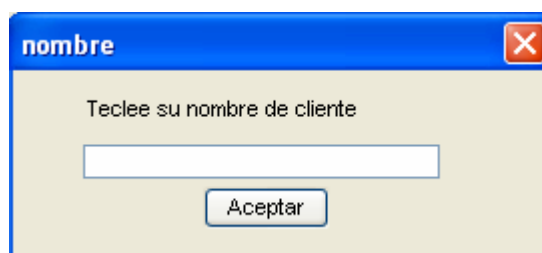


Figura A5.23 Cuadro de diálogo de introducción de nombre de locutor

3. Aparecerá un cuadro de diálogo como el siguiente para preguntar el modo en el que quiere realizar la operación:

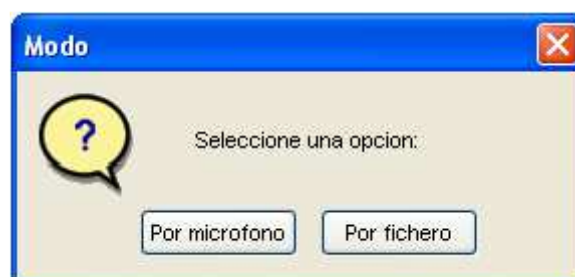


Figura A5.24 Cuadro de diálogo de selección de opción

En este cuadro de diálogo hay que seleccionar la opción Por fichero.

4. Seleccione el fichero con la muestra de voz que quiere verificar, con la ayuda del siguiente cuadro de diálogo:

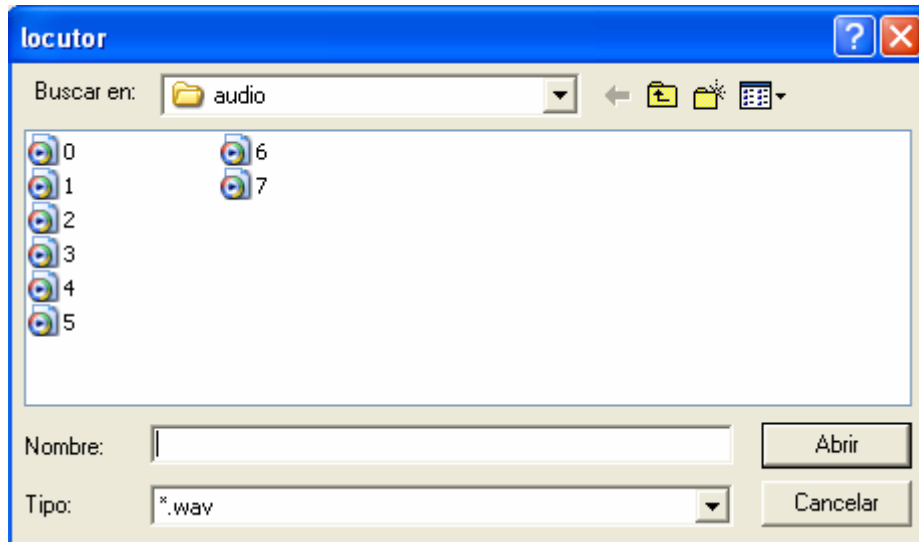


Figura A5.25 Cuadro de diálogo de selección de fichero de voz

5. En pantalla aparecerá un cuadro de diálogo con el resultado de la operación.

NOTA: Para poder llevar a cabo esta operación, es necesario que el nombre de locutor introducido exista en el sistema.

A5.2.10.2 Por micrófono

La configuración con la que se entrena el sistema, tiene que ser la misma que con la que se realizan las pruebas, y se carga desde seral.ini (fichero que guarda la configuración del entrenamiento).

SERAL permite realizar la operación de verificación de locutor mediante un fichero de voz o mediante una señal de voz adquirida desde el micrófono en tiempo real.

Para realizar la verificación de locutor con una señal de voz adquirida en tiempo real hay que seguir los siguientes pasos:

1. Seleccione la opción Verificación dentro del menú Operaciones ⇒ Reconocimiento.



Figura A5.26 Menú Operaciones -> Reconocimiento -> Verificación

2. Introduzca el nombre del locutor que quiere verificar en el siguiente cuadro de diálogo:

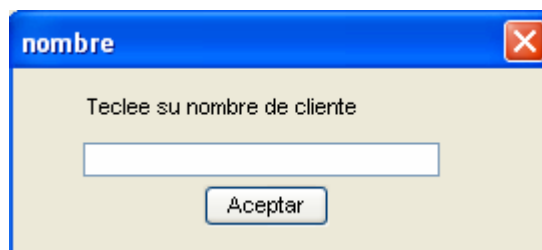


Figura A5.27 Cuadro de diálogo de introducción de nombre de locutor

3. Aparecerá un cuadro de diálogo como el siguiente para preguntar el modo en el que quiere realizar la operación:



Figura A5.28 Cuadro de diálogo de selección de opción

En este cuadro de diálogo hay que seleccionar la opción Por microfono.

4. Una vez introducido el nombre del locutor, aparecerá un cuadro de diálogo como el siguiente, y en cuanto pulse Aceptar, se grabará por micrófono una muestra de voz de tres segundos.

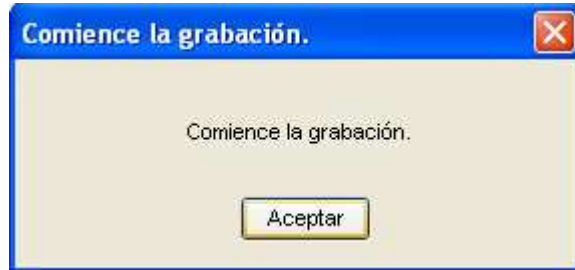


Figura A5.29 Cuadro de diálogo de inicio de grabación

5. En pantalla aparecerá un cuadro de diálogo con el resultado de la operación.

NOTA: Para poder llevar a cabo esta operación, es necesario que el nombre de locutor introducido exista en el sistema.

A5.2.11 Identificación

A5.2.11.1 Por fichero

La configuración con la que se entrena el sistema, tiene que ser la misma que con la que se realizan las pruebas, y se carga desde seral.ini (fichero que guarda la configuración del entrenamiento).

SERAL permite realizar la operación de identificación de locutor mediante un fichero de voz o mediante una señal de voz adquirida desde el micrófono en tiempo real.

Para realizar la identificación de locutor con un fichero de voz hay que seguir los siguientes pasos:

1. Seleccione la opción Identificación dentro del menú Operaciones ⇒ Reconocimiento.



Figura A5.30 Menú Operaciones -> Reconocimiento -> Identificación

2. Aparecerá un cuadro de diálogo como el siguiente para preguntar el modo en el que quiere realizar la operación:



Figura A5.31 Cuadro de diálogo de selección de opción

En este cuadro de diálogo hay que seleccionar la opción Por fichero.

3. Seleccione el fichero con la muestra de voz que quiere identificar, con la ayuda del siguiente cuadro de diálogo:

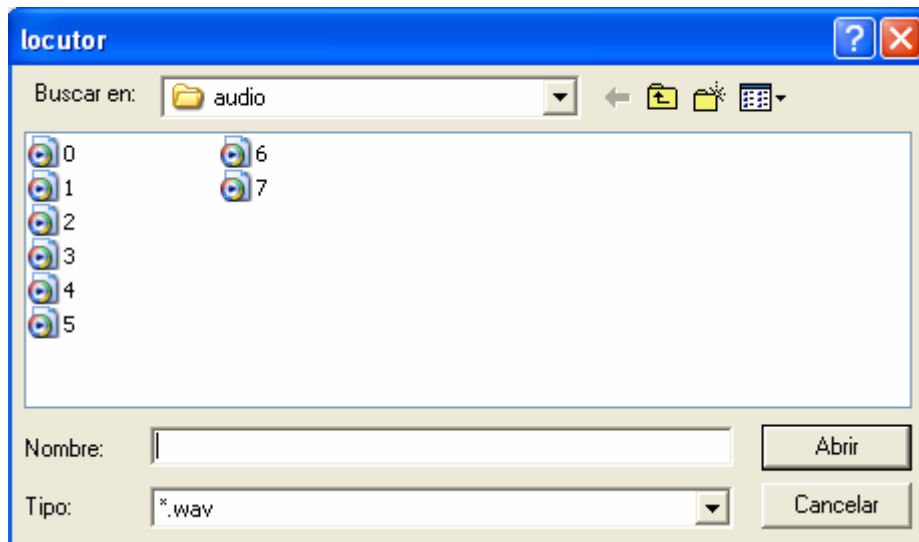


Figura A5.32 Cuadro de diálogo de selección de fichero de voz

4. En pantalla aparecerá un cuadro de diálogo con el resultado de la operación.

NOTA: Para poder llevar a cabo esta operación, es necesario que existan locutores en el sistema.

A5.2.11.2 Por micrófono

La configuración con la que se entrena el sistema, tiene que ser la misma que con la que se realizan las pruebas, y se carga desde seral.ini (fichero que guarda la configuración del entrenamiento).

SERAL permite realizar la operación de identificación de locutor mediante un fichero de voz o mediante una señal de voz adquirida desde el micrófono en tiempo real.

Para realizar la identificación de locutor con una señal de voz adquirida en tiempo real hay que seguir los siguientes pasos:

1. Seleccione la opción Identificación dentro del menú Operaciones ⇒ Reconocimiento.



Figura A5.33 Menú Operaciones -> Reconocimiento -> Identificación

2. Aparecerá un cuadro de diálogo como el siguiente para preguntar el modo en el que quiere realizar la operación:



Figura A5.34 Cuadro de diálogo de selección de opción

En este cuadro de diálogo hay que seleccionar la opción Por microfono.

3. Una vez indicado que la identificación se realizará por micrófono, aparecerá un cuadro de diálogo como el siguiente, y en cuanto pulse Aceptar, se grabará por micrófono durante tres segundos la muestra de voz a identificar.

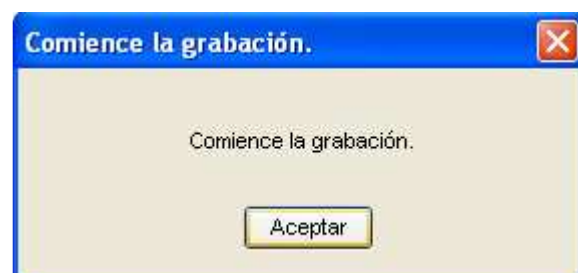


Figura A5.35 Cuadro de diálogo de inicio de grabación

4. En pantalla aparecerá un cuadro de diálogo con el resultado de la operación.

NOTA: Para poder llevar a cabo esta operación, es necesario que existan locutores en el sistema.

A5.2.12 Detección de inicio y fin de palabras

El sistema SERAL permite detectar el inicio y fin de las palabras que componen una señal de voz.

Para realizar esta detección hay que realizar los siguientes pasos:

1. Pulsar el botón “bordes” situado en la parte superior de la aplicación.
2. El sistema calcula automáticamente el inicio y fin de las palabras y las delimita con barras en el gráfico presentado en el osciloscopio, como muestra la figura.

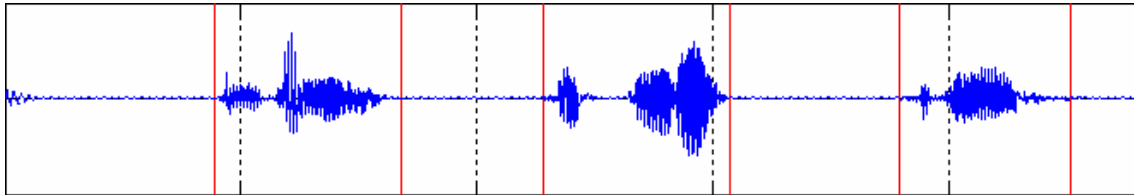


Figura A5.36 Detección de inicio y fin de palabras

A5.2.13 Etiquetado

El sistema SERAL permite asociar a ciertas zonas de la señal de voz unas etiquetas definidas por el usuario

Para realizar el etiquetado de una señal de voz hay que realizar los siguientes pasos:

1. Pulsar el botón “etiquetar” situado en la parte superior de la aplicación.
2. Seleccionar con el ratón el inicio y fin de la etiqueta, como se muestra en la siguiente figura.

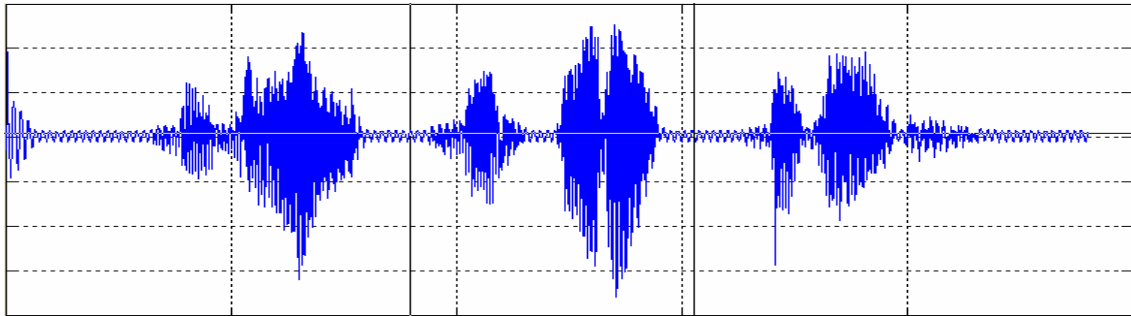


Figura A5.37 Selección de dos puntos de la señal de voz

3. Introducir la etiqueta asociada a la zona marcada mediante el siguiente cuadro de diálogo.

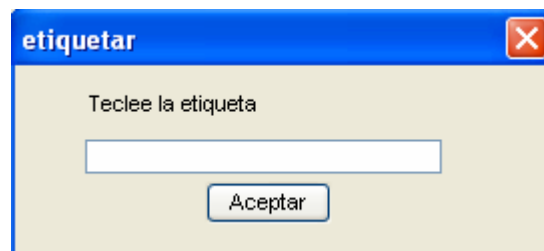


Figura A5.38 Cuadro de diálogo de etiquetar

4. Las etiquetas se mostrarán en la pantalla.

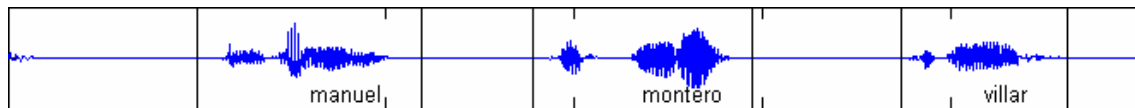


Figura A5.39 Señal de voz etiquetada

NOTA: Si actualmente no está viendo las etiquetas, puede elegir si quiere añadir la nueva etiqueta al fichero de etiquetas existente o si quiere sobrescribirlo con la nueva etiqueta.

A5.2.14 Alternancia del espectrograma

El sistema SERAL, permite ver el espectrograma de la señal de voz, calculado con la TRF o con los coeficientes LPC.

Para alternar los espectrogramas, hay que pulsar los botones Espectro TRF (que muestra el calculado con la TRF) y Espectro LPC (que muestra el calculado con los coeficientes LPC).



Figura A5.40 Botones Espectro TRF y Espectro LPC

La pulsación de cada uno de estos botones, hace que se muestre en pantalla los siguientes espectrogramas:

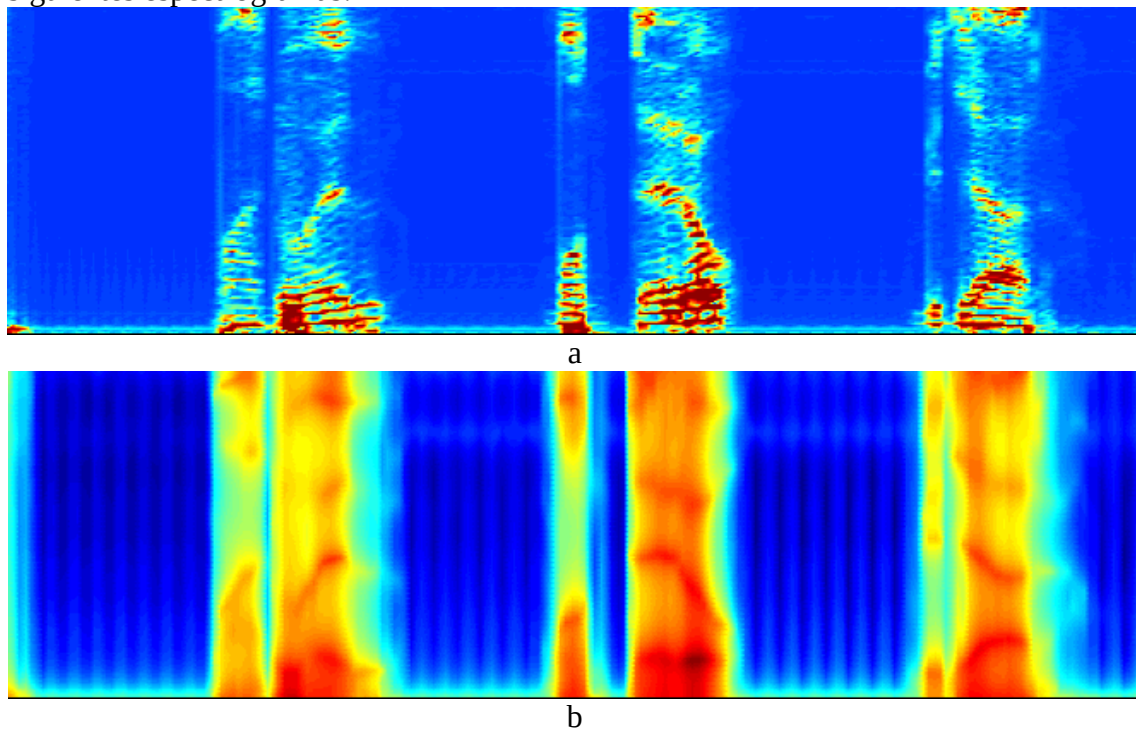


Figura A5.41 (a)Espectrograma calculado con TRF
(b)Espectrograma calculado con LPC

A5.2.15 Ayuda

El sistema SERAL dispone de una ayuda para facilitar su utilización a los distintos usuarios.

La ayuda se presenta a través del menú ayuda.

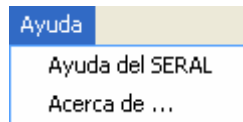


Figura A5.42 Menú Ayuda

Este menú consta de dos opciones, si se selecciona la primera (ayuda del SERAL), se desplegará este documento que contiene la ayuda necesaria para poder realizar todas las operaciones que permite el sistema SERAL.

Si se selecciona la segunda opción (acerca de ...), se mostrara un cuadro de diálogo con información acerca del sistema SERAL como el que aparece a continuación:



Figura A5.43 Cuadro de diálogo de Ayuda Acerca de...

Acrónimos

RAM	Random Access Memory
CPU	Central Unit Processor
FR	Falsa Rechazo
FA	Falsa Aceptación
FI	Falsa Identificación
FNI	Falsa No Identificación
RNA	Redes Neuronales Artificiales

Acrónimos

SERAL	Sistema Experimental de Reconocimiento Automático de Locutor
RAV	Reconocimiento Automático de la Voz
RAH	Reconocimiento Automático el Habla
RAL	Reconocimiento Automático de Locutor
IBP	Identificación Biométrica Personal
VAL	Verificación Automática de Locutor
IAL	Identificación Automática de Locutor
FR	Falsa Rechazo
FA	Falsa Aceptación
FI	Falsa Identificación
FNI	Falsa No Identificación
TEE	Tasa de EquiError
RAL-DT	RAL-Dependientes del Texto
RAL-IT	RAL-Independientes del Texto
CPL	Coeficientes de Predicción Lineal
EBF	Energía por Bandas de Frecuencia
RNA	Redes Neuronales Artificiales
MATLAB	MATrix LABoratory
TRF	Transformada Rápida de Fourier
TCC	Tasa de cruces por cero
ADT	Análisis Dinámico Temporal
MOM	Modelos Ocultos de Markov
PERT	Program Evaluation and Review Technique
VAN	Valor Actual Neto
TIR	Tasa Interna de Rentabilidad
RF	Requisito funcional
RNF	Requisito no funcional
RI	Requisito de información
ASCII	American Standard Code for Information Interchange
DFD	Diagramas de Flujo de Datos
MB	MegaByte
RAM	Random Access Memory
CPU	Central Unit Processor
LPC	Linear Prediction Coefficients
GUI	Graphic User Interface
DET	Data Error Trade-off

Glosario

Reconocimiento Automático del locutor: método por el cual una máquina (generalmente un ordenador) puede saber quién es el locutor que está pronunciando unas palabras.

Verificación Automática del locutor: método que consiste en decidir si una muestra de voz desconocida pertenece o no a un locutor que proporciona su identidad al sistema.

Identificación Automática del locutor: método que consiste en decidir a que locutor de los reconocidos por el sistema pertenece una muestra de voz desconocida.

Falsa Rechazo: En la VAL, cuando un usuario auténtico es rechazado por el sistema.

Falsa Aceptación: En la VAL, cuando un usuario falso (impostor) es aceptado por el sistema.

Falsa Identificación: En la IAL, cuando un usuario o impostor es identificado erróneamente, es decir, se le asigna al locutor una identidad que no le corresponde.

Falsa No Identificación: En la IAL, cuando un usuario del sistema no es identificado por el mismo.

Tasa de EquiError: En la VAL, umbral para el cual la tasa de Falsos Rechazos es igual a la de Falsas Aceptaciones. En la IAL, umbral para el cual la tasa de Falsas Identificaciones es igual a la de Falsas No Identificaciones.

Bibliografía

[Doddington, 1985] Doddington G.R., "Speaker Recognition – Identifying people by their voices". Proceedings of IEEE, Vol 73, No. 11 1985 pp.1651 – 1664

[Soong, Rosenberg y Juang, 1987] Soong F. K., Rosenberg A. E. y Juang B. H., "A vector Quantization Approach to Speaker Recognition". AT&T technical journal, Vol. 66 Issue 2 1987 pp. 14 – 26

[O'Shaughnessy, 1987] O'Shaughnessy D., "Speech communication: human and machine". Adison-Wesley Publishing Company USA 1987 p.568

[Atal y Hanauer, 1971] Atal B. S. and Hanauer S. L. "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave", Journal of the Acoustical Society of America. Vol 50 No 2 1971 pp.637 – 655

[Rabiner y Schafer, 1978] Rabiner L. R. and Schafer R. W., "Digital Signal Processing of Speech Signals". Prentice Hall Inc, Englewood Cliffs, New Jersey, USA 1978 p. 512

[Kröse y Smagt, 1993] Ben J. A. Kröse and P. Patrick van der Smagt "An Introduction to Neural Networks" Faculty of Mathematics and Computer Science, University of Amsterdam, The Netherlands 1993. p.113

[Rabiner y Juang, 1993] Rabiner L. R. and Juang B. H., "Fundamentals of Speech Recognition" Prentice Hall Signal Processing series, Englewood Cliffs, New Jersey 1993 p.507

[Lamel, Rabiner, Rosenberg y Wilpon, 1991] Lamel L. F., Rabiner L. R., Rosenberg A. E. and Wilpon J. G., "An Improved Endpoint Detector for Isolated Word Recognition". IEEE Transactions on ASSP, Vol 29 No 4 1981 pp.777 – 785

[Dayhoff, 1990] J.E. Dayhoff, "Neural Network Architectures", Van Nostrand Reinhold, New York 1990

[Tapiador y Sigüenza, 2005] Tapiador Mateos M. y Sigüenza Pizarro J.A. "Tecnologías biométricas aplicadas a la seguridad" Ra-Ma Madrid 2005 pp. 167 – 192

[Vivaracho, Cardeñoso y González, 2006] Vivaracho Pascual C.E., Cardeñoso Payo V, y González Ferreras C., "Mejoras en un Sistema de Reconocimiento de Locutor Basado en RNA Mediante entrenamiento con Normalización de Canal" Dpto. Informática, Universidad de Valladolid, España, 2006

[Furi,1993] Furi S., "Cepstral Analysis Techniques in Speech Recognition". Proceedings of the IEEE, Vol 81, No 9, pp1214-1246, September 1993.

OTRAS FUENTES CONSULTADAS

- 📖 Junca J.C., Mak B., and Reaves B. “A Robust Algorithm for Word Boundary Detection in the Presence of Noise”, IEEE transactions on Speech and Audio Processing Vol. 2 No 3 1994 pp. 406 – 412
- 📖 Qi Y. and Hunt B.R., “Voiced-Unvoiced-Silence Classifications of Speech Using Hybrid Features and a Network Classifier”, IEEE transactions on Speech and Audio Processing Vol. 1 No 2 1993 pp. 250 – 255
- 📖 Zelinski R. and Class F. “A Segmentation Algorithm for Connected Word Recognition Based on Estimation Principles”, IEEE Transactions on ASSP, Vol. 31 No. 4 1983 pp. 818 – 827
- 📖 Silva Varela H. “Esquemas Experimentales para el Reconocimiento Automático del Locutor y la Decodificación Acústico-Fonética” Dpto. Informática, Universidad de Valladolid, España, 2000
- 📖 Martin A., Doddington G., Kamm T., Ordowski M., Prybocki M., “The DET Curve in Assessment of Detection Task Performance” National Institute of Standards and Technology, USA.
- 📖 Pressman R.S., "Ingeniería del Software, Un enfoque práctico" Mc Graw Hill Mexico 2005 pp. 58 – 60, 211 – 218
- 📖 Piattini M.G., "Análisis y Diseño de Aplicaciones Informáticas de Gestión, Una perspectiva de Ingeniería del Software" Ra-Ma Madrid 2003 pp.70 – 74, 111 – 132, 175 – 182
- 📖 F. J. García Peñalvo, J. M. Maudes Raedo, M. G. Piattini Velthius, J. R. García-Bermejo Giner y M. N. Moreno García “Proyecto de final de carrera en la Ingeniería Técnica en Informática, Guía de realización y documentación”, Departamento de Informática y Automática Universidad de Salamanca, España, 2000
- 📖 Martínez García-Moreno, C.A. and Silva Varela, H. “An Experimental System for Automatic Speaker Recognition”, Proceedings of the ICSPAT’93 Santa Clara, California USA, Vol2, pp.1499-1506, 1993

Alguno de los artículos mencionados en este apartado, están presentes en las siguientes direcciones Web:

- 🌐 <http://www.shsconsultores.es/jornadas/>
- 🌐 <http://www.infor.uva.es/~cesargf/investigacion/publicaciones.html>

Índice alfabético

A:

Análisis, 32, 33, 34, 35, 36, 40, 49, 68, 120

B:

Biométrica, 14

C:

Coeficientes de Predicción Lineal, 23, 24, 34, 36, 40, 162, 197

D:

Diagramas de Flujo de Datos, 101, 107

F:

Falsa Aceptación, 15, 197

Falsa Identificación, 16, 197

Falsa No Identificación, 16, 197

Falsa Rechazo, 15, 197

I:

Identificación, 14, 15, 16, 20, 33, 67, 69, 71, 72, 73, 78, 82, 84, 95, 122, 128, 129, 143, 144, 181, 197, 198, 216, 218

Impostor 15, 16, 209

L:

Locutor, 11, 13, 14, 15, 16, 18, 20, 21, 22, 33, 35, 36, 38, 39, 40, 67, 69, 71, 72, 73, 78, 82, 84, 95, 97, 98, 101, 102, 106, 109, 110, 112, 120, 122, 129, 130, 143, 144, 145, 149, 174, 175, 176, 181, 182, 185, 197, 198, 207, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219

M:

MATLAB, 33, 34, 35, 36, 40, 53, 67, 148, 160, 161, 165, 184, 196

P:

Patrón, 14, 16, 17, 29, 35, 36, 40, 70, 71, 78, 82, 104, 105, 106, 108, 109, 110, 111, 112, 197, 198

R:

Reconocimiento, 11, 13, 14, 18, 27, 33, 36, 39, 40, 67, 82, 95, 101, 106, 111, 120, 122, 128, 129, 181, 182, 212, 214, 216, 218

Redes Neuronales Artificiales, 27, 33, 34, 35, 36, 39, 40, 69, 97, 122, 163, 164, 174, 175, 176, 185, 209, 211

Ruido, 23, 35

S:

SERAL, 24, 40, 83, 101, 122, 127, 135, 143, 180, 184, 196, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 211, 212, 214, 216, 218, 220, 221, 222, 223

Seguridad, 20, 53

T:

Tasa de EquiError, 16, 17

Transformada Rápida de Fourier, 24, 34, 36, 40, 162, 185, 197

U:

Umbral, 16, 17, 28, 197

V:

Verificación, 15, 69, 71, 72, 73, 82, 84, 95, 122, 128, 129, 143, 144, 181, 197, 198, 212, 214

Voz, 11, 13, 14, 15, 16, 17, 20, 21, 22, 23, 25, 27, 34, 39, 40, 67, 69, 70, 71, 73, 74, 78, 79, 80, 81, 82, 84, 95, 97, 98, 99, 100, 101, 102, 104, 105, 106, 107, 108, 108, 110, 111, 112, 113, 114, 115, 116, 121, 122, 135, 137, 138, 139, 140, 141, 143, 144, 145, 149, 150, 162, 165, 166, 167, 168, 169, 170, 171, 172, 173, 177, 178, 179, 180, 183, 185, 186, 196, 197, 200, 201, 202, 203, 204, 205, 206, 210, 212, 213, 214, 215, 216, 218, 220, 221, 222