

UNIVERSIDAD CARLOS III DE MADRID

Escuela Politécnica Superior



Ingeniería en Informática

PROYECTO FIN DE CARRERA

Desarrollo de una plataforma para la grabación y análisis de somniloquias: una aproximación usando técnicas de análisis de voz

Autor: Virginia Aparicio Paniagua

Tutor: Israel González Carrasco

Co-director: Alejandro Rodríguez González

Leganés, Octubre de 2012





Título: Desarrollo de una plataforma para la grabación y análisis de somniloquias: una aproximación usando técnicas de análisis de voz.

Autor: Virginia Aparicio Paniagua

Tutor: Israel González Carrasco

Co-director: Alejandro Rodríguez González

TRIBUNAL

Presidente:

Vocal:

Secretario:

Realizado el acto de defensa y lectura del Proyecto Fin de Carrera el día ___ de octubre de 2012 en Leganés, en la Escuela Politécnica Superior de la Universidad Carlos III de Madrid, acuerda otorgarle la CALIFICACIÓN de

VOCAL

SECRETARIO

PRESIDENTE



Agradecimientos

En primer lugar, me gustaría dar las gracias a mi familia por apoyarme siempre y por creer en mí de una manera incondicional, sin vosotros nada de esto sería posible.

A mi compañero y amigo Jorge, por todos los buenos momentos que hemos compartido y los que están aún por llegar.

A mi grupo de amigos de Getafe: Sergio, Raquel, Pacheco, Rosi, Pilar y Paola, por sacarme siempre una sonrisa.

A mis amigos de Telecomunicaciones: Chemi, Pablete, Miguel, Elena, Aída, Adal, Rober, Lisar, Jesús, Carlitos, Samu, Dieguete y en especial a David por toda la ayuda ofrecida.

A mis amigos de universidad, por todos estos años juntos.

A Sito, por sus ánimos y por su cariño.

A Raquel, porque siempre has estado cuando te necesitaba.

Por último, a toda la gente que ha colaborado en este proyecto aportando su voz.

Muchas gracias.



Autores

Este proyecto de fin de carrera ha sido realizado por los alumnos Virginia Aparicio Paniagua y Jorge Pérez Muñoz de la Universidad Carlos III de Madrid.

En la página siguiente se muestra una tabla con la repartición del trabajo entre los miembros del equipo, así como la dedicación que se ha tenido para cada tarea. La medición de las horas hace referencia al desarrollo del presente documento.



Tarea		Virginia Aparicio	Jorge Pérez	Total
Introducción y Objetivos	Introducción	1	0	1
	Motivación	1	0	1
	Objetivos	1	0	1
	Posibles aplicaciones	1	0	1
	Medios empleados	1	0	1
	Esquema de la memoria	1	0	1
Estado del arte	Estado del arte en análisis de emociones	7	0	7
	Estado del arte en reconocimiento automático del habla	6	0	6
	Estado del arte en la memoria de Jorge Pérez	0	13	13
Tecnologías utilizadas	Android	0	5	5
	SQLite	0	2	2
	Java	3	3	6
	J2EE y Servlet	4	0	4
	Apache Tomcat	2	0	2
	Mysql	1	0	1
	SoX	1	0	1
Análisis y Diseño	Introducción	1	0	1
	Metodología de desarrollo	0	1	1
	Casos de uso	5	4	9
	Requisitos	11	11	22
	Arquitectura	0	1	1
	Diagrama de clases	4	5	9
Implementación	Módulo Core Functionality	0	16	16
	Módulo Wav	0	15	15
	Módulo Fourier Transform	0	4	4
	Módulo Management	0	2	2
	Módulo Communication	2	0	2
	Módulo Servlet	2	0	2
	Módulo Analysis	15	0	15
	Módulo Transcription	15	0	15
	Módulo Snore	0	4	4
	Módulo Database Management	1	0	1
Presupuesto	Diagrama de tareas	2	0	2
	Desglose de costes	2	0	2
Conclusiones		0	2	2
Líneas Futuras		0	2	2
Glosario	Definiciones	1	1	2
	Acrónimos	1	1	2



Referencias	3	3	6
Anexos	17	17	34
Horas totales	112	112	224



Resumen

La somniloquia es una parasomnia referida al hablar en voz alta durante el sueño. Esta parasomnia ocurre tanto en la fase de sueño REM como en la NREM. La grabación de estas parasomnias puede ser útil para ayudar en el diagnóstico de ciertas patologías psicológicas, dado que pueden reflejar el estado de ansiedad o algunos comportamientos que podrían ser identificados como criterios de diagnóstico psicológico.

En este proyecto desarrollamos una plataforma móvil que únicamente graba los momentos en los que una persona habla mientras está dormida (excluyendo otros ruidos o sonidos) y los analiza para identificar la emoción principal de la voz, además de realizar una transcripción de la conversación que se presenta.

Palabras clave: psicología, somniloquia, Android, análisis emocional, grabación inteligente, transcripción



Abstract

Somniloquy is a parasomnia that refers to talking aloud while sleep. This parasomnia occurs in REM and NREM sleep phase. The recording of these parasomnias could be useful to help in the diagnosis of certain psychological pathologies given that they can reflect anxiety status or some behaviors which could be identified as psychological diagnosis criterions.

In this project we develop a mobile platform which only records sleep-talking (excluding other noises or sounds) and analyze them to identify the main emotion in the voice, allowing to make a transcription of the conversation is presented.

Keywords: psychology, somniloquy, Android, emotions analysis, smart recording, transcription.



Índice general

1. INTRODUCCIÓN Y OBJETIVOS.....	24
1.1 Introducción.....	25
1.2 Motivación	26
1.3 Objetivos	27
1.4 Posibles aplicaciones del sistema	27
1.5 Medios empleados	28
1.6 Esquema de la memoria	30
2. ESTADO DEL ARTE	32
2.1 Estado del arte en análisis de emociones	33
2.1.1 La emoción y sus clasificaciones	33
2.1.2 Técnicas usadas en reconocimiento de emociones.....	34
2.1.3 Software para el reconocimiento de emociones en la voz	36
2.2 Estado del arte en reconocimiento automático del habla.....	37
2.2.1 Introducción	37
2.2.2 Técnicas más utilizadas en el Reconocimiento del Habla	38
2.2.3 Sistemas de Reconocimiento de Voz.....	39
3. TECNOLOGÍAS UTILIZADAS.....	46
3.1 Java	47
3.1.1 Historia.....	47
3.1.2 Características	48
3.1.3 Máquina virtual de Java (JVM).....	48
3.1.4 Kit de desarrollo y entorno de ejecución	49
3.1.5 Versiones.....	49
3.1.6 Distribuciones de Java	52



3.2	J2EE y Servlet	54
3.2.1	J2EE	54
3.2.2	Servlet	59
3.3	Apache Tomcat.....	63
3.3.1	Introducción	63
3.3.2	Historia.....	63
3.3.3	Estructura	63
3.3.4	Componentes	64
3.3.5	Características y evolución	64
3.4	Mysql	66
3.4.1	Introducción	66
3.4.2	Historia.....	66
3.4.3	Características	67
3.5	SoX	68
3.5.1	Introducción	68
3.5.2	Historia.....	68
3.5.3	Características	69
4.	ANÁLISIS Y DISEÑO	71
4.1	Introducción.....	72
4.2	Metodología de desarrollo.....	72
4.3	Casos de Uso	72
4.3.1	Casos de uso de Gestión de Usuarios	73
4.3.2	Casos de uso de Gestión de Configuración de los valores de grabación	75
4.3.3	Casos de uso de la funcionalidad principal	77
4.3.4	Casos de uso navegación.....	80
4.4	Requisitos.....	81
4.4.1	Requisitos de Usuario	81
4.4.2	Requisitos de Software	95
4.5	Arquitectura.....	118
4.6	Diagrama de clases	119
4.6.1	Módulo Core Functionality	119
4.6.2	Módulo Wav.....	122
4.6.3	Módulo Fourier Transform	123
4.6.4	Módulo Management.....	123



4.6.5	Módulo Communication	124
4.6.6	Módulo Servlet	125
4.6.7	Módulo Analysis	126
4.6.8	Módulo Transcription	128
4.6.9	Módulo Snore	128
4.6.10	Módulo Database Management.....	129
5.	IMPLEMENTACIÓN	131
5.1	Dispositivo móvil	132
5.1.1	Módulo de funcionalidad principal.....	132
5.1.2	Módulo de gestión	132
5.1.3	Módulo de comunicación.....	132
5.2	Servidor.....	134
5.2.1	Módulo Servlet	134
5.2.2	Módulo de gestión de la base de datos	136
5.2.3	Módulo de eliminación de ruidos	137
5.2.4	Modulo de Transcripción	137
5.2.5	Módulo de análisis de emociones.....	147
6.	PRESUPUESTO	162
6.1	Diagrama de tareas	163
6.2	Desglose de costes	165
6.2.1	Personal.....	165
6.2.2	Material	165
6.2.3	Transporte	167
6.2.4	Resumen de costes.....	167
6.2.5	Totales.....	167
7.	CONCLUSIONES.....	170
8.	LÍNEAS FUTURAS.....	174
9.	GLOSARIO DE TÉRMINOS Y ACRÓNIMOS	178
9.1	Glosario	179
9.2	Acrónimos	180
10.	BIBLIOGRAFÍA Y REFERENCIAS.....	184
11.	ANEXOS.....	189
11.1	Anexo 1	190
11.2	Anexo 2.....	220



Índice de ilustraciones

Ilustración 1. Componentes de Java SE 7	52
Ilustración 2. Arquitectura J2EE.....	55
Ilustración 3. Ciclo de vida de un servlet	60
Ilustración 4. Casos de uso de Gestión de Usuarios	73
Ilustración 5. Casos de uso de configuración de valores de grabación.....	75
Ilustración 6. Casos de uso de la funcionalidad principal	77
Ilustración 7. Casos de uso de navegación	80
Ilustración 8. Arquitectura cliente-servidor.....	118
Ilustración 9. Diagrama de clases de Core Functionality	119
Ilustración 10. Diagrama de clases de Wav	122
Ilustración 11. Diagrama de clases de Fourier Transform	123
Ilustración 12. Diagrama de clases de Management.....	123
Ilustración 13. Diagrama de clases de Communication	124
Ilustración 14. Diagrama de clases de Servlet	125
Ilustración 15. Diagrama de clases de Analysis	126
Ilustración 16. Diagrama de clases de Transcription	128
Ilustración 17. Diagrama de clases de Snore.....	128
Ilustración 18. Diagrama de clases de Database Management.....	129
Ilustración 19. Peticiones al servlet	132
Ilustración 20. Funcionalidad del Servlet Simple	134
Ilustración 21. Funcionalidad del Servlet Ficheros	135
Ilustración 22. Voxforge	141
Ilustración 23. Subtítulos en YouTube	144
Ilustración 24. Musicg en svn.....	149



Ilustración 25. Forma canónica del formato de fichero WAVE.....	150
Ilustración 26. Flujo del reconocimiento de emociones	157
Ilustración 27. Frecuencias de la voz humana	158
Ilustración 28. Tareas del proyecto	163
Ilustración 29. Diagrama de Gantt.....	164



Índice de tablas

Tabla 1. Hardware utilizado	28
Tabla 2. Software utilizado	30
Tabla 3. Tecnologías Java EE de la capa web	56
Tabla 4. Tecnologías Java EE de la capa de negocio	57
Tabla 5. Tecnologías Java EE de la capa de sistemas de información	58
Tabla 6. CU-01	74
Tabla 7. CU-02	74
Tabla 8. CU-03	75
Tabla 9. CU-04	75
Tabla 10. CU-05	76
Tabla 11. CU-06	76
Tabla 12. CU-07	77
Tabla 13. CU-08	78
Tabla 14. CU-09	78
Tabla 15. CU-10	79
Tabla 16. CU-11	79
Tabla 17. CU-12	79
Tabla 18. CU-13	80
Tabla 19. CU-14	80
Tabla 20. CU-15	80
Tabla 21. CU-16	81
Tabla 22. CU-17	81
Tabla 23. RUC-A01	82
Tabla 24. RUC-A02	82
Tabla 25. RUC-A03	82
Tabla 26. RUC-A04	82
Tabla 27. RUC-A05	82



Tabla 28. RUC-A06	83
Tabla 29. RUC-A07	83
Tabla 30. RUC-A08	83
Tabla 31. RUC-A09	83
Tabla 32. RUC-A10	83
Tabla 33. RUC-A11	83
Tabla 34. RUC-A12	84
Tabla 35. RUC-A13	84
Tabla 36. RUC-A14	84
Tabla 37. RUC-A15	84
Tabla 38. RUC-A16	84
Tabla 39. RUC-A17	84
Tabla 40. RUC-A18	85
Tabla 41. RUC-A19	85
Tabla 42. RUC-A20	85
Tabla 43. RUC-A21	85
Tabla 44. RUC-A22	85
Tabla 45. RUC-A24	85
Tabla 46. RUC-W01	86
Tabla 47. RUC-W02	86
Tabla 48. RUC-W03	86
Tabla 49. RUC-W04	86
Tabla 50. RUC-W05	86
Tabla 51. RUC-W06	87
Tabla 52. RUC-W07	87
Tabla 53. RUC-W08	87
Tabla 54. RUC-W09	87
Tabla 55. RUC-W10	87
Tabla 56. RUC-W11	87
Tabla 57. RUC-W12	88
Tabla 58. RUC-W13	88
Tabla 59. RUC-W14	88
Tabla 60. RUC-W15	88
Tabla 61. RUC-W16	88
Tabla 62. RUR-A01	89
Tabla 63. RUR-A02	89
Tabla 64. RUR-A03	89
Tabla 65. RUR-A04	89
Tabla 66. RUR-A05	89
Tabla 67. RUR-A06	90
Tabla 68. RUR-A07	90
Tabla 69. RUR-A08	90



Tabla 70. RUR-A09	90
Tabla 71. RUR-A10	90
Tabla 72. RUR-A11	91
Tabla 73. RUR-A12	91
Tabla 74. RUR-A13	91
Tabla 75. RUR-A14	91
Tabla 76. RUR-A15	91
Tabla 77. RUR-A16	92
Tabla 78. RUR-A17	92
Tabla 79. RUR-A18	92
Tabla 80. RUR-A19	92
Tabla 81. RUR-A20	92
Tabla 82. RUR-W01	93
Tabla 83. RUR-W02	93
Tabla 84. RUR-W03	93
Tabla 85. RUR-W04	93
Tabla 86. RUR-W05	93
Tabla 87. RUR-W06	94
Tabla 88. RUR-W07	94
Tabla 89. RUR-W08	94
Tabla 90. RUR-W09	94
Tabla 91. RUR-W10	94
Tabla 92. RUR-W11	95
Tabla 93. RUR-W12	95
Tabla 94. RUR-W13	95
Tabla 95. RUR-W14	95
Tabla 96 RSF-A01	96
Tabla 97 RSF-A02	96
Tabla 98 RSF-A03	96
Tabla 99 RSF-A04	97
Tabla 100 RSF-A05	97
Tabla 101 RSF- A06	97
Tabla 102 RSF- A07	97
Tabla 103 RSF- A08	97
Tabla 104 RSF- A09	98
Tabla 105 RSF- A10	98
Tabla 106 RSF- A11	98
Tabla 107 RSF- A12	98
Tabla 108 RSF-A13	98
Tabla 109 RSF-A14	99
Tabla 110 RSF-A15	99
Tabla 111 RSF-A16	99



Tabla 112 RSF-A17	99
Tabla 113 RSF-A18	99
Tabla 114 RSF-A19	100
Tabla 115 RSF-A20	100
Tabla 116 RSF-A21	100
Tabla 117 RSF-A22	100
Tabla 118. RSF-A23	100
Tabla 119. RSF-A24	101
Tabla 120. RSF-A25	101
Tabla 121. RSF-A26	101
Tabla 122 RSF-W01	101
Tabla 123 RSF-W02	102
Tabla 124 RSF-W02	102
Tabla 125 RSF-W02	102
Tabla 126 RSF-W05	102
Tabla 127 RSF-W06	102
Tabla 128 RSF-W07	103
Tabla 129. RSF-W08	103
Tabla 130. RSF-W09	103
Tabla 131. RSF-W10	103
Tabla 132. RSF-W11	103
Tabla 133. RSF-W12	104
Tabla 134. RSF-W13	104
Tabla 135. RSF-W14	104
Tabla 136. RSF-W15	104
Tabla 137. RSF-W16	104
Tabla 138. RSF-W17	105
Tabla 139. RSF-W18	105
Tabla 140. RSF-W19	105
Tabla 141. RSNF-A01	105
Tabla 142 RSNF-A02	106
Tabla 143 RSNF-A03	106
Tabla 144. RSNF-A04	106
Tabla 145. RSNF-A05	106
Tabla 146. RSNF-A06	106
Tabla 147 RSNF-A07	107
Tabla 148 RSNF-A08	107
Tabla 149 RSNF-A09	107
Tabla 150. RSNF-A10	107
Tabla 151. RSNF-A11	108
Tabla 152 RSNF-A12	108
Tabla 153 RSNF-A13	108



Tabla 154 RSNF-A14	108
Tabla 155 RSNF-A15	108
Tabla 156 RSNF-A16	109
Tabla 157 RSNF-A17	109
Tabla 158 RSNF-A18	109
Tabla 159 RSNF-A19	109
Tabla 160 RSNF-A20	109
Tabla 161. RSNF-A20	110
Tabla 162 RSNF-A21	110
Tabla 163. RSNF-A22	110
Tabla 164. RSNF-A23	110
Tabla 165. RSNF-A23	111
Tabla 166. RSNF-A24	111
Tabla 167. RSNF-A25	111
Tabla 168. RSNF-A26	111
Tabla 169. RSNF-A27	111
Tabla 170. RSNF-A28	112
Tabla 171. RSNF-A29	112
Tabla 172. RSNF-A30	112
Tabla 173 RSNF- A31	112
Tabla 174 RSNF- A32	113
Tabla 175. RSNF-A33	113
Tabla 176. RSNF-A34	113
Tabla 177 RSNF-W01	113
Tabla 178 RSNF-W02	114
Tabla 179 RSNF-W03	114
Tabla 180 RSNF-W04	114
Tabla 181 RSNF-W05	114
Tabla 182 RSNF-W06	114
Tabla 183 RSNF-W07	115
Tabla 184 RSNF-W08	115
Tabla 185 RSNF-W09	115
Tabla 186 RSNF-W10	116
Tabla 187 RSNF-W11	116
Tabla 188. RSNF-W12.....	117
Tabla 189. RSNF-W13.....	117
Tabla 190. RSNF-W14.....	118
Tabla 191. RSNF-W15.....	118
Tabla 192. Sphinx - transcripción primera prueba	141
Tabla 193. Sphinx - transcripción segunda prueba	142
Tabla 194. Sphinx - transcripción tercera prueba	142
Tabla 195. Sphinx - transcripción cuarta prueba	142



Tabla 196. Sphinx - transcripción quinta prueba.....	142
Tabla 197. Sphinx - transcripción sexta prueba	143
Tabla 198. Sphinx - transcripción séptima prueba	143
Tabla 199. ASR Google - transcripción primera prueba.....	146
Tabla 200. ASR Google - transcripción segunda prueba	146
Tabla 201. ASR Google - transcripción tercera prueba	146
Tabla 202. ASR Google - transcripción cuarta prueba	146
Tabla 203. ASR Google - transcripción quinta prueba	146
Tabla 204. ASR Google - transcripción sexta prueba.....	146
Tabla 205. ASR Google - transcripción séptima prueba.....	147
Tabla 206. Comparativa de los valores neutro y alegría	154
Tabla 207. Comparativa de los valores neutro y enfado	154
Tabla 208. Comparativa de los valores neutro y tristeza	155
Tabla 209. Comparativa de los valores neutro y miedo.....	155
Tabla 210. Medias de todos los actores por emociones.....	155
Tabla 211. Porcentaje de éxito de las pruebas.....	159
Tabla 212. Porcentaje de éxito de la emoción alegría	159
Tabla 213. Porcentaje de éxito de la emoción alegría	160
Tabla 214. Porcentaje de éxito de la emoción alegría	160
Tabla 215. Porcentaje de éxito de la emoción tristeza.....	160
Tabla 216. Coste de los equipos	165
Tabla 217. Coste de las licencias	166
Tabla 218. Resumen de los costes	167
Tabla 219. Coste total	167
Tabla 221. Glosario.....	180
Tabla 222. Acrónimos.....	182
Tabla 223. Actor 1, valores neutro y alegría	190
Tabla 224. Actor 1, comparaciones entre neutro y alegría.....	190
Tabla 225. Actor 1, valores neutro y enfado	191
Tabla 226. Actor 1, comparaciones entre neutro y enfado.....	191
Tabla 227. Actor 1, valores neutro y enfado	191
Tabla 228. Actor 1, comparaciones entre neutro y enfado.....	192
Tabla 229. Actor 1, valores neutro y miedo	192
Tabla 230. Actor 1, comparaciones entre neutro y miedo	192
Tabla 231. Actor 2, valores neutro y alegría	193
Tabla 232. Actor 2, comparaciones entre neutro y alegría.....	193
Tabla 233. Actor 2, valores neutro y enfado	193
Tabla 234. Actor 2, comparaciones entre neutro y enfado.....	194
Tabla 235. Actor 2, valores neutro y miedo	194
Tabla 236. Actor 2, comparaciones entre neutro y miedo	194
Tabla 237. Actor 2, valores neutro y tristeza	195
Tabla 238. Actor 2, comparaciones entre neutro y miedo	195



Tabla 239. Actor 3, valores neutro y alegría	195
Tabla 240. Actor 3, comparaciones entre neutro y alegría.....	196
Tabla 241. Actor 3, valores neutro y enfado	196
Tabla 242. Actor 3, comparaciones entre neutro y enfado.....	196
Tabla 243. Actor 3, valores neutro y miedo	197
Tabla 244. Actor 3, comparaciones entre neutro y miedo	197
Tabla 245. Actor 3, valores neutro y tristeza	197
Tabla 246. Actor 3, comparaciones entre neutro y tristeza	198
Tabla 247. Actor 4, valores neutro y alegría	198
Tabla 248. Actor 4, comparaciones entre neutro y alegría.....	198
Tabla 249. Actor 4, valores neutro y enfado	199
Tabla 250. Actor 4, comparaciones entre neutro y enfado.....	199
Tabla 251. Actor 4, valores neutro y miedo	199
Tabla 252. Actor 4, comparaciones entre neutro y miedo	200
Tabla 253. Actor 4, valores neutro y tristeza	200
Tabla 254. Actor 4, comparaciones entre neutro y tristeza	200
Tabla 255. Actor 5, valores neutro y alegría	201
Tabla 256. Actor 5, comparaciones entre neutro y alegría.....	201
Tabla 257. Actor 5, valores neutro y enfado	201
Tabla 258. Actor 5, comparaciones entre neutro y enfado.....	202
Tabla 259. Actor 5, valores neutro y miedo	202
Tabla 260. Actor 5, comparaciones entre neutro y miedo	202
Tabla 261. Actor 5, valores neutro y tristeza	203
Tabla 262. Actor 5, comparaciones entre neutro y tristeza	203
Tabla 263. Actor 6, valores neutro y alegría	203
Tabla 264. Actor 6, comparaciones entre neutro y alegría.....	204
Tabla 265. Actor 6, valores neutro y enfado	204
Tabla 266. Actor 6, comparaciones entre neutro y enfado.....	204
Tabla 267. Actor 6, valores neutro y miedo	205
Tabla 268. Actor 6, comparaciones entre neutro y miedo	205
Tabla 269. Actor 6, valores neutro y tristeza	205
Tabla 270. Actor 6, comparaciones entre neutro y tristeza	206
Tabla 271. Actor 7, valores neutro y alegría	206
Tabla 272. Actor 7, comparaciones entre neutro y alegría.....	206
Tabla 273. Actor 7, valores neutro y enfado	207
Tabla 274. Actor 7, comparaciones entre neutro y enfado.....	207
Tabla 275. Actor 7, valores neutro y miedo	207
Tabla 276. Actor 7, comparaciones entre neutro y miedo	208
Tabla 277. Actor 7, valores neutro y tristeza	208
Tabla 278. Actor 7, comparaciones entre neutro y tristeza	208
Tabla 279. Actor 8, valores neutro y alegría	209
Tabla 280. Actor 8, comparaciones entre neutro y alegría.....	209



Tabla 281. Actor 8, valores neutro y enfado	209
Tabla 282. Actor 8, comparaciones entre neutro y enfado.....	210
Tabla 283. Actor 8, valores neutro y miedo	210
Tabla 284. Actor 8, comparaciones entre neutro y tristeza	210
Tabla 285. Actor 8, valores neutro y tristeza	211
Tabla 286. Actor 8, comparaciones entre neutro y tristeza	211
Tabla 287. Actor 9, valores neutro y alegría	211
Tabla 288. Actor 9, comparaciones entre neutro y alegría.....	212
Tabla 289. Actor 9, valores neutro y enfado	212
Tabla 290. Actor 9, comparaciones entre neutro y enfado.....	212
Tabla 291. Actor 9, valores neutro y miedo	213
Tabla 292. Actor 9, comparaciones entre neutro y miedo	213
Tabla 293. Actor 9, valores neutro y tristeza	213
Tabla 294. Actor 9, comparaciones entre neutro y tristeza	214
Tabla 295. Actor 10, valores neutro y alegría	214
Tabla 296. Actor 10, comparaciones entre neutro y alegría.....	214
Tabla 297. Actor 10, valores neutro y enfado	215
Tabla 298. Actor 10, comparaciones entre neutro y enfado.....	215
Tabla 299. Actor 10, valores neutro y miedo	215
Tabla 300. Actor 10, comparaciones entre neutro y miedo	216
Tabla 301. Actor 10, valores neutro y tristeza.....	216
Tabla 302. Actor 10, comparaciones entre neutro y tristeza	216
Tabla 303. Actor 11, valores neutro y alegría	217
Tabla 304. Actor 11, comparaciones entre neutro y alegría.....	217
Tabla 305. Actor 11, valores neutro y enfado	217
Tabla 306. Actor 11, comparaciones entre neutro y enfado.....	218
Tabla 307. Actor 11, valores neutro y miedo	218
Tabla 308. Actor 11, comparaciones entre neutro y miedo	218
Tabla 309. Actor 11, valores neutro y tristeza.....	219
Tabla 310. Actor 11, comparaciones entre neutro y tristeza	219
Tabla 311. Actor 1 – prueba 1, matriz de confusión para alegría.....	220
Tabla 312. Actor 1 – prueba 1, matriz de confusión para enfado.....	220
Tabla 313. Actor 1 – prueba 1, matriz de confusión para miedo	220
Tabla 314. Actor 1 – prueba 1, matriz de confusión para tristeza	220
Tabla 315. Actor 1 – prueba 2, matriz de confusión para alegría.....	220
Tabla 316. Actor 1 – prueba 2, matriz de confusión para enfado.....	221
Tabla 317. Actor 1 – prueba 2, matriz de confusión para miedo	221
Tabla 318. Actor 1 – prueba 2, matriz de confusión para tristeza	221
Tabla 319. Actor 1 – prueba 3, matriz de confusión para alegría.....	221
Tabla 320. Actor 1 – prueba 3, matriz de confusión para enfado.....	221
Tabla 321. Actor 1 – prueba 3, matriz de confusión para miedo	221
Tabla 322. Actor 1 – prueba 3, matriz de confusión para tristeza	221



Tabla 323. Actor 2 – prueba 4, matriz de confusión para alegría.....	222
Tabla 324. Actor 2 – prueba 4, matriz de confusión para enfado.....	222
Tabla 325. Actor 2 – prueba 4, matriz de confusión para miedo	222
Tabla 326. Actor 2 – prueba 4, matriz de confusión para tristeza.....	222
Tabla 327. Actor 2 – prueba 5, matriz de confusión para alegría.....	222
Tabla 328. Actor 2 – prueba 5, matriz de confusión para enfado.....	222
Tabla 329. Actor 2 – prueba 5, matriz de confusión para miedo	222
Tabla 330. Actor 2 – prueba 5, matriz de confusión para tristeza.....	223
Tabla 331. Actor 3 – prueba 6, matriz de confusión para alegría.....	223
Tabla 332. Actor 3 – prueba 6, matriz de confusión para enfado.....	223
Tabla 333. Actor 3 – prueba 6, matriz de confusión para miedo	223
Tabla 334. Actor 3 – prueba 6, matriz de confusión para tristeza.....	223
Tabla 335. Actor 3 – prueba 7, matriz de confusión para alegría.....	223
Tabla 336. Actor 3 – prueba 7, matriz de confusión para enfado.....	223
Tabla 337. Actor 3 – prueba 7, matriz de confusión para miedo	224
Tabla 338. Actor 3 – prueba 7, matriz de confusión para tristeza.....	224



Capítulo 1

Introducción y objetivos

“La esperanza es el sueño del hombre despierto.”

Aristóteles (384 AC-322 AC)

1.1 Introducción

En la actualidad, la sociedad lleva un ritmo de vida frenético, existiendo una necesidad de realizar un gran número de actividades en el menor tiempo posible, llevando a las personas a situaciones estresantes y pudiendo causar problemas de salud y psicológicos. Además, en los tiempos de crisis que corren, con las grandes dificultades financieras y laborales que sufre la población, el número de individuos con ansiedad o problemas emocionales se está viendo elevado.

Existen diversos estudios, de los que hablaremos a continuación, en los que se muestra la relación que hay entre problemas psicológicos con trastornos del sueño, como la somniloquia.

Por definición, la somniloquia [1] consiste en la emisión de palabras durante el sueño, sin que exista noción subjetiva y simultánea por parte del sujeto. Suele ser un proceso poco frecuente y de corta duración, que puede aparecer en cualquier persona sin importar la edad y el sexo aunque predomina en la población infantil y la adolescente y afecta mayoritariamente a hombres.

La somniloquia se ha estudiado poco en relación con otros trastornos del sueño. Sin embargo, se plantea que factores de maduración y desarrollo son de gran importancia dentro de las causas. Se han observado antecedentes hereditarios en algunas personas y puede verse asociado a la enuresis (orinarse dormido) y el sonambulismo (caminar y, también, realizar diferentes actos mientras se duerme). Puede estar condicionado al consumo de sustancias psicoactivas, la fiebre, la sobreexcitación, el estrés emocional y trastornos afectivos.

El hablar en sueños puede aparecer tanto en la fase REM como en la NREM, y hay varios estudios, como los que realizaron Arkin [2] y Rechtschaffen [3], en los que se explica que existe relación entre la somniloquia producida en la fase REM con momentos de carácter emocional. Según estos autores, los sueños que ocurren en la fase NREM tienden a relacionarse con acontecimientos recientes, sin contenido afectivo, siendo habitualmente situaciones reales de la vida diaria.

Por otra parte, Andriani [4] en 1892, también declaró que el contenido del habla durante un sueño es normalmente emocional, y revela un deseo insatisfecho, un placer inesperado, un lamento, o más frecuentemente un estado de miedo, ansiedad, angustia o terror.

El sistema propuesto en este proyecto pretende principalmente ayudar a profesionales del área de la psicología como herramienta de ayuda para diagnosticar ciertas patologías, como la ansiedad o el estrés.

1.2 Motivación

Como se ha descrito en el punto [anterior](#), la somniloquia se relaciona de una manera muy estrecha con problemas emocionales de la persona en la que se presenta. Además, el habla en sí misma también contiene información extra-lingüística sobre características físicas, estados fisiológicos y emocionales. Darwin señaló [5] que las emociones y su expresión son innatas e instintivas y que la parte no verbal de la voz puede ser un medio de expresión de emociones. Esto se debe a que los diferentes estados emocionales de un hablante producen cambios fisiológicos en el aparato fonador y se reflejan en características de la señal de la voz.

Debido a la gran demanda de personas que buscan ayuda en manos de un psicólogo para intentar mitigar y solucionar sus problemas psicológicos y emocionales, es razonable tener a disposición herramientas o medios que faciliten la labor de estos profesionales. Sin embargo, actualmente este mercado no posee un gran desarrollo por lo que es necesaria la creación de recursos.

Este proyecto se centra en la creación de una plataforma de grabación de personas somnílocuas y el posterior análisis de los audios obtenidos para intentar clasificarlos por emociones primarias.

Se propone un modelo de grabación inteligente de las somniloquias, obteniendo únicamente los fragmentos de audio en los que se detecten palabras o frases de una determinada persona durante el proceso del sueño. Esto aporta mejoras frente a la grabación tradicional, ya que permite ahorrar en tamaño de almacenamiento, debido a que se descartan los periodos de silencio así como los ruidos que no corresponden a palabras. Esto hace que se reduzca el tiempo de duración por lo que el usuario que desee oír sus grabaciones perderá menos tiempo en este proceso.

Tal y como se comentó anteriormente, la voz contiene información emocional que se puede extraer mediante un análisis de la señal que la representa. En este proyecto se realizará un módulo encargado de la predicción de la emoción predominante en el audio.

Por último, con el objetivo de aportar una mayor cantidad de información a los profesionales de la psicología, la aplicación proporcionará un método de transcripción a



texto del archivo de audio, de forma que éstos puedan realizar una interpretación leída de las somniloquias.

1.3 Objetivos

El objetivo de este proyecto es crear un sistema que permita grabar el habla de las personas mientras duermen de un modo eficiente, para posteriormente procesar acústicamente los archivos de audios y predecir, en función de unas reglas previamente definidas, la emoción predominante del archivo y generar una transcripción del mismo.

En base a este objetivo principal, se proponen los siguientes objetivos parciales:

- Crear una aplicación en una plataforma móvil encargada de la grabación de somniloquias.
- Grabar audios de una manera optimizada evitando la generación de archivos que contengan grandes silencios y que por lo tanto sean muy pesados.
- Eliminación de ruidos en los archivos grabados.
- Manejar un sistema de comunicación entre la aplicación de la plataforma móvil y un servidor de aplicaciones.
- Detección de la emoción predominante entre las cuatro emociones primarias de enfado, alegría, tristeza y miedo.
- Transcripción a texto del audio interpretado.
- Presentar un sistema fácil e intuitivo para el usuario final.

Para lograr estos objetivos es necesario estudiar los siguientes aspectos:

- Adquirir conocimientos sobre análisis acústico de una señal.
- Adquirir conocimientos sobre la programación en Android.
- Adquirir conocimientos en programación J2EE.

1.4 Posibles aplicaciones del sistema

Al tratarse de un sistema con tanta funcionalidad, las aplicaciones que puede tener son muy variadas, entre las que podemos destacar:

- El sistema podría utilizarse para grabar a personas mientras duermen para saber si roncan o si padecen apneas, o ayudar a prevenir cualquier otro problema que pueda detectarse con audios.

- El sistema podría aplicarse para realizar grabaciones en exteriores, ya que al tratarse de ambientes con ruidos, se les facilitarían la eliminación de sonidos innecesarios.
- La aplicación podría interesar a entidades con call centers automáticos, para detectar la frustración del locutor por medio de la emoción del audio y por tanto la necesidad de remitirlo a un operador humano, o simplemente para detectar la satisfacción que puede estar produciendo la calidad del servicio.
- Atención de llamadas en servicios de emergencia con el fin de cuantificar la gravedad del asunto de la llamada.
- El sistema puede ayudar en investigaciones de psicología, psiquiatría, o neurología donde el reconocimiento de emociones puede mejorar la calidad de los estudios obteniendo mayor fiabilidad en las medidas y mayor velocidad en tareas manuales de procesamiento de datos sobre el comportamiento emocional.

1.5 Medios empleados

A lo largo del proyecto ha sido necesario emplear una serie de dispositivos físicos y aplicaciones informáticas con el fin de recopilar información, realizar tareas de análisis, diseño e implementación de código, llevar a cabo simulaciones y pruebas, documentar los distintos procesos, etc. A continuación se muestran los distintos medios utilizados de manera detallada y clasificados en función de su tipo.

Hardware:
















Hardware	Características	
	Ordenador portátil Toshiba Satellite L655	Intel Core i3-370M, 2.40 GHz, 4GB DDR3, 500GB
	Ordenador portátil Toshiba Satellite A300	Intel Core 2 Duo, 2.00GHz, 4GB DDR2, 320GB
	Teléfono móvil Samsung Galaxy i9000	Pantalla WVGA Super AMOLED de 4 pulgadas, procesador de 1GHz y el sistema operativo Android 2.1.
	Micrófono ordenador	Grabación

Tabla 1. Hardware utilizado

Software:

Software	Características	
	Windows 7 Professional	Sistema Operativo
	Windows 7 Home Premium	Sistema Operativo
	Microsoft Office Word 2007	Editor de textos
	Microsoft Excel 2007	Hoja de cálculo
	Microsoft Project Professional 2007	Planificación
	Microsoft Office Powerpoint 2007	Presentación y diagramas
	SoX	Procesamientos de audio
	Eclipse Helios	Programación
	Eclipse Índigo IDE	Programación
	Android SDK	Programación
	MySQL 5.5	Base de datos




	SQLite	Base de datos
	Apache Tomcat	Servidor web
	Java Platform (JDK) 7	Plataforma de programación

Tabla 2. Software utilizado

1.6 Esquema de la memoria

Esta sección está destinada a ofrecer una breve explicación de cada uno de los capítulos de los que consta el presente documento.

- **Estado del arte.** En este apartado hablaremos de las distintas tecnologías y estudios relacionados con el sistema propuesto.
- **Tecnologías utilizadas.** Este capítulo está dedicado a explicar las distintas herramientas y tecnologías empleadas en la realización del sistema.
- **Análisis y Diseño.** El objetivo de esta sección es la especificación de la funcionalidad a través de los casos de uso y los requisitos. Además, se explicará el diseño del sistema mediante su arquitectura y su diagrama de clases.
- **Implementación.** Este apartado está dedicado a explicar cómo se ha realizado el sistema de una manera detallada. Se especifican las decisiones que se han tomado y las dificultades encontradas en el proceso.
- **Presupuesto.** En este apartado se hace un análisis de los costes relacionados con las distintas fases de desarrollo en función del tiempo y recursos invertidos en cada una de ellas.
- **Conclusiones.** En este capítulo se comentan las propiedades del sistema y se realiza una valoración final del sistema.
- **Líneas futuras.** Para finalizar se tratarán aquellas funcionalidades que pueden ser incluidas en el futuro para convertir nuestro sistema en una herramienta más completa.





Capítulo 2

Estado del arte

*“Antes de iniciar la labor de cambiar el mundo,
da tres vueltas por tu propia casa.”*

Proverbio chino

2.1 Estado del arte en análisis de emociones

2.1.1 La emoción y sus clasificaciones

La entonación con que se emite una frase, la presencia de pausas, los puntos de inflexión y otras características presentes en la comunicación verbal, pueden transmitirnos más información acerca del estado emocional del emisor que el propio contenido lingüístico de mensaje. Además, muchos estudios han demostrado que por medio de la voz se puede reconocer varios aspectos del estado físico, tales como la edad, el sexo y la apariencia.

Algunas de las preguntas que surgen al estudiar el tema de las emociones pueden ser: ¿qué es una emoción?, o ¿todas las personas expresan de la misma manera las emociones? A continuación explicaremos brevemente, la teoría sobre la evolución de las emociones o la clasificación de las emociones en primarias y secundarias.

La definición del término emoción es la base para cualquier tipo de investigación en esta área. Según Scherer [6] las emociones son definidas como: “Episodios de cambios coordinados en varios componentes (incluyendo al menos activación neuropsicológica, expresión motriz, y sentimientos subjetivos pero posiblemente también tendencias a la acción y procesos cognitivos) en respuesta a eventos externos o internos de mayor significancia para el organismo.”

Las personas conocemos nuestras emociones gracias a su intromisión en nuestras mentes conscientes, pero las emociones no evolucionaron como sentimientos conscientes sino que evolucionaron como el resultado de especializaciones de la conducta y fisiológicas. La mayoría de los especialistas están de acuerdo en trazar una línea en el conjunto de las emociones humanas y distinguir aquellas que son primarias evolutivamente, de las que son secundarias.

Para el célebre psicólogo Theodore Ribot [7], nuestra personalidad envuelve en su profundidad el origen de la gran trinidad afectiva constituida por el miedo, la cólera y el deseo. Estos tres instintos nacieron directamente de la vida orgánica: instinto defensivo, instinto ofensivo, instinto nutricional. Desde este punto de partida se mantuvieron el miedo y la cólera, y se agregaron la alegría y la tristeza, cuatro emociones que poseen también los mamíferos superiores, y quedó entonces conformado un cuadro de cuatro emociones primarias, con su respectiva variedad de manifestaciones:

- Cólera: enojo, mal genio, atropello, fastidio, molestia, furia, resentimiento, hostilidad, animadversión, impaciencia, indignación, ira, irritabilidad, violencia y odio patológico.

- **Alegría:** disfrute, felicidad, alivio, capricho, extravagancia, deleite, dicha, diversión, estremecimiento, éxtasis, gratificación, orgullo, placer sensual, satisfacción y manía patológica.
- **Miedo:** ansiedad, desconfianza, fobia, nerviosismo, inquietud, terror, preocupación, aprehensión, remordimiento, sospecha, pavor y pánico patológico.
- **Tristeza:** aflicción, autocompasión, melancolía, desaliento, desesperanza, pena, duelo, soledad, depresión y nostalgia.

Hay otro enfoque de clasificación muy influyente, que es el de Robert Plutchik [8] en 2001, el cual expone que las emociones primarias son ocho: ira, miedo, tristeza, asco, sorpresa, la anticipación, la confianza y alegría. Plutchik propuso que estas emociones son biológicamente primitivas y han evolucionado con el fin de aumentar la capacidad reproductiva del animal.

La clasificación varía en función del investigador que la realice y no se ha llegado a un conjunto de emociones primarias validado por la comunidad, por lo que para nuestro proyecto cogeremos las cuatro emociones presentes en casi todos ellos: alegría, ira, miedo y tristeza.

Si bien hemos hablado de las emociones primarias también hablaremos de las emociones secundarias, entre las que podemos encontrar:

- **Pena:** cuidado, aflicción, dolor o sentimiento interior grande.
- **Ternura:** afectivo, cariñoso, amable.
- **Ironía:** burla.
- **Sorpresa:** asombro, estupefacción, maravilla y shock.
- **Vergüenza:** arrepentimiento, humillación, culpa, mortificación, pena, remordimiento, y vergüenza.
- **Aversión:** repulsión, asco, desdén, desprecio, menosprecio y aberración.

Otras emociones secundarias: como el temor, la queja, el anhelo, el aburrimiento, la satisfacción, la impaciencia o el ensueño también han sido objeto de estudio.

2.1.2 Técnicas usadas en reconocimiento de emociones

En esta sección se hablará de las técnicas de reconocimiento de emociones en el habla más importantes estudiadas en la actualidad. La mayoría de estas técnicas ofrecen buenos resultados en tareas de reconocimiento de locutor y de idioma.

GMM

La técnica de Modelos de Mezcla de Gaussianas (GMM) aplicada al reconocimiento automático de emociones se basa en el principio de que las emociones tienen diferentes sonidos y que la frecuencia de aparición de los sonidos (\vec{x}) es diferente de una emoción a otra. Los GMM modelan la distribución de probabilidad de los parámetros de un fragmento de audio. La función de densidad de probabilidad de un GMM se define como una suma ponderada de M densidades y está dada por la siguiente ecuación:

$$p(\vec{x}|\lambda) = \sum_{i=1}^M p_i b_i(\vec{x})$$

Donde (\vec{x}) es un vector de D-dimensiones aleatorias, $b_i(\vec{x})$ donde $i = 1-M$, son las componentes de densidad y p_i donde $i = 1-M$ son las medias de la mezcla. Cada componente de densidad es una función gaussiana D-variada de la forma:

$$b_i(\vec{x}) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(\vec{x} - \vec{\mu}_i)' \Sigma_i^{-1} (\vec{x} - \vec{\mu}_i)\right\}$$

Con un vector de medias $\vec{\mu}_i$ y matriz de covarianza Σ_i . El peso de la mezcla satisface que $\sum_{i=1}^M p_i = 1$.

La densidad de toda la mezcla de Gaussianas es parametrizada con el vector de medias, la matriz de covarianzas y la mezcla de medias de todos los componentes de densidad.

SVM

Las Maquinas de Vectores Soporte (SVM o Support Vector Machines) son un tipo de clasificador de patrones binarios, desarrollados por Vladimir Vapnik [9] y su equipo en los laboratorios AT&T, cuyo objetivo es asignar cada patrón a una clase.

La SVM mapean los puntos de entrada a un espacio de características de una dimensión mayor, para luego encontrar el hiperplano que los separe y maximice el margen entre las clases.

Pertenecen a la familia de clasificadores lineales puesto que inducen separadores lineales o hiperplanos en espacios de características de muy alta dimensionalidad (introducidos por funciones núcleo o kernel) con un sesgo inductivo muy particular.

Modelos Ocultos de Márkov (HMM)

Un modelo oculto de Márkov [10] es un modelo estadístico en el que se asume que el sistema a modelar es un proceso de Márkov de parámetros desconocidos. El objetivo es determinar los parámetros desconocidos u ocultos de dicha cadena a partir de los parámetros observables. Un HMM puede considerarse como una red bayesiana dinámica muy simple.

El proceso comienza en un estado inicial, específicamente diseñado para ello, y cada uno de los estados va a tener asociado un conjunto de probabilidades sobre un grupo de símbolos salientes. Por cada una de las ejecuciones, se va a elegir una transición hacia un estado nuevo, y se va a generar un símbolo de salida relacionado con dicho estado. La elección de las ejecuciones de cada transición y símbolo se va a realizar en función de probabilidades, y por tanto va a ser una elección aleatoria. La característica principal de los modelos de Márkov es que no se conoce el conjunto de estados por los que el proceso ha realizado el recorrido hasta llegar al conjunto de símbolos obtenidos en la salida, y este es el motivo fundamental por el que se le conoce como Modelo oculto de Márkov.

Dado el buen funcionamiento de esta técnica en tareas como el reconocimiento de habla, también se ha aplicado al reconocimiento de emociones combinado con otras técnicas de clasificación como los GMM o SVM.

2.1.3 Software para el reconocimiento de emociones en la VOZ

Aunque no hay muchos programas que realicen esta labor, a continuación exponemos algunos de ellos:



EmoSpeech

EMOSpeech [11] es una familia de productos de software orientados a call centers que deseen incrementar la eficiencia de sus procesos de control de calidad y mejorar el desempeño de sus campañas.

La tecnología está basada en el reconocimiento de emociones y es fácilmente configurable dependiendo de las necesidades del usuario y provee información valiosa para operadores telefónicos y analistas de calidad.

Esta plataforma ha sido desarrollada por los Doctores Luis Villaseñor y Carlos Reyes, Aarón Pancardo y Humberto Pérez. Emospeech está orientado -por el momento-

a cubrir los requerimientos de los cerca de 3,000 call centers en México y en donde se llevan a cabo campañas telefónicas para más de 20,000 empresas, y tiene como objetivo crear estrategias que incrementen y mejoren la efectividad y calidad de los servicios de los call centers, de estas campañas y que se mejore la relación con los clientes/usuarios, a través de la interpretación de las emociones.



eXaudios [12] ha desarrollado unos programas informáticos que pueden analizar la voz y determinar tanto la personalidad instintiva-emocional del hablante como su actitud emocional.

El proceso de cálculo de estos programas se basa en vínculos establecidos entre las actitudes emocionales y parámetros físicamente medibles en la voz. El sistema es preciso y rápido, y unos 30-segundos de audio son suficientes para realizar el análisis.

Puede aplicarse como una solución en áreas de comercialización, servicio y ventas o para impulsar el rendimiento de un call center.

2.2 Estado del arte en reconocimiento automático del habla

2.2.1 Introducción

Un sistema de reconocimiento de voz es una herramienta computacional que realiza el procesamiento de la señal de voz emitida por el ser humano y reconoce la información contenida en ésta. Estos datos puede o bien convertirlo en texto o bien emitir órdenes para que actúen sobre un proceso. En un sistema de reconocimiento de voz intervienen diversas disciplinas, tales como: la fisiología, la acústica, el procesamiento de señales, la inteligencia artificial y la ciencia de la computación.

Si bien el proceso tiene como función obtener la secuencia de palabras asociada a la frase en lenguaje natural de entrada, las personas, al hablar, normalmente lo hacen de forma continua, es decir, sin pausas entre algunas palabras. A menudo, también existen problemas gramaticales, que incluyen elementos propios del habla espontánea como son las interjecciones, falsos comienzos, repeticiones, etc. Esto supone que el proceso de reconocimiento de habla, presente problemas de cobertura léxica y sintáctica. Por estas razones, la tarea del sistema de reconocimiento de habla es difícil, y además, es costosa, tanto en memoria como en cálculo.

El principal problema en los sistemas de reconocimiento de voz ha sido determinar cuáles son las posibles causas que hacen tan difícil llevar a cabo este proceso en condiciones generales. Algunas de las causas que se han conseguido averiguar son:

- La variación fonética de los hablantes. Ninguna persona habla igual que otra, es decir, los sonidos producidos por diferentes personas no suenan igual.
- Las ambigüedades acústicas. A veces no es posible mapear eventos acústicos a sus símbolos fonéticos correspondientes, no es posible realizar una buena decodificación ya que no se dispone de todas las características o fuentes de conocimiento que una persona utiliza durante el proceso de una conversación.
- Mala pronunciación de algunas palabras. A veces ciertas palabras de duración breve se omiten, como preposiciones o conjunciones, o se transforman en sonidos extraños.
- Rapidez al pronunciar las palabras. Con esto se consigue que no haya una clara transición entre las diferentes sílabas, llegando a la fusión u omisión de algunas de ellas.
- Coarticulación. Las características acústicas de los sonidos se ven afectadas por el contexto en el que se encuentran. La mayoría de estos efectos se traducen en alófonos que tiene cada fonema. Ello supone la necesidad de tener múltiples patrones que tengan en cuenta estas variaciones.
- Variaciones temporales. La duración de una palabra e incluso de los sonidos puede cambiar, generando la necesidad de realizar alineamientos dinámicos, que permitan tener en cuenta a estas posibles variaciones.
- Ruidos e interferencias. Las personas podemos hablar en ambientes no idóneos, con ruidos o en presencia de otros sonidos interferentes.

2.2.2 Técnicas más utilizadas en el Reconocimiento del Habla

Los estudios científicos asociados a esta área de conocimiento aportan diversas técnicas de clasificación de patrones de voz. A continuación, explicaremos aquellas que han obtenido los mejores resultados y que más prometedoras parecen:

- **Comparación de Plantillas o Patrones utilizando técnicas de Programación Dinámica (DTW).** Esta técnica consiste en realizar una comparación entre la señal acústica de entrada con los patrones o plantillas de las que dispone el sistema. Para realizar esta tarea se parametriza la señal recibida y se transforma en coeficientes espectrales, los cuales se cotejan con los patrones almacenados. Esta técnica, es utilizada tanto para resolver problemas de reconocimiento de habla continua como aislada e incluso con una cierta independencia del locutor. Sin embargo esta técnica suele tener algunos problemas, ya que: la duración de

la palabra puede variar, por lo que puede que no coincida con la de la plantilla; y el ritmo con el que se realiza la pronunciación no tiene que mantenerse constante por lo que no se ajustará a la plantilla en ese sentido, ya que este depende de la persona.

- **Modelos Ocultos de Márkov (HMM).** Al igual que en el análisis de emociones, los HMM también pueden aplicarse en el reconocimiento automático del habla. De tal manera, que cada estado va a indicar cuáles son aquellos sonidos más probables para cada segmento del habla, mientras que las transiciones van a ser restricciones temporales para cada uno de esos sonidos, indicando cuáles son sus secuencias de apariciones. Esta técnica soluciona problema respecto al modelo anterior y proporciona los mejores resultados hasta la fecha tanto para el reconocimiento de habla aislada como continua y para independencia del locutor.
- **Redes Neuronales (NN).** Las redes neuronales son estructuras de procesamiento paralelo de información, formadas por numerosos nodos simples conectados entre sí mediante pesos y agrupados en diferentes capas, entre las que se deben distinguir la capa de entrada y la capa de salida. Debido a su naturaleza intrínsecamente no lineal, a su capacidad de clasificación, y sobre todo a la capacidad que tienen para aprender una determinada tarea a partir de pares observación-objetivo sin hacer suposición alguna sobre el modelo subyacente, se han convertido en una de las herramientas más atractivas para la solución del problema del reconocimiento de habla. Sin embargo, al principio el estudio de las redes neuronales fue abandonado, debido a que no se podía llevar a cabo su entrenamiento con algoritmos que fuesen eficientes. En la actualidad ha quedado perfectamente demostrado, que los modelos basados en las redes neuronales cuentan con una gran potencia desde el punto de vista computacional y se han conseguido resultados comparables a los obtenidos con otros métodos ya clásicos como los HMM. Sin embargo, las NN presentan diferentes problemas o inconvenientes como el desconocimiento a priori de la estructura de capas y el número de nodos necesarios, un tiempo a veces muy elevado para su entrenamiento y la posibilidad de que se mantenga en los mínimos locales de las funciones de coste usadas durante el entrenamiento de la red. Esto implica que se haga necesario combinar dichos sistemas con técnicas basadas en programación dinámica y en modelos ocultos de Márkov [10].

2.2.3 Sistemas de Reconocimiento de Voz

Los sistemas de reconocimiento automático del habla comenzaron a utilizarse en la década de los 90s, cuando empresas tan influyentes como AT&T o Blue Cross y Blue Shield la utilizaron en sus procesos. Pero ha sido ahora, con la llegada de los nuevos dispositivos móviles y la incorporación a navegadores, cuando esta tecnología ha cobrado de nuevo una gran importancia.

A continuación, mostraremos una lista resumida de los sistemas de reconocimiento de voz de uso más frecuente en computadores personales y en dispositivos móviles.



Dragon NaturallySpeaking

Es una de las herramientas más potentes del mercado y una de las más usadas en la computación personal para el dictado de texto y de comandos, es así mismo uno de los más antiguos y más reconocidos del mercado.

Nuance Communications, desarrolló para su primera versión, DragonDictate para DOS, un sistema basado en modelos ocultos de Márkov [10] que funcionaba en computadores cuya capacidad de procesamiento era mínima, aunque con una tasa de precisión del reconocimiento baja y restricciones específicas sobre la pronunciación de las palabras.

Dragon Systems lanzó la primera versión de NaturallySpeaking 1.0 [13] como producto de dictado continuo en 1997. Su última versión, lanzada en junio de 2011, cuenta con capacidades mejoradas para el control por voz de computadores y el dictado casi en tiempo real. Para agosto de 2012 anunciaron la versión 12.0, aunque de momento en su página web no aparece a la venta.

Las principales características que la empresa Nuance destaca de este software son:

- Permite interactuar con su PC mediante la voz, y puede crear textos tres veces más rápido que con el teclado, con una precisión hasta de un 99%.
- Aumentar la productividad y reducir costes en una organización.
- Dictar, modificar y dar formato a documentos y hojas de cálculo.
- Redactar, enviar y administrar mensajes de correo electrónico.
- Realizar búsquedas en Internet o en su escritorio.
- Crear citas y reuniones.
- Personalizar el vocabulario y los comandos para que reflejen la terminología y el flujo de trabajo exclusivos de su organización.
- Crear comandos de voz personalizados para automatizar los procesos comerciales repetitivos.
- Utilizar una grabadora de voz digital para capturar notas y transcribirlas automáticamente al regresar a su PC.

Este software es comercial, y se consigue para Windows con el nombre de Dragon NaturallySpeaking y para Macintosh bajo el nombre de DragonDictate. También puede usarse en Linux a través de Wine aunque puede dar diversos errores en función de la

versión que se utilice. Nuance, ha sacado recientemente aplicaciones para dispositivos móviles Android e iOS.



CMU Sphinx

CMU Sphinx [14] o comúnmente conocido como Sphinx, es un grupo de sistemas opensource de reconocimiento de voz desarrollado en la Universidad de Carnegie Mellon. Incluye una serie de programas para reconocimiento de voz y un entrenador de modelo acústico.

Sphinx ha desarrollado a lo largo de los años diferentes productos y versiones para el reconocimiento de voz. Los diferentes decodificadores pueden considerarse como ramas independientes entre sí ya que difieren en lenguaje en el que se escriben y en las opciones. La decisión acerca de cuál es la versión que debe utilizar un usuario, depende de cómo esté familiarizado con los lenguajes que utilizan, ya que por ejemplo PocketSphinx está desarrollado con C/Python o Sphinx4 en Java.

Los decodificadores de voz vienen con modelos acústicos y aplicaciones de ejemplo. Los recursos disponibles incluyen además el software para el entrenamiento de modelos acústicos, la compilación de un modelo de lenguaje y un diccionario de pronunciación de dominio público llamado cmudict.

PocketSphinx

PocketSphinx CMU es el más rápido de los sistemas de reconocimiento de voz de Sphinx. Es una librería escrita en C, por lo que es óptima para el desarrollo de aplicaciones en este lenguaje. Además en tiempo real es el motor más preciso, y por lo tanto es una buena opción para aplicaciones interactivas.

Puede utilizarse para aplicaciones de escritorio, de control y de dictado donde los objetivos son una respuesta rápida y un bajo consumo de recursos. También incluye soporte para dispositivos embebidos y se utiliza con éxito en el iPhone, en los dispositivos Nokia Maemo y en Windows Mobile.

Sphinx-4

Sphinx-4 está escrito completamente en el lenguaje de programación Java, con el objetivo de proporcionar un marco más flexible para la investigación en reconocimiento de voz. Además al estar escrito en Java puede utilizarse en una gran diversidad de sistemas operativos. Sin embargo, el sistema ha evolucionado más allá que una simple reescritura para crecer de forma sustancial, y se han marcado como objetivos:

- El desarrollo de un nuevo entrenador (modelo acústico).
- Implementación de adaptación de la persona que está hablando.

- Mejora de la gestión de configuración.
 - La creación de una interfaz de usuario gráfica basada en el diseño gráfico de sistemas.

Sphinx-3

Sphinx-3 es el sistema de reconocimiento de voz de CMU con más vocabulario. Es el decodificador más viejo basado en C que siguen manteniendo. Es el decodificador más preciso para tareas con vocabularios muy grandes, pero no reconocimiento en tiempo real. La evolución reciente ha hecho que Sphinx funcione casi en tiempo real, aunque todavía no es adecuado para las aplicaciones interactivas. Sphinx 3 está en desarrollo y en colaboración con SphinxTrain proporciona acceso a una serie de técnicas de modelado modernas, como LDA / MLLT, MLLR y VTLN, que mejoran la precisión en el reconocimiento.

Sphinx-2

En la wiki de CMU Sphinx aparece Sphinx-2 como obsoletos, y no recomiendan su uso, aunque entienden que siga teniendo interés para los investigadores de reconocimiento de voz.

Sphinx-2 es un sistema de reconocimiento de voz rápido, el predecesor de PocketSphinx. No se está desarrollando activamente en este momento, pero es todavía ampliamente utilizado en las aplicaciones interactivas. Utiliza modelos ocultos de Márkov [10] (HMM) y un lenguaje de modelado estadístico de n-gramas. Aunque no es tan preciso como Sphinx-3 o Sphinx-4, se ejecuta en tiempo real, y por lo tanto es una buena elección para aplicaciones interactivas.



Julius

Julius [15] es una aplicación para reconocimiento de voz publicado como software libre con licencia BSD. Es uno de los motores de reconocimiento experimentales más antiguos y con mayor desarrollo.

Julius nace de la necesidad de crear un sistema que permitiese el reconocimiento del habla japonesa, mediante tecnologías experimentales que, en su momento, no habían sido probadas. El proyecto surge en 1997 como un experimento de la Universidad de Kyoto para demostrar la capacidad de reconocer un gran vocabulario en muestras de voz que contuviesen habla continua en japonés (LVCSR). El trabajo continuó con el proyecto de herramientas de dictado japonesas IPA (1997-2000), el Consorcio de Reconocimiento de Habla Continua (CSRC), Japón (2000-2003) y actualmente con el Consorcio de Tecnologías Interactivas de Voz (ISTC). La versión más reciente de Julius es la 4.2.2, disponible desde el primero de agosto de 2012.

Julius está basado en N-gramas de palabras y HMM dependientes del contexto e incorpora las técnicas de búsquedas más importantes. Entre sus principales características destacan:

- Tiempo real, alta velocidad, reconocimiento preciso.
- Requisitos de memoria baja.
- Altamente configurable.
- Soporta LM de N-gram, gramática, y palabras aisladas.

El proyecto tiene una licencia de código abierto, aunque el nombre tiene derechos de autor y ciertas partes de su código tienen restricciones de modificación para uso comercial. Este motor funciona en Windows y en Linux, y proporciona al usuario o desarrollador un modelo acústico y de lenguaje para que pueda trabajar con Julius.



Windows Speech Recognition

Es la aplicación de reconocimiento de voz incluida en las versiones Vista y 7 del sistema operativo Windows de Microsoft.

Microsoft ofreció un motor de reconocimiento de voz en otros sistemas operativos como el Windows 2000 y XP llamado Speech API, pero desafortunadamente, este sistema era solamente una librería de programación para que otros desarrolladores la usaran en sus programas.

El sistema de reconocimiento de voz de Windows permite a los usuarios interactuar con sus ordenadores a través de la voz. Fue diseñado para las personas que quieren limitar significativamente el uso del ratón y del teclado, y mantener o aumentar su productividad. Esta herramienta puede dictar documentos y correos electrónicos, usar comandos de voz para iniciar y cambiar entre aplicaciones, controlar el sistema operativo, e incluso rellenar formularios en la Web.

Windows Speech Recognition [16] posee un esquema de licenciamiento privativo, pero accesible de forma gratuita, solo está disponible para ciertas versiones y ediciones de los sistemas operativos Windows.



Siri

Siri [17] es un asistente personal inteligente que ayuda a realizar tareas mediante preguntas. Se puede utilizar para enviar mensajes de voz, programar reuniones, hacer llamadas telefónicas, enviar un tweet, postear en Facebook, etc.


La aplicación utiliza procesamiento del lenguaje natural para responder preguntas, hacer recomendaciones y realizar acciones mediante la delegación de las solicitudes a un conjunto de servicios web. Apple afirma que el software se adapta a las preferencias individuales del usuario con el tiempo y además que personaliza los resultados.

Siri se introdujo originalmente como una aplicación iOS disponible en la App Store por Siri, Inc., que fue adquirida por Apple el 28 de abril de 2010. Siri, Inc. ha anunciado que su software estará disponible para teléfonos BlackBerry y Android. Siri está disponible para el iPhone 5, iPhone 4S, iPad (3^a generación) e iPod touch (5^a generación).



Google Voice search

Google Voice Search es un producto de Google que permite usar el buscador de Google por voz en un teléfono móvil o un ordenador.

Está integrado en el navegador Google Chrome y se usa en el propio buscador, acoplado un micrófono al ordenador y pulsando sobre el icono . También han sacado la aplicación de Google Search para Android que soporta el Google Voice Action a partir de la versión 2.2. Google Voice Action [18] permite enviar mensajes de texto, escuchar música mediante búsquedas por artista o canción, realizar llamadas, utilizar el navegador, etc.





Capítulo 3

Tecnologías utilizadas

"Nunca ande por el camino trazado, porque él sólo conduce a donde ya fueron otros."

Alexander Graham Bell (1847-1922)

3.1 Java

3.1.1 Historia

Surgió en 1991, cuando un grupo de ingenieros de Sun Microsystems trataron de diseñar un nuevo lenguaje de programación destinado a electrodomésticos. La reducida potencia de cálculo y memoria de los electrodomésticos llevó a desarrollar un lenguaje sencillo capaz de generar código de tamaño reducido.

Java [19] fue diseñado para soportar múltiples arquitecturas, debido a que desarrollaron un código independiente del tipo de electrodoméstico, el cual se ejecutaba sobre una máquina virtual denominada Java Virtual Machine (JVM). Era la JVM quien interpretaba el código convirtiéndolo a código particular dependiendo de la CPU utilizada. Esto permitía lo que luego se ha convertido en el principal lema de lenguaje: "Write Once, Run Everywhere". A pesar de los esfuerzos realizados por sus creadores, ninguna empresa de electrodomésticos se interesó por el nuevo lenguaje.

Con la llegada del navegador web Mosaic el equipo reorientó el proyecto hacia internet. Para ello, el código compilado tenía que poder transportarse a través de redes, operar en cualquier cliente, y asegurar al cliente que el código era seguro para ejecutarse.

La popularización de la World Wide Web hizo que estos atributos fueran mucho más interesantes. Los navegadores web permitían a millones de personas navegar por la red y acceder a contenido multimedia de una manera sencilla. Sin embargo los usuarios de internet pronto descubrieron que el formato HTML era demasiado limitado. Las extensiones HTML, tales como formularios, sólo pusieron de relieve las limitaciones, dejando claro que ningún navegador podía incluir todas las características que los usuarios querían.

El navegador de Sun, HotJava™, fue el primero en mostrar las interesantes propiedades que poseía la plataforma y el lenguaje de programación Java para integrar programas dentro de páginas HTML. Los programas eran descargados de forma transparente en el navegador junto con las páginas HTML en las que aparecían. Al igual que las páginas HTML, los programas compilados eran independientes de la red y del host.

En 1994, se hizo una demostración de HotJava y la plataforma Java a los ejecutivos de Sun. Java 1.0 pudo descargarse por primera vez en 1994, pero hubo que esperar al 23 de mayo de 1995, durante las conferencias de SunWorld, a que vieran la luz pública Java y HotJava, el navegador Web. El acontecimiento fue anunciado por John Gage, el

Director Científico de Sun Microsystems. El acto estuvo acompañado por una pequeña sorpresa adicional, el anuncio por parte de Marc Andreessen, Vicepresidente Ejecutivo de Netscape, de que Java sería soportado en sus navegadores. El 9 de enero de 1996, Sun fundó el grupo empresarial JavaSoft para que se encargase del desarrollo tecnológico. Dos semanas más tarde la primera versión de Java fue publicada.

3.1.2 Características

Las principales características pueden resumirse en:

- Java es un lenguaje de programación de propósito general, concurrente y orientado a objetos.
- Su sintaxis es similar a C y C ++, pero omite muchas de las características que hacen que C y C ++ sea complejo, confuso e inseguro.
- Independiente de la plataforma de ejecución.
- Lenguaje compilado e interpretado.
- Lenguaje orientado a internet.
- JVM se implementa en código dependiente de la plataforma.
- Fuertemente tipado.
- Gestión de memoria por la JVM: para ello utiliza el operador new para la creación de objetos y la liberación de recursos es gestionada por el recolector de basura.

3.1.3 Máquina virtual de Java (JVM)

La máquina virtual de Java es la piedra angular de la plataforma Java. Es el componente de la tecnología responsable del hardware y de que sea independiente del sistema operativo, así como del pequeño tamaño de su código compilado, y de proteger a los usuarios de programas maliciosos.

La JVM es una máquina de computación abstracta. Al igual que una máquina de computación real, tiene un conjunto de instrucciones y manipula varias áreas de memoria diferentes en tiempo de ejecución.

La implementación del primer prototipo de la máquina virtual de Java, hecho en Sun Microsystems, Inc., emuló el conjunto de instrucciones de la JVM en un software alojado en un dispositivo de mano. Las implementaciones actuales de Oracle emulan la máquina virtual de Java en dispositivos móviles, en ordenadores personales y en servidores, pero la máquina virtual de Java no asume ninguna tecnología de implementación particular, hardware o sistema operativo del host donde se aloje.

La máquina virtual de Java no entiende el lenguaje de programación Java, sólo de un formato binario especial, el formato de los ficheros “.class”. Un fichero de clase contiene instrucciones de la máquina virtual de Java y una tabla de símbolos, así como otra información complementaria. Por razones de seguridad, la máquina virtual Java impone fuertes restricciones sintácticas y estructurales en el código en un archivo “.class”.

3.1.4 Kit de desarrollo y entorno de ejecución

El Java Development Kit (JDK) es un entorno de desarrollo para aplicaciones Java. JDK no es un entorno integrado de desarrollo en el sentido de aglutinar todas las herramientas bajo la misma interfaz gráfica si no que se ejecutan en modo de comando. Los comandos principales son:

- javac: compila el código y genera los archivos class.
- java: intérprete de la JVM. Ejecuta los bytecodes de los ficheros class.
- javadoc: genera documentación
- AppletViewer: visualiza páginas HTML.
- jdb: depura aplicaciones Java.

El Java Runtime Environment (JRE) es el software necesario para ejecutar cualquier aplicación desarrollada para la plataforma Java, y está incluido dentro del JDK. Puede también obtenerse como un paquete independiente, y puede considerarse como el entorno necesario para ejecutar una aplicación Java, mientras que un desarrollador debe además contar con otras facilidades que ofrece el JDK.

3.1.5 Versiones

A continuación especificaremos el historial de versiones de Java.

JDK 1.0	Nombre clave: OAK Fecha de lanzamiento: 23 de enero de 1996 Características: <ul style="list-style-type: none">• Nació como herramienta para los sistemas operativos Solaris, Windows, Mac Os y Linux.• La primera versión estable fue la JDK 1.0.2.
JDK 1.1	Nombre clave: -- Fecha de lanzamiento: 19 de febrero de 1997 Mejoras: <ul style="list-style-type: none">• Reestructuración intensiva del modelo de eventos AWT.• Clases internas.



	<ul style="list-style-type: none">• JavaBeans.• JDBC.• RMI.
J2SE 1.2	<p>Nombre clave: Playground</p> <p>Fecha de lanzamiento: 8 de diciembre de 1998</p> <p>Nomenclatura:</p> <ul style="list-style-type: none">• Esta y las siguientes versiones fueron recogidas bajo la denominación Java 2 y el nombre J2SE, reemplazó a JDK para distinguir la plataforma base de J2EE y J2ME. <p>Mejoras:</p> <ul style="list-style-type: none">• Reflexión en la programación.• La palabra reservada strictfp.• La API gráfica fue integrada en las clases básicas.• La máquina virtual fue equipada con un compilador JIT por primera vez.• Java Plug-in.• Java IDL, una implementación de IDL para la interoperabilidad con CORBA• Colecciones.
J2SE 1.3	<p>Nombre clave: Kestrel</p> <p>Fecha de lanzamiento: 8 de mayo de 2000</p> <p>Mejoras:</p> <ul style="list-style-type: none">• Inclusión de la máquina virtual de HotSpot JVM.• RMI fue cambiado para que se basara en CORBA.• JavaSound• Inclusión del JNDI en el paquete de librerías principales.• JPDA
J2SE 1.4	<p>Nombre clave: Merlin</p> <p>Fecha de lanzamiento: 6 de febrero de 2002</p> <p>Mejoras:</p> <ul style="list-style-type: none">• Palabra reservada assert.• Expresiones regulares modeladas al estilo de las expresiones regulares Perl.• Encadenación de excepciones Permite a una excepción encapsular la excepción de bajo nivel original.• Non-blocking.• Logging API.• API para la lectura y escritura de imágenes en formatos como JPEG o PNG.



	<ul style="list-style-type: none">• Parser XML integrado y procesador XSLT.• Seguridad integrada y extensiones criptográficas.• Java Web Start.
Java SE 6	<p>Nombre clave: Mustang</p> <p>Fecha de lanzamiento: 11 de diciembre de 2006</p> <p>Nomenclatura:</p> <ul style="list-style-type: none">• Sun cambió el nombre J2SE por Java SE y eliminó el ".0" del número de versión <p>Mejoras:</p> <ul style="list-style-type: none">• Incluye un nuevo marco de trabajo y APIs que hacen posible la combinación de Java con lenguajes dinámicos como PHP, Python, Ruby y JavaScript.• Incluye el motor Rhino, de Mozilla, una implementación de Javascript en Java.• Incluye un cliente completo de Servicios Web y soporta las últimas especificaciones para Servicios Web.• Mejoras en la interfaz gráfica y en el rendimiento.
Java SE 7	<p>Nombre clave: Dolphin</p> <p>Fecha de lanzamiento: julio de 2011</p> <p>Mejoras:</p> <ul style="list-style-type: none">• Soporte para XML dentro del propio lenguaje.• Un nuevo concepto de super paquete.• Soporte para closures.• Introducción de anotaciones estándar para detectar fallos en el software.• Java Module System.• Java Kernel.• Nueva API para el manejo de Días y Fechas, la cual reemplazara las antiguas clases Date y Calendar.• Posibilidad de operar con clases BigDecimal usando operadores.

Los componentes de Java han ido cambiando a lo largo de su historia, pero en la versión actual, la Java SE 7, estos son los por los que contiene:

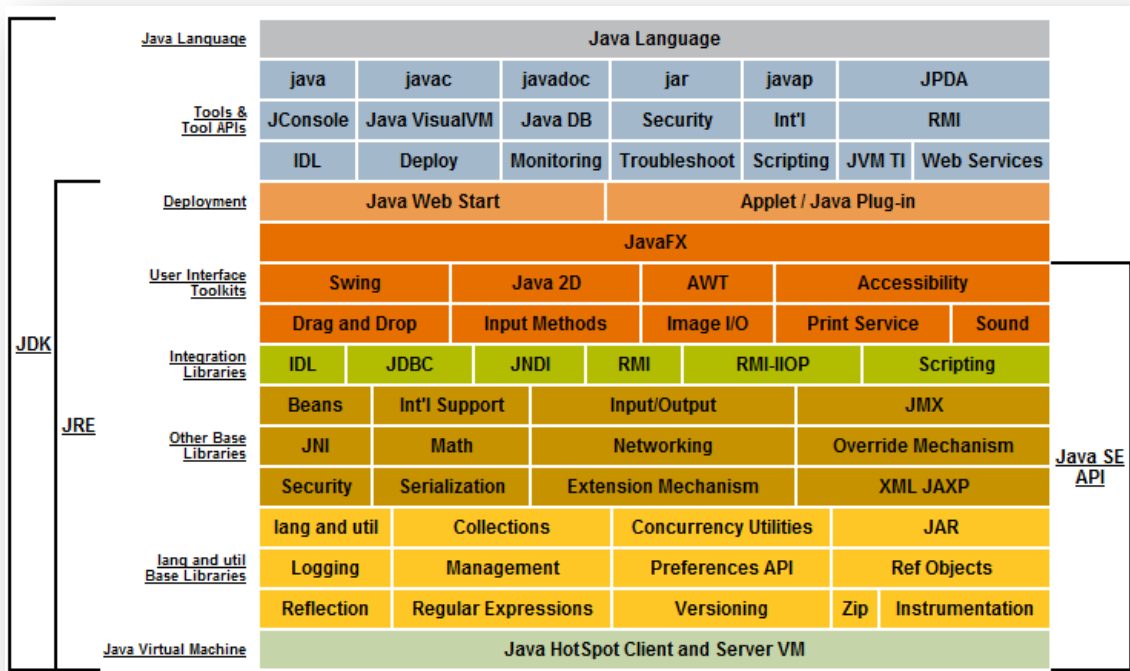


Ilustración 1. Componentes de Java SE 7

3.1.6 Distribuciones de Java

Existen cuatro plataformas del lenguaje de programación Java:

- Java Platform, Standard Edition (Java SE)
- Java Platform, Enterprise Edition (Java EE)
- Java Platform, Micro Edition (Java ME)
- JavaFX

Cada plataforma Java proporciona una máquina virtual y una interfaz de programación de aplicaciones, esto permite que las aplicaciones escritas en una plataforma puedan ejecutarse en cualquier sistema compatible con todas las ventajas del lenguaje de programación Java: independencia de la plataforma, potencia, estabilidad, facilidad de desarrollo y seguridad.

Java SE

Es la base para las otras distribuciones Java. Su principal componente es el API, el cual proporciona la funcionalidad básica del lenguaje de programación Java. Define todo lo necesario, desde los tipos básicos hasta clases que se utilizan para la creación de redes, seguridad, acceso a base de datos, interfaz gráfica de usuario (GUI) desarrollo y análisis XML.



Además de la API, la plataforma Java SE dispone de una máquina virtual, herramientas de desarrollo, tecnologías de implementación, y otras bibliotecas de clases y herramientas usadas en aplicaciones de tecnología Java.

Java EE

La plataforma Java EE está construida sobre la plataforma Java SE. La plataforma Java EE proporciona una API y un entorno de ejecución para desarrollar y ejecutar aplicaciones de red a gran escala, a varios niveles y escalables.

Java ME

La plataforma Java ME proporciona un API y una máquina virtual de pequeñas dimensiones para la ejecución de aplicaciones en dispositivos pequeños como los móviles. Este API es un subconjunto de la API de Java SE, junto con las bibliotecas especiales de clases útiles para el desarrollo de aplicación en dispositivos pequeños. Las aplicaciones JavaME a menudo son clientes de servicios de la plataforma Java EE.

Java FX

JavaFX es una plataforma para la creación de aplicaciones dinámicas de Internet usando una interfaz de usuario ligera. Las aplicaciones de JavaFX utilizan gráficos de aceleración hardware y motores de comunicación para aprovechar los clientes de mayor rendimiento y además aporta un moderno look-and-feel y APIs de alto nivel para la conexión de fuentes de datos en red. Aplicaciones JavaFX pueden ser clientes de servicios de la plataforma Java EE.

3.2 J2EE y Servlet

3.2.1 J2EE

Historia

A principio de la década de los 90s, los sistemas tradicionales de información de las empresas empezaron a responder a las necesidades del cliente pasando de las aplicaciones construidas en dos capas, aplicaciones implementadas con el modelo cliente-servidor, a modelos de aplicación más flexible de tres niveles o de varios niveles. Los nuevos modelos separaban la lógica de negocio de los servicios de sistema y de la interfaz de usuario, colocándolo en una capa intermedia entre los dos. La evolución de los nuevos servicios de middleware, le dio un impulso adicional a esta nueva arquitectura. Al mismo tiempo, el creciente uso de internet e intranets en las aplicaciones empresariales contribuyó a enfatizar en aplicaciones que fuesen más fáciles de desplegar en los clientes y que fuesen más ligeras.

El diseño multi-capa simplifica el desarrollo, despliegue y mantenimiento de las aplicaciones empresariales. Permite a los desarrolladores centrarse en los detalles de la programación de la lógica de negocio, contando con servicios de backend para proporcionar la infraestructura y aplicaciones cliente (tanto independientes como dentro de los navegadores web) para proporcionar la interacción del usuario. Una vez desarrollado, la lógica de negocio puede ser desplegada en los servidores apropiados de una organización. Sin embargo, a pesar de estas claras ventajas, el modelo limita la capacidad de los desarrolladores para construir aplicaciones a partir de componentes estandarizados, para desplegar una aplicación en diferentes plataformas, o para escalar fácilmente aplicaciones que satisfagan las cambiantes condiciones del negocio.

Dentro de Sun Microsystems, hubo varios enfoques de desarrollo que marcaron el camino hacia lo que se convertiría en la tecnología J2EE [20]. En primer lugar, la tecnología Java Servlets mostró que los desarrolladores querían crear CGI que pudieran ejecutarse en cualquier servidor web que soportase la plataforma Java. En segundo lugar, la tecnología JDBC proporcionaba un modelo que une el lema de Java “Write Once, Run Anywhere” con los sistemas de gestión de base de datos existentes. Finalmente, el éxito de la arquitectura de componentes JavaBeans demostró la utilidad de encapsular el comportamiento dentro de componentes fácilmente configurables y reutilizables. La convergencia de estos tres conceptos: comportamiento del lado del servidor escrito en Java, conectores para permitir el acceso a sistemas empresariales existentes y componentes modulares y fáciles de desplegar; fue lo que condujo al estándar J2EE.

Características

La plataforma Java EE está diseñada para ayudar a los desarrolladores a crear aplicaciones escalables, con varias capas, distribuidas, seguras y fiables. Un nombre común para estas aplicaciones es el de "aplicaciones empresariales", llamado así porque estas aplicaciones están diseñadas para resolver los problemas encontrados por las grandes empresas. Sin embargo, estas aplicaciones no sólo son útiles para las grandes corporaciones, agencias y gobiernos, ya que los beneficios de estas aplicaciones son útiles, incluso esenciales, para los desarrolladores individuales y pequeñas organizaciones.

Las características que hacen que las aplicaciones empresariales sean importantes, como la seguridad y la fiabilidad, a menudo las convierten en complejas. Java EE está diseñado para reducir la complejidad de la implementación de aplicaciones, proporcionando un entorno de desarrollo, una API, y un entorno de ejecución, permitiendo a los desarrolladores concentrarse únicamente en la funcionalidad.

Arquitectura en capas

En una aplicación multicapa, la funcionalidad de la aplicación está separada mediante capas aisladas. Por lo general, este tipo de aplicaciones tienen una capa cliente, una capa intermedia y una capa de dato. El nivel de cliente se compone de un programa cliente que realiza peticiones a la capa intermedia. Las funciones del nivel intermedio de negocios son manejar peticiones de clientes y datos de la aplicación, y su almacenamiento en una base de datos permanente en el nivel de datos.

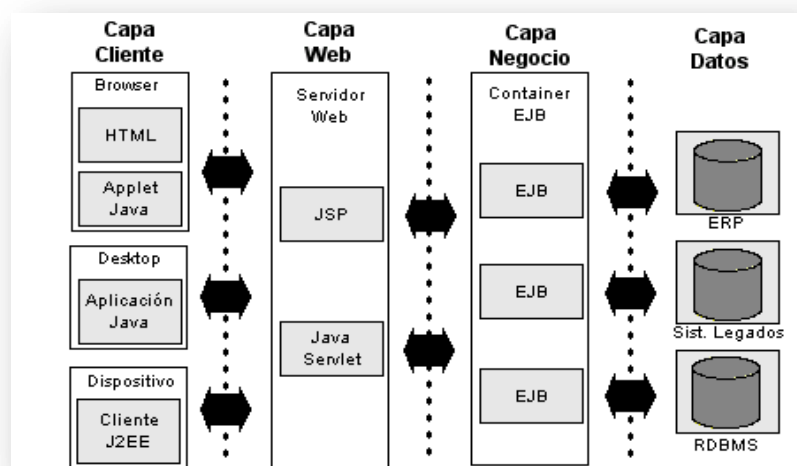


Ilustración 2. Arquitectura J2EE

El desarrollo de aplicaciones Java EE se concentra en el nivel intermedio para que la gestión de aplicaciones empresariales sea más fácil, más robusta y más segura.

Capa Cliente

Este nivel se compone de aplicaciones clientes que acceden a un servidor Java EE y que se encuentran normalmente en una máquina distinta al servidor. Los clientes hacen peticiones al servidor, y el servidor procesa las peticiones y da una respuesta al cliente. Existen varios tipos de aplicaciones que pueden ser clientes Java EE, y no son siempre, aplicaciones Java. Los clientes pueden ser un navegador web, una aplicación independiente, u otros servidores, y pueden ejecutarse en un equipo diferente del servidor Java EE.

Capa Web

Se encuentra en el servidor web y contiene la lógica de presentación que se utiliza para generar una respuesta al cliente. Recibe los datos del usuario desde la capa cliente y basado en éstos genera una respuesta apropiada a la solicitud. J2EE utiliza en esta capa las componentes Java Servlets y JavaServer Pages para crear los datos que se enviarán al cliente. En la siguiente tabla se muestra un resumen de las tecnologías de esta capa.

Tecnología	Descripción
Servlets	Clases del lenguaje de programación Java que procesan solicitudes de manera dinámica y construye respuestas que normalmente van dirigidas a páginas HTML.
Tecnología JavaServer Faces	Componente del sistema dedicado a las interfaces de usuario para aplicaciones web, que te permite incluir elementos en una página (por ejemplo, campos y botones), convertir y validar datos de la interfaz, guardar información de componentes en los almacenes de datos del servidor, y mantener el estado del componente.
Lenguaje de expresiones	Un conjunto de etiquetas estándar que se utilizan en las páginas JSP para referirse a los componentes de Java EE.
JSP	Documentos de texto que se compilan en los servlets y definen cómo el contenido dinámico puede ser añadido a páginas estáticas, como las páginas HTML.
Librería estándar de etiquetas JSP	Biblioteca de etiquetas que encapsula la funcionalidad básica común a las páginas JSP.
Componentes JavaBeans	Objetos que actúan como almacenes de datos temporales para las páginas de una aplicación.

Tabla 3. Tecnologías Java EE de la capa web

Capa Negocio

Se encuentra en el servidor de aplicaciones y contiene el núcleo de la lógica del negocio de la aplicación. Provee las interfaces necesarias para utilizar el servicio de

componentes del negocio. Las componentes del negocio interactúan con la capa de datos y son típicamente implementadas como componentes EJB. Las siguientes tecnologías se utilizan en la capa de negocio de aplicaciones Java EE:

Tecnología	Descripción
Componentes de negocio JavaBeans	Son componentes que encapsulan la funcionalidad principal de la aplicación.
Servicios web JAX-RS RESTful	API para crear servicios web que responden a los métodos HTTP (GET o POST). Los servicios web JAX-RS son desarrollados de acuerdo a los principios de REST, o transferencia de estado representacional.
Servicios web endpoints JAX-WS	API para crear y consumir servicios web SOAP.
Entidades de API Java Persistence	API para acceder a la información de almacenes de datos subyacentes y mapear esos datos a objetos Java.
Beans gestionados de Java EE	Componentes que proporcionan la lógica de negocio de una aplicación, pero no requieren de las características transaccionales o de seguridad de un beans de negocio.

Tabla 4. Tecnologías Java EE de la capa de negocio

Capa de Sistemas de Información Empresariales

Este nivel consta de servidores de bases de datos, sistemas de planificación de recursos empresariales, y otras fuentes de datos existentes, como los mainframes. Estos recursos se encuentran típicamente en un equipo distinto al del servidor Java EE, y son accedidos por los componentes de la capa de negocio. Las tecnologías que se utilizan en esta capa son:

Tecnología	Descripción
JDBC	API de bajo nivel para el acceso y recuperación de información de los almacenes de datos. Un uso común de JDBC es hacer consultas SQL en una base de datos particular.
API de persistencia de Java	API para acceder a la información de los almacenes de datos y mapear esos datos a objetos de Java. Este API es de alto nivel, en comparación con la API de JDBC, y oculta la complejidad de JDBC para el usuario.
Conector de la arquitectura de Java EE	Un API para conectarse a otros recursos empresariales, como a la planificación de recursos o al software de gestión de clientes del sistema.
API de transacciones de Java	API para definir y gestionar las transacciones, incluyendo las transacciones distribuidas o las transacciones que cruzan múltiples fuentes de datos.



Tabla 5. Tecnologías Java EE de la capa de sistemas de información

Servidores Java EE

Un servidor Java EE es un servidor de aplicaciones que implementa el API de la plataforma Java EE y proporciona los servicios estándar de J2EE. Los servidores J2EE son normalmente llamados servidores de aplicaciones, porque permiten servir datos de aplicación a los clientes, al igual que los servidores web servir páginas web a los navegadores web.

Los servidores albergan varios tipos de componentes de aplicación, que se corresponden con las capas de la arquitectura multicapa de Java EE. Además, proporcionan servicios a estos componentes mediante un contenedor.

Contenedores de Java EE

Los contenedores de Java EE son la interfaz entre el componente y la funcionalidad de bajo nivel proporcionado por la plataforma para soportar ese componente. La funcionalidad del contenedor está definida por la plataforma, y es diferente para cada tipo de componente. Sin embargo, el servidor permite que los diferentes tipos de componentes trabajen juntos para proporcionar la funcionalidad de una aplicación de empresa.

Contenedor Web

El contenedor web es la interfaz entre los componentes web y el servidor web. Un componente web puede ser un servlet, o una página JSP. El contenedor gestiona el ciclo de vida de los componentes, despacha las peticiones a los componentes de la aplicación, y proporciona interfaces para los datos del contexto, así como la información sobre la solicitud actual.

Contenedor de Aplicaciones Cliente

El contenedor de aplicaciones cliente es la interfaz entre las aplicaciones cliente de Java EE, que son aplicaciones especiales de Java SE que utilizan componentes del servidor y el propio servidor Java EE. El contenedor de la aplicación cliente se ejecuta en la máquina cliente, y es la puerta de enlace entre las aplicaciones cliente y los componentes del servidor Java EE que utiliza el cliente.

Contenedor EJB

El contenedor EJB es la interfaz entre los beans de empresas, los cuales proporcionan la lógica de negocio en una aplicación, y el servidor Java EE. El contenedor EJB se ejecuta en un servidor y controla la ejecución de los beans de empresa de las aplicaciones.

3.2.2 Servlet

Historia

La palabra servlet deriva de otra anterior, applet, que se refería a pequeños programas que se ejecutan en el contexto de un navegador web. Por contraposición, un servlet es un programa que se ejecuta en un servidor.

La especificación original de Servlets fue creada por Sun Microsystems (la versión 1.0 fue terminada en junio de 1997). Comenzando con la versión 2.3, la especificación de Servlet fue desarrollada siguiendo el Proceso de la Comunidad Java (Java Community Process).

Descripción

Un servlet es una clase del lenguaje de programación Java utilizado para ampliar las capacidades de los servidores que alojan aplicaciones accesibles a través de un modelo de programación de petición-respuesta. Aunque los servlets pueden responder a cualquier tipo de solicitud, se utilizan comúnmente para extender las aplicaciones alojadas en servidores web.

Los paquetes `javax.servlet` y `javax.servlet.http` proporcionar interfaces y clases para escribir servlets. Todos los servlets deben implementar la interfaz `Servlet`, ya que define los métodos del ciclo de vida de un servlet. La clase `HttpServlet` proporciona métodos, tales como `doGet` y `doPost`, para el manejo de servicios específicos de HTTP.

Los servlets a menudo tienen el mismo propósito que los programas implementados utilizando CGI, pero los servlets ofrecen varias ventajas frente a CGI:

- El rendimiento es mucho mejor.
- Los servlets se ejecutan dentro del espacio de direcciones de un servidor Web. No es necesario para crear un proceso independiente para controlar cada solicitud de cliente.
- Los servlets son independientes de la plataforma, ya que está escrito en Java.
- El controlador de la seguridad en Java impone una serie de restricciones para proteger los recursos en un equipo servidor.
- La funcionalidad completa de las bibliotecas de clases de Java está disponible para un servlet. Se puede comunicar con los applets, bases de datos u otros programas a través de mecanismos RMI.

Ciclo de vida

El ciclo de vida de los servlets es manejado por el contenedor de servlets y contiene los pasos siguientes:

- Paso 1: Carga de la clase Servlet.
- Paso 2: Creación de la instancia del servlet.
- Paso 3: Llamada al método `init()` del servlet.
- Paso 4: Llamada al método `service()` del servlet.
- Paso 5: Llamada al método `destroy()` del servlet.

Los pasos 1, 2 y 3 se ejecutan sólo una vez, cuando el servlet se carga inicialmente. Por defecto, el servlet no se carga hasta que se recibe la primera solicitud. Sin embargo, se puede forzar el contenedor para cargar el servlet cuando el contenedor se ejecute.

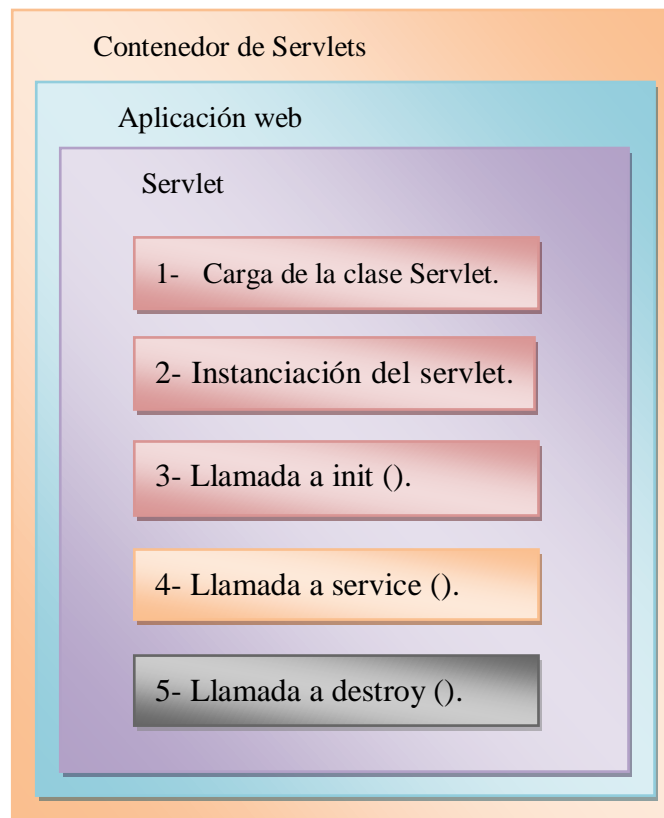


Ilustración 3. Ciclo de vida de un servlet

Cargar la clase Servlet

Antes de que un servlet pueda invocar el contenedor de servlets primero debe cargar su definición de clase.

Crear la instancia del Servlet

Cuando la clase servlet está cargada, el contenedor de servlets crea una instancia del servlet. Normalmente, sólo se crea una única instancia, y todas las solicitudes simultáneas se ejecutan en la misma instancia del servlet.

Llamada al método init del Servlet

Cuando la instancia está creada, su método `init()` se invoca. Este método permite que el servlet se inicialice antes de la que primera solicitud sea procesada. Se pueden obtener los parámetros iniciales con “`getInitParameter`” de “`ServletConfig`”.

Llamada al método service del Servlet

El contenedor de servlets llama al método `service ()` para manejar las solicitudes procedentes del cliente (navegador) y escribir la respuesta al cliente. Cada vez que el servidor recibe una petición para un servlet, el servidor genera un nuevo hilo y una llamada a `service`. El método `service ()` comprueba el tipo de petición HTTP (GET, POST, PUT, DELETE, etc) y llama a los métodos `doGet`, `doPost`, `doPut`, `doDelete`, etc., según corresponda. Los métodos que más comúnmente son usados es el `doGet()` y el `doPost()`. La diferencia entre ambos métodos es que en `doPost` los parámetros se pasan en la cabecera del mensaje, mientras que en `doGet` los parámetros están en la URL, y son visibles para todo el mundo.

Llamada al método destroy del Servlet

Si ya no necesita el servlet para procesar las peticiones, el contenedor de servlets llama al método `destroy ()`. Al igual que el método `init ()`, este método también se llama una sola vez en todo el ciclo de vida del servlet. Al llamar al método `destroy ()` se indica al contenedor de servlets que no envíe ninguna solicitud a `service` y que el servlet libere todos los recursos asociados. La máquina virtual de Java reclama la memoria asociada a los recursos para la recolección de basura.

Características

Las características más importantes de los Servlets de Java se pueden resumir en:

- Manejo de Sesiones: Se puede hacer seguimiento de usuarios a través de distintos servlets a través de la creación de sesiones.
- Utilización de Cookies: Las cookies son pequeños datos en texto plano que pueden ser guardados en el cliente. La API de servlets permite un manejo fácil y limpio de ellas.
- Multi-thread: Los servlets soportan el acceso concurrente de los clientes, aunque hay que tener especial cuidado con las variables compartidas a menos que se utilice la interfaz `SingleThreadModel`.
- Programación en Java: Se obtienen las características de multiplataforma o acceso a APIs como JDBC, RMI, etc.



Versiones

A continuación especificaremos el historial de versiones de Servlets.

Servlet 3.0	Fecha de lanzamiento: diciembre 2009. Plataforma: JavaEE 6, JavaSE 6. Cambios importantes: <ul style="list-style-type: none">• Facilidad de desarrollo.• Servlets asíncronos.• Carga de archivos.
Servlet 2.5	Fecha de lanzamiento: septiembre 2005. Plataforma: JavaEE 5, JavaSE 5. Cambios importantes: <ul style="list-style-type: none">• Requiere JavaSE 5.• Soporta anotación.
Servlet 2.4	Fecha de lanzamiento: noviembre 2003. Plataforma: J2EE 1.4, J2SE 1.3. Cambios importantes: <ul style="list-style-type: none">• El archivo “web.xml” usa el esquema XML.
Servlet 2.3	Fecha de lanzamiento: agosto 2001. Plataforma: J2EE 1.3, J2SE 1.2. Cambios importantes: <ul style="list-style-type: none">• Adición del filtro.
Servlet 2.2	Fecha de lanzamiento: agosto 1999. Plataforma: J2EE 1.2. Cambios importantes: <ul style="list-style-type: none">• Se convierte en parte de J2EE.• Introducción de aplicaciones web en ficheros “.war”.
Servlet 2.1	Fecha de lanzamiento: noviembre 1998. Cambios importantes: <ul style="list-style-type: none">• Dispatcher.• ServletContext.
Servlet 2.0	Plataforma: JDK 1.1. Cambios importantes: <ul style="list-style-type: none">• Parte del Kit de Desarrollo Java Servlet 2.0.
Servlet 1.0	Fecha de lanzamiento: junio 1997.

3.3 Apache Tomcat

3.3.1 Introducción

Apache Tomcat [21], también conocido como simplemente Tomcat o Jakarta Tomcat, es un servidor web multiplataforma que funciona como contenedor de servlets y que se desarrolla bajo el proyecto denominado Jackarta perteneciente a la Apache Software Foundation. Tomcat implementa las especificaciones de los servlets y de JavaServer Pages o JSP de Sun Microsystem. El servidor es mantenido y desarrollado por miembros de la fundación y por voluntarios independientes, que tienen libre acceso al código fuente bajo los términos establecidos por la Apache Software Foundation.

3.3.2 Historia

Apache Tomcat comenzó siendo una implementación de servlets iniciada por James Duncan Davidson, que trabajaba como arquitecto software en Sun Microsystem y que posteriormente ayudó al proyecto de código abierto. Duncan, inicialmente pensó que el proyecto se convertiría en software de código abierto y además quiso ponerle un nombre de animal, en este caso Tomcat (gato) ya que, de algún modo pretendía trasladar la posibilidad de cuidarse por sí mismo, es decir, de ser independiente.

Las primeras distribuciones de Apache Tomcat fueron las 3.0.x aunque las versiones estables más recientes son las 6.0.30 y la 7.0.6 que implementan las especificaciones de Servlet 2.5 y JSP 2.1.

3.3.3 Estructura

La jerarquía de directorios de Apache Tomcat es la siguiente:

- bin: arranque, cierre, scripts y ejecutables.
- common: clases comunes que puede utilizar Catalina (contenedor de servlets) y las aplicaciones web.
- conf: ficheros XML y la correspondiente DTD para la configuración de Apache Tomcat.
- logs: logs del contenedor de servlets y de las aplicaciones.
- server: clases usadas por el contenedor de servlets.
- shared: clases compartidas por todas las aplicaciones web.
- webapps: directorio que contiene las aplicaciones web.
- work: almacenamiento temporal de ficheros y directorios.

3.3.4 Componentes

A partir de la versión 4.x Tomcat fue lanzado con el contenedor de servlets "Catalina", con el contenedor HTTP "Coyote" y un motor para JSP llamado "Jasper". Las principales características de estos tres componentes se especifican a continuación:

Catalina

Este componente implementa las especificaciones de servlets y JSP. Para Apache Tomcat el elemento principal es una base de datos de nombres de usuarios, password y roles a estos usuarios permitiendo a Catalina integrarse donde ya existe información de autenticación como describe la especificación de servlets.

Coyote

Componente conector que admite el protocolo HTTP 1.1 para el servidor web. Escucha en un puerto TCP especificado por el servidor y envía la solicitud al motor Tomcat para que éste procese la solicitud y envíe una respuesta al cliente.

Jasper

Jasper analiza archivos JSP para compilar el código Java y, si se producen cambios, éste los vuelve a compilar. Desde la versión 5.x se usa Jasper 2 que es JavaServer para webs 2.0. Jasper 2 ha introducido las siguientes novedades:

- Recompilación al producirse un cambio.
- Incorpora el compilador JDT de Java.
- Puesta en común de etiquetas JSP.

3.3.5 Características y evolución

Como se ha dicho anteriormente, Apache Tomcat es un servidor web que da soporte a servlets y JSPs de modo que, no es un servidor de aplicaciones. Dado que dicho producto fue desarrollado en Java, éste puede ejecutarse sobre cualquier sistema operativo, previa instalación de la máquina virtual de Java aunque, también se puede usar con MAMPP (Mac OS X), LAMPP (GNU/Linux), WAMPP (Windows) o XAMPP (cualquier sistema operativo). Además, puede funcionar como servidor web por sí mismo. Sin embargo, en sus inicios se pensaba que dicho servidor era recomendable usarse en entornos de desarrollo con requisitos mínimos de velocidad. En la actualidad no existe esta percepción y por esto, es usado como servidor web autónomo en entornos con alto nivel de tráfico y alta disponibilidad.



La evolución del producto y las principales características de cada una de las versiones son:

Tomcat 3.x	Es la distribución inicial. Implementación a partir de Servlet 2.2 y JSP 1.1 Recarga de servlets. Funciones básicas de HTTP.
Tomcat 4.x	Implementado a partir de las especificaciones Servlet 2.3 y JSP 1.2 Contenedor de servlets rediseñado como Catalina. Motor JSP rediseñado con Jasper. Conector Coyote. Java Management Extensions (JMX), JSP y administración basada en Struts.
Tomcat 5.x	Implementado a partir de las especificaciones Servlet 2.4 y JSP 2.0 Recolección de basura reducida. Capa envolvente nativa para Windows y Unix para la integración de las plataformas. Análisis rápido JSP.
Tomcat 6.x	Implementado a partir de Servlet 2.5 y JSP 2.1. Soporte para Unified Expression Language 2.1. Diseñado para funcionar en Java SE 5.0 y posteriores. Soporte para Comet a través de la interfaz CometProcessor.
Tomcat 7.x	Implementado a partir de Servlet 3.0 y JSP 2.2. Mejoras para detectar y prevenir "fugas de memoria" en las aplicaciones web. Limpieza interna de código. Soporte para la inclusión de contenidos externos directamente en una aplicación web.

Las principales características que soporta la última versión estable son:

- Autenticación de acceso básico.
- Negociación de credenciales.
- HTTPS.
- Alojamiento compartido.
- CGI o interfaz de entrada común.
- Servlets de Java.
- SSI.
- Consola de administrador.

3.4 Mysql

3.4.1 Introducción

El software MySQL® [22] proporciona un servidor de base de datos SQL (Structured Query Language) muy rápido, multi-threaded, multi usuario y robusto. El servidor MySQL está diseñado para entornos de producción críticos, con alta carga de trabajo así como para integrarse en software para ser distribuido. MySQL es una marca registrada de MySQL AB

MySQL, el sistema de gestión de bases de datos SQL Open Source más popular, lo desarrolla, distribuye y soporta MySQL AB. MySQL AB es una compañía comercial, fundada por los desarrolladores de MySQL. Es una compañía Open Source de segunda generación que une los valores y metodología Open Source con un exitoso modelo de negocio.

3.4.2 Historia

El origen de MySQL se remonta a la década de los ochenta. Michael Widenius, también conocido como Monty, un joven programador que realizaba complejas aplicaciones en el lenguaje BASIC, al no encontrar un sistema de almacenamiento bueno, pensó en construir el suyo propio.

Años después, en 1995, y en colaboración con David Axmark, Widenius desarrolló un producto que básicamente era el resultado de sus investigaciones, más dos aportaciones nuevas: el uso del lenguaje SQL y la accesibilidad a través de Internet. Así nació MySQL y también la empresa MySQL AB.

Widenius comenzó a usar mSQL para conectar a tablas utilizando sus propias rutinas rápidas de bajo nivel (ISAM). Sin embargo y tras algunas pruebas, llegó a la conclusión que mSQL no era lo suficientemente rápido o flexible para sus necesidades. Esto provocó la creación de una nueva interfaz SQL para nuestra base de datos pero casi con la misma interfaz API que mSQL. Esta API fue diseñada para permitir código de terceras partes que fue escrito para poder usarse con mSQL para ser fácilmente portado para el uso con MySQL.

MySQL surgió alrededor de la década del 90, Michael Windenis comenzó a usar mSQL para conectar tablas usando sus propias rutinas de bajo nivel. Tras unas primeras pruebas, llegó a la conclusión de que mSQL no era lo bastante flexible ni rápido para lo



que necesitaba, por lo que tuvo que desarrollar nuevas funciones. Esto resultó en una interfaz SQL a su base de datos, totalmente compatible a mSQL.

El origen del nombre MySQL no se sabe con certeza de donde proviene, por un lado se dice que en sus librerías han llevado el prefijo “my” durante los diez últimos años, por otra parte, la hija de uno de los desarrolladores se llama My. Así que no está claramente definido cuál de estas dos causas han dado lugar al nombre de este conocido gestor de bases de datos.

El nombre del delfín de MySQL es "Sakila", que fue elegido por los fundadores de MySQL AB de una gran lista de nombres sugerida por los usuarios en el concurso "Name the Dolphin" (ponle nombre al delfín). El nombre ganador fue enviado por Ambrose Twebaze, un desarrollador de software Open Source de Swaziland, África. Según Ambrose, el nombre femenino de Sakila tiene sus raíces en SiSwate, el idioma local de Swaziland. Sakila también es el nombre de una ciudad en Arusha, Tanzania, cerca del país de origen de Ambrose, Uganda.

3.4.3 Características

Inicialmente, MySQL carecía de algunos elementos esenciales en las bases de datos relacionales, tales como integridad referencial y transacciones. A pesar de esto, atrajo a los desarrolladores de páginas web con contenido dinámico, debido a su simplicidad, de tal manera que los elementos faltantes fueron complementados por la vía de las aplicaciones que la utilizan. Poco a poco estos elementos faltantes, están siendo incorporados tanto por desarrolladores internos, como por desarrolladores de software libre.

En las últimas versiones se pueden destacar las siguientes características principales:

- El principal objetivo de MySQL es la velocidad y la robustez.
- Soporta gran cantidad de tipos de datos para las columnas.
- Gran portabilidad entre sistemas, puede trabajar en distintas plataformas y sistemas operativos.
- Cada base de datos cuenta con 3 archivos: Uno de estructura, uno de datos y uno de índice y soporta hasta 32 índices por tabla.
- Aprovecha la potencia de sistemas multiproceso, gracias a su implementación multihilo.
- Flexible sistema de contraseñas (passwords) y gestión de usuarios, con un muy buen nivel de seguridad en los datos.
- El servidor soporta mensajes de error en distintas lenguas.

3.5 SoX

3.5.1 Introducción

SoX [23] es una utilidad que funciona por línea de comando y es multiplataforma (Windows, Linux, MacOS X). SoX lee y escribe archivos de audio en los formatos más populares y, opcionalmente, puede aplicar efectos en ellos. Puede combinar varias fuentes de entrada de audio, sintetizar y, en muchos sistemas, actuar como un reproductor de audio de propósito general o un grabador de audio multipista. También tiene capacidad para dividir la entrada en varios archivos de salida.

Toda la funcionalidad de SoX está disponible utilizando el comando “sox”. Para simplificar la reproducción y grabación de audio, si SoX es invocado como reproductor, el archivo de salida se ajusta automáticamente para ser el dispositivo de sonido por defecto, y si se invoca como grabador, el dispositivo de sonido predeterminado se utiliza como fuente de entrada. Además, SoX posee comandos para consultar información de la cabecera del archivo del archivo de audio.

SoX es una herramienta de procesamiento de audio mediante línea de comandos, especialmente adecuado para hacer rápida y sencilla la edición y el procesamiento batch.

3.5.2 Historia

SoX fue creado en julio de 1991 por Lance Norskog y enviado al grupo Usenet como “Aural eXchange: Sound sample translator”. Con la segunda versión (en noviembre de ese mismo año) fue rebautizado como Sound Exchange. Norskog siguió manteniendo y lanzando SoX via Usenet, FTP, y luego en la web hasta principios de 1995, cuando SoX estaba en su versión 11 (gamma). En mayo de 1996, Chris Bagwell comenzó a mantener y publicar las versiones actualizadas de SoX, comenzando con la versión sox-11gamma-cb. En septiembre de 2000, Bagwell registró el proyecto en SourceForge con el nombre de "sox". El registro fue anunciado el 4 de septiembre de 2000 y SoX 12.17 salió el 7 de septiembre de 2000.

A lo largo de su historia, SoX ha tenido muchos colaboradores, Guido van Rossum, más conocido como el creador del lenguaje de programación Python, fue un factor importante en los primeros días de SoX.



3.5.3 Características

Las principales características de SoX son:

- Conversión de archivos de audio.
- Edición de archivos de audio: concatenación, borrado, repetición, volumen, normalización, etc.
- Cambio de los atributos de audio.
- Adición de efectos de audio.
- Multiplataforma.
- Posee multitud de opciones de manipulación de sonido.
- Lectura y escritura de una gran variedad de formatos de audio como: AU, WAV, AIFF, Ogg Vorbis, FLAC, RAW, etc.
- Grabación de audio.
- Reproducción de audio, incluso mediante URL.
- Procesamiento mediante retardos, cambios de fase, compresión, filtros de paso y paso alto, etc.
- Análisis estadísticos.
- Fusión de archivos.





Capítulo 4

Análisis y diseño

"Nada os pertenece en propiedad más que vuestros sueños."

Friedrich Nietzsche (1844-1900)

4.1 Introducción

El objetivo de esta sección es explicar la metodología de desarrollo seguida a lo largo del proyecto y la especificación de la funcionalidad a través de los casos de uso y los requisitos. Además, se comentará el diseño del sistema mediante su arquitectura y su diagrama de clases.

4.2 Metodología de desarrollo

Durante la creación de este proyecto se ha seguido como metodología de desarrollo software una versión adaptada de Métrica 3. Los procesos de la estructura principal de Métrica Versión 3 son los siguientes:

- Planificación de Sistemas de Información.
- Desarrollo de Sistemas de Información.
- Mantenimiento de Sistemas de Información.

Métrica v.3 ha sido concebida para abarcar el desarrollo completo de Sistemas de Información sea cual sea su complejidad y magnitud, por lo cual su estructura responde a desarrollos máximos. La ventaja de esta metodología es que puede adaptarse y dimensionarse en cada momento de acuerdo a las características particulares de un proyecto. Es por ello que para el desarrollo de nuestro sistema nos hemos centrado únicamente en los dos primeros procesos, descartando la fase de mantenimiento de Sistemas de Información porque entendemos esta fase como un proceso a realizar en el futuro.

4.3 Casos de Uso

En este apartado se expondrán los casos de uso de nuestro sistema. Un caso de uso representa un uso típico del sistema que permite identificar y definir los requisitos que una aplicación debe cumplir así como la interacción que existe entre el usuario y el sistema. Es una técnica muy utilizada para mostrar la futura capacidad del sistema de una forma simple e intuitiva.

4.3.1 Casos de uso de Gestión de Usuarios

En la siguiente ilustración se muestran los casos de usos de la aplicación relacionados con la gestión de usuarios y sesiones.

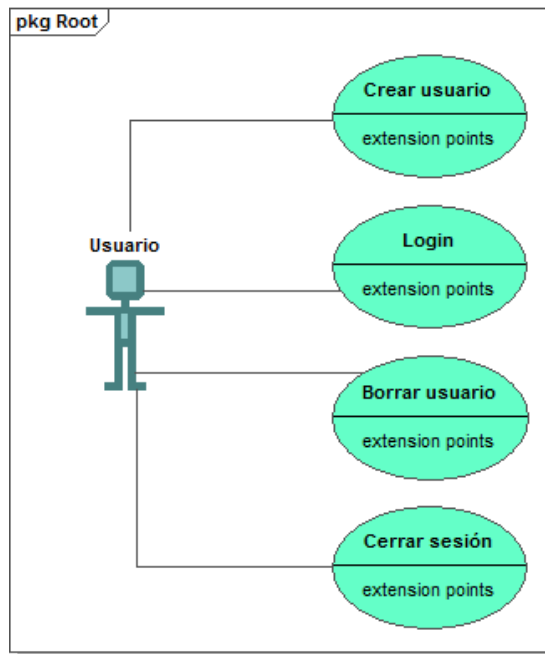


Ilustración 4. Casos de uso de Gestión de Usuarios

Identificador	CU-01
Nombre	Crear usuario
Actores	Usuario sin registrar
Objetivos	El usuario crea un perfil y se guarda la información del usuario.
Precondiciones	El usuario debe acceder a la aplicación y pulsar el botón de “Nuevo Usuario”.
Flujo Normal	<ol style="list-style-type: none"> 1. El usuario introduce el nombre de usuario y la contraseña (dos veces) y pulsa el botón “Siguiente”. 2. Se envían los datos al servidor junto con un SPL y un tiempo por defecto. 3. Se comprueba que no exista un usuario con ese nombre de usuario. 4. Se almacena el nuevo usuario en la base de datos del servidor. 5. Se envía una respuesta afirmativa al dispositivo móvil. 6. Se guarda en la base de datos del móvil. 7. Se muestra un mensaje por pantalla informando al usuario que su perfil ha sido creado correctamente.
Flujo Alternativo	<ol style="list-style-type: none"> 1. El usuario pulsa el botón “Atrás” y se muestra la pantalla de login.
Flujo Alternativo	<ol style="list-style-type: none"> 4. Se comprueba que no exista un usuario con ese nombre de usuario, pero previamente ya ha sido creado. 5. Se envía una respuesta de error al dispositivo móvil.

	6. Se muestra un mensaje por pantalla informando al usuario que su perfil ha sido creado correctamente.
Postcondiciones	El usuario ha sido registrado en el sistema. Se crea la jerarquía de carpetas de usuario en el dispositivo móvil.

Tabla 6. CU-01

Identificador	CU-02
Nombre	Login
Actores	Usuario registrado y sin registrar
Objetivos	Identificar al usuario y permitir el acceso al sistema.
Precondiciones	El usuario debe acceder al sistema.
Flujo Normal	<ol style="list-style-type: none"> 1. Se muestra al usuario una pantalla para que introduzca su usuario y contraseña. 2. El usuario escribe los datos y pulsa el botón “Entrar”. 3. Se envían los datos al servidor. 4. Se comprueba que exista un usuario con ese nombre de usuario y contraseña. 5. Se envía una respuesta afirmativa al dispositivo móvil. 6. Se muestra un mensaje por pantalla informando al usuario que el login es correcto.
Flujo Alternativo	<ol style="list-style-type: none"> 3. No se pueden enviar los datos al servidor porque se produce un fallo en la conexión. 4. Se muestra un mensaje por pantalla informando al usuario que no es posible realizar la conexión.
Flujo Alternativo	<ol style="list-style-type: none"> 7. Se comprueba el usuario y contraseña y no coincide. 8. Se envía una respuesta de error al dispositivo móvil. 9. Se muestra un mensaje por pantalla informando al usuario que no existe ese usuario y contraseña en el sistema.
Postcondiciones	Se crea la jerarquía de carpetas de usuario en el dispositivo móvil, si no existe. Se accede a la pantalla principal.

Tabla 7. CU-02

Identificador	CU-03
Nombre	Borrar usuario.
Actores	Usuario registrado.
Objetivos	Eliminar el usuario del sistema.
Precondiciones	El usuario debe haber accedido a la pantalla de configuración.
Flujo Normal	<ol style="list-style-type: none"> 1. Se muestra un mensaje por pantalla preguntando al usuario la confirmación de la acción. 2. El usuario pulsa el botón “Sí”. 3. Se envía al servidor el nombre de usuario 4. El servidor busca el nombre de usuario en la base de datos y elimina su registro y su jerarquía de carpetas. 5. El servidor envía una respuesta afirmativa. 6. El dispositivo móvil busca el nombre de usuario en la base de datos y elimina su registro y su jerarquía de carpetas.
Flujo Alternativo	<ol style="list-style-type: none"> 2. El usuario pulsa el botón “No”. 3. Se muestra la pantalla de configuración.

Postcondiciones	Se muestra la pantalla de login.
------------------------	----------------------------------

Tabla 8. CU-03

Identificador	CU-04
Nombre	Cerrar sesión.
Actores	Usuario registrado.
Objetivos	Cerrar sesión de usuario y salir del sistema.
Precondiciones	El usuario debe acceder al Menú de la pantalla principal y pulsar sobre el botón “Cerrar sesión”.
Flujo Normal	<ol style="list-style-type: none"> 1. Se muestra un mensaje por pantalla preguntando al usuario la confirmación de la acción. 2. El usuario pulsa el botón “Sí”.
Flujo Alternativo	<ol style="list-style-type: none"> 2. El usuario pulsa el botón “No”. 3. Se muestra la pantalla principal.
Postcondiciones	Se muestra la pantalla de Login.

Tabla 9. CU-04

4.3.2 Casos de uso de Gestión de Configuración de los valores de grabación

En la ilustración de más abajo se muestran los casos de usos relacionados con la gestión de los valores de SPL y tiempo, utilizados en la grabación de somniloquias.

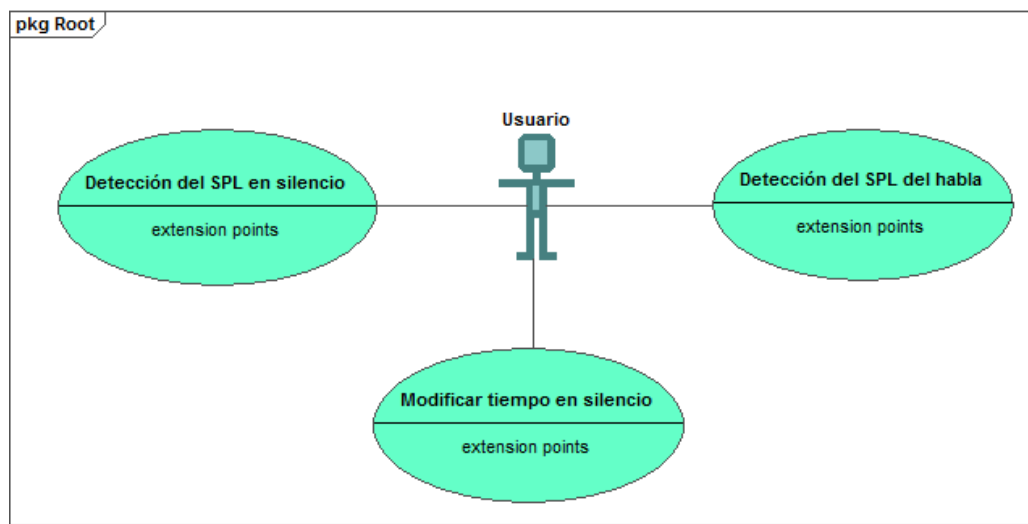


Ilustración 5. Casos de uso de configuración de valores de grabación

Identificador	CU-05
Nombre	Detección del SPL en silencio
Actores	Usuario registrado
Objetivos	Calcular el valor del SPL cuando no hay ruido ambiental
Precondiciones	El usuario debe pulsar el botón “OK” tras haber creado su perfil o

	acceder a la pantalla de configuración.
Flujo Normal	<ol style="list-style-type: none"> 1. Se muestra al usuario la pantalla de detección en silencio con un texto informativo. 2. El usuario pulsa el botón de detección. 3. El usuario permanece callado para que el sistema pueda realizar los cálculos del SPL en silencio. 4. El usuario pulsa el botón “→”.
Flujo Alternativo	<ol style="list-style-type: none"> 4. El usuario pulsa de nuevo el botón de detección. 5. El usuario permanece callado para que el sistema pueda realizar los cálculos del SPL en silencio. 6. El usuario pulsa el botón “→”.
Postcondiciones	Se guarda el valor del SPL internamente para usos posteriores.

Tabla 10. CU-05

Identificador	CU-06
Nombre	Detección del SPL del habla
Actores	Usuario registrado
Objetivos	Calcular el valor del SPL cuando el usuario está hablando.
Precondiciones	El usuario debe pulsar el botón “→” en la pantalla de detección del SPL en silencio.
Flujo Normal	<ol style="list-style-type: none"> 1. Se muestra al usuario la pantalla de detección del habla con un texto informativo. 2. El usuario pulsa el botón de detección. 3. El usuario habla durante un tiempo determinado para que el sistema pueda realizar los cálculos del SPL con su voz. Simultáneamente el sistema graba un fichero de audio para el posterior análisis de la emoción neutra del usuario. 4. El sistema envía el SPL calculado al servidor junto con el nombre de usuario. 5. El servidor busca el registro del usuario y lo actualiza. 6. El servidor envía una respuesta afirmativa. 7. El sistema envía el fichero de audio al servidor junto con el nombre de usuario. 8. El servidor realiza un análisis del archivo, obtiene sus valores prosódicos y los almacena. 9. El servidor envía una respuesta afirmativa. 10. El usuario pulsa el botón “→”.
Flujo Alternativo	10. El usuario pulsa de nuevo el botón de detección y se realiza el flujo normal de nuevo.
Postcondiciones	<p>El SPL se almacena en la base de datos del servidor y del dispositivo móvil.</p> <p>Se muestra la pantalla principal al usuario.</p>

Tabla 11. CU-06

Identificador	CU-07
Nombre	Modificar tiempo en silencio
Actores	Usuario registrado.
Objetivos	Modificar el valor del tiempo de silencio entre archivos.
Precondiciones	El usuario debe haber accedido a la pantalla de configuración.

Flujo Normal	<ol style="list-style-type: none"> 1. El usuario escribe el tiempo en el cuadro de texto correspondiente. 2. El usuario pulsa el botón de guardar. 3. El sistema comprueba el formato del tiempo introducido. 4. Si el formato es correcto establece una conexión con el servidor y le envía el valor junto con el nombre de usuario 5. El servidor busca el registro del usuario y lo actualiza. 6. El servidor envía una respuesta afirmativa.
Flujo Alternativo	<ol style="list-style-type: none"> 3. Si el formato es incorrecto se muestra un mensaje por pantalla informando al usuario de que el formato no es correcto. 4. El usuario pulsa el botón OK. 5. Se muestra la pantalla de configuración.
Postcondiciones	El SPL se almacena en la base de datos del servidor y del dispositivo móvil.

Tabla 12. CU-07

4.3.3 Casos de uso de la funcionalidad principal

En la siguiente ilustración se muestran los casos de usos de la aplicación relacionados con la principal funcionalidad del sistema.

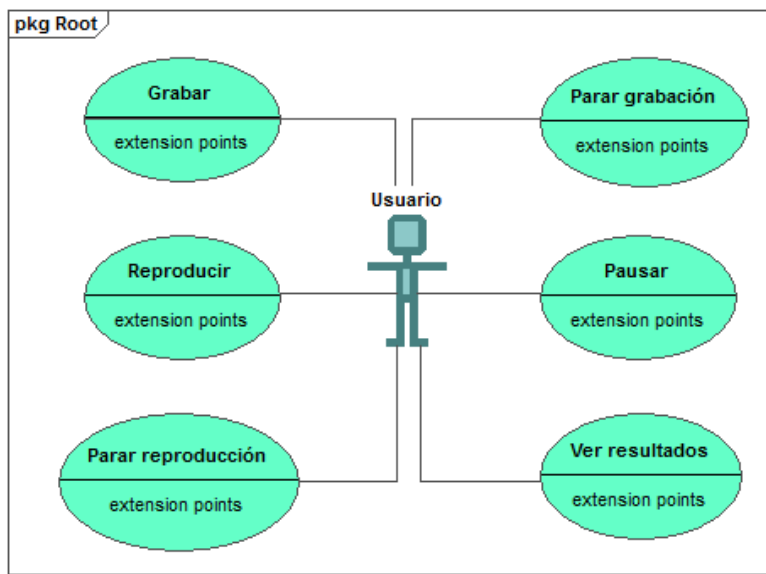


Ilustración 6. Casos de uso de la funcionalidad principal

Identificador	CU-08
Nombre	Grabar
Actores	Usuario registrado.
Objetivos	Grabar las somniloquias del usuario.
Precondiciones	El usuario debe acceder a la pestaña de grabación de la pantalla principal y pulsar el botón de grabación.
Flujo Normal	<ol style="list-style-type: none"> 1. El sistema inicia el proceso de grabación de somniloquias. 2. La aplicación crear archivos de audio tras detectar sonidos que superen el umbral de SPL del usuario.

	<ol style="list-style-type: none"> 3. Para cada somniloquia: <ol style="list-style-type: none"> a. Se establece una comunicación el servidor y se envía el fichero. b. El servidor lo procesa y elimina los ruidos c. El servidor envía el fichero post-procesado al dispositivo móvil. d. El dispositivo móvil realiza la petición para la obtención de resultados. e. El servidor realiza un análisis acústico y una transcripción a texto. f. El servidor envía los resultados al dispositivo móvil.
Postcondiciones	<p>Se almacenan los archivos en sus correspondientes rutas del servidor y del dispositivo móvil.</p> <p>Se desactiva el botón de reproducción.</p>

Tabla 13. CU-08

Identificador	CU-09
Nombre	Parar grabación.
Actores	Usuario registrado.
Objetivos	Parar la grabación de las somniloquias del usuario.
Precondiciones	El sistema debe haber iniciado el proceso de grabación.
Flujo Normal	<ol style="list-style-type: none"> 1. El sistema termina la grabación de las somniloquias. 2. El sistema almacena el archivo de grabación en curso. 3. Se comprueba si hay somniloquias sin enviar y para cada una de ellas: <ol style="list-style-type: none"> a. Se establece una comunicación el servidor y se envía el fichero. b. El servidor lo procesa y elimina los ruidos c. El servidor envía el fichero post-procesado al dispositivo móvil. d. El dispositivo móvil realiza la petición para la obtención de resultados. e. El servidor realiza un análisis acústico y una transcripción a texto. f. El servidor envía los resultados al dispositivo móvil.
Postcondiciones	<p>Finaliza el proceso de grabación.</p> <p>Se almacenan los archivos en sus correspondientes rutas del servidor y del dispositivo móvil.</p> <p>Se activa el botón de reproducción.</p>

Tabla 14. CU-09

Identificador	CU-10
Nombre	Reproducir
Actores	Usuario registrado.
Objetivos	Reproducir un archivo de audio propio.
Precondiciones	El usuario debe haber accedido a la pestaña de grabación de la pantalla principal y haber pulsado el botón de reproducción.
Flujo Normal	<ol style="list-style-type: none"> 1. El sistema muestra al usuario un listado de sus carpetas.

	<ol style="list-style-type: none"> 2. El usuario accede a una de las carpetas y selecciona un archivo de audio. 3. El sistema reproduce el archivo permitiendo la pausa del mismo.
Postcondiciones	<p>La reproducción del archivo finaliza al terminar de leer el fichero de audio.</p> <p>El sistema activa el botón de pausa y desactiva el botón de grabación.</p>

Tabla 15. CU-10

Identificador	CU-11
Nombre	Pausar
Actores	Usuario registrado.
Objetivos	Pausar un archivo de audio propio.
Precondiciones	El sistema debe estar reproduciendo un archivo de audio.
Flujo Normal	<ol style="list-style-type: none"> 1. El usuario pulsa el botón de pausar. 2. La reproducción se detiene.
Postcondiciones	El sistema activa el botón de reproducción.

Tabla 16. CU-11

Identificador	CU-12
Nombre	Parar reproducción
Actores	Usuario registrado.
Objetivos	Pausar un archivo de audio propio.
Precondiciones	El sistema debe estar reproduciendo un archivo de audio.
Flujo Normal	<ol style="list-style-type: none"> 1. El usuario pulsa el botón de pausar. 2. La reproducción se detiene.
Postcondiciones	El sistema activa el botón de reproducción y activa el botón de grabación.

Tabla 17. CU-12

Identificador	CU-13
Nombre	Ver resultados
Actores	Usuario registrado.
Objetivos	Mostrar al usuario los resultados calculados para un archivo de audio: frecuencia fundamental, transcripción y análisis de emociones.
Precondiciones	El usuario debe haber accedido a la pestaña de resultados de la pantalla principal.
Flujo Normal	<ol style="list-style-type: none"> 1. Si se trata del primer acceso a esta pantalla, el sistema muestra al usuario un listado de sus archivos de resultados. 2. El usuario selecciona un archivo. 3. Se muestra una pantalla con los resultados del archivo y un botón "Cambiar" para cambiar de archivo.
Flujo Alternativo	<ol style="list-style-type: none"> 1. Si no se trata del primer acceso a esta pantalla, el sistema muestra al usuario los resultados de la selección anterior. 2. El usuario pulsa el botón "Cambiar". 3. El sistema muestra al usuario un listado de sus archivos de resultados. 4. El usuario selecciona un archivo.

	5. Se muestra una pantalla con los resultados del archivo y un botón “Cambiar” para cambiar de archivo.
Postcondiciones	Se muestran al usuario los resultados obtenidos de una somniloquia.

Tabla 18. CU-13

4.3.4 Casos de uso navegación

La siguiente ilustración refleja un esquema con los casos de usos del sistema relacionados a la navegación que el usuario puede realizar en la aplicación.

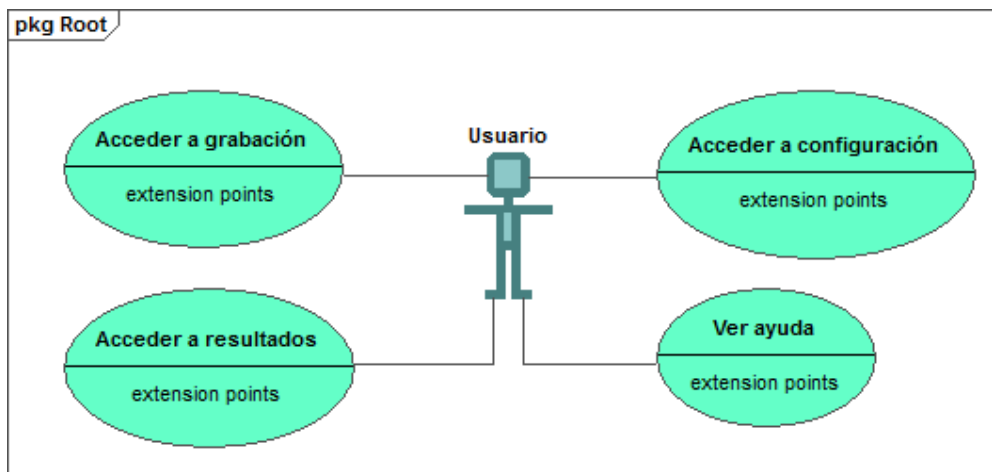


Ilustración 7. Casos de uso de navegación

Identificador	CU-14
Nombre	Acceder a grabación
Actores	Usuario registrado.
Objetivos	Acceder a la pantalla de grabación dentro de la pantalla principal.
Precondiciones	El usuario debe acceder a la pantalla principal.
Flujo Normal	1. El usuario pulsa la pestaña de grabación.
Postcondiciones	Se muestra la pantalla de grabación.

Tabla 19. CU-14

Identificador	CU-15
Nombre	Acceder a configuración
Actores	Usuario registrado.
Objetivos	Acceder a la pantalla de configuración dentro de la pantalla principal.
Precondiciones	El usuario debe acceder a la pantalla principal.
Flujo Normal	1. El usuario pulsa la pestaña de configuración.
Postcondiciones	Se muestra la pantalla de configuración.

Tabla 20. CU-15

Identificador	CU-16
Nombre	Acceder a resultados
Actores	Usuario registrado.
Objetivos	Acceder a la pantalla de resultados dentro de la pantalla principal.
Precondiciones	El usuario debe acceder a la pantalla principal.
Flujo Normal	1. El usuario pulsa la pestaña de resultados.
Postcondiciones	Se muestra la pantalla de resultados.

Tabla 21. CU-16

Identificador	CU-17
Nombre	Ver ayuda.
Actores	Usuario registrado.
Objetivos	Mostrar una pantalla de ayuda al usuario.
Precondiciones	El usuario debe acceder al Menú de la pantalla principal.
Flujo Normal	1. El usuario pulsa sobre el botón “Ayuda”.
Postcondiciones	El sistema muestra la pantalla de Ayuda con un texto informativo.

Tabla 22. CU-17

4.4 Requisitos

4.4.1 Requisitos de Usuario

La plantilla que se utilizará para especificar los requisitos de usuario consta de los siguientes campos que se pasan a describir a continuación:

- **Identificador:** Representa unívocamente a cada requisito. La nomenclatura utilizada será:
 - RUC-AXX: para los requisitos de usuario de capacidad Android.
 - RUC-WXX para los requisitos de usuario de capacidad del servicio web.
 - RUR-AXX para los requisitos de usuario de restricción Android.
 - RUR-WXX para los requisitos de usuario de restricción del servicio web.

XX será un número desde el 01 hasta el 99.

- **Nombre:** Nombre representativo del requisito.
- **Descripción:** Fragmento en el que se explica brevemente en qué consiste el requisito.
- **Fuente:** Hace referencia a la procedencia del requisito.
- **Prioridad:** Grado de urgencia con el que debe cumplirse el requisito. Puede ser Alta, Media o Baja.



Requisitos de Usuario de Capacidad

Android

Identificador	RUC-A01
Nombre	<u>Crear un nuevo usuario</u>
Descripción	Dar de alta en el sistema a un nuevo usuario.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 23. RUC-A01

Identificador	RUC-A02
Nombre	<u>Eliminar a un usuario registrado</u>
Descripción	Eliminar a un usuario del sistema, así como toda la información asociada a su cuenta de usuario.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 24. RUC-A02

Identificador	RUC-A03
Nombre	<u>Modificar usuario registrado</u>
Descripción	Un usuario podrá modificar algunos valores de su cuenta.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 25. RUC-A03

Identificador	RUC-A04
Nombre	<u>Iniciar sesión</u>
Descripción	Iniciar sesión por parte de los usuarios registrados para hacer uso de las funcionalidades del sistema.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 26. RUC-A04

Identificador	RUC-A05
Nombre	<u>Cerrar sesión</u>
Descripción	Cerrar sesión cuando el usuario registrado desee abandonar la aplicación.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 27. RUC-A05



Identificador	RUC-A06
Nombre	<u>Grabar</u>
Descripción	El usuario podrá grabar somniloquias.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 28. RUC-A06

Identificador	RUC-A07
Nombre	<u>Parar grabación</u>
Descripción	El usuario podrá parar la grabación.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 29. RUC-A07

Identificador	RUC-A08
Nombre	<u>Reproducir</u>
Descripción	El usuario podrá reproducir una grabación.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 30. RUC-A08

Identificador	RUC-A09
Nombre	<u>Pausar</u>
Descripción	El usuario podrá pausar la reproducción de un audio.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 31. RUC-A09

Identificador	RUC-A10
Nombre	<u>Parar reproducción</u>
Descripción	El usuario podrá parar la reproducción de un audio.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 32. RUC-A10

Identificador	RUC-A11
Nombre	<u>Acceder grabación</u>
Descripción	El usuario podrá acceder a las grabaciones realizadas
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 33. RUC-A11



Identificador	RUC-A12
Nombre	<u>Navegación</u>
Descripción	El usuario podrá navegar entre las pantallas y pestañas del programa.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 34. RUC-A12

Identificador	RUC-A13
Nombre	<u>Envío de archivos</u>
Descripción	La aplicación móvil podrá enviar los archivos de audio generados a la plataforma web.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 35. RUC-A13

Identificador	RUC-A14
Nombre	<u>Creación de jerarquías</u>
Descripción	La aplicación móvil podrá crear la jerarquía de directorios para cada usuario.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 36. RUC-A14

Identificador	RUC-A15
Nombre	<u>Comunicación</u>
Descripción	El dispositivo móvil podrá comunicarse con la plataforma web.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 37. RUC-A15

Identificador	RUC-A16
Nombre	<u>Almacenamiento de archivos</u>
Descripción	La aplicación móvil almacenará los archivos asociados a cada usuario.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 38. RUC-A16

Identificador	RUC-A17
Nombre	<u>Obtención del nivel de presión sonora de forma automática</u>
Descripción	El sistema calculará el nivel de presión sonora de un usuario.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 39. RUC-A17



Identificador	RUC-A18
Nombre	<u>Ver resultados</u>
Descripción	El usuario podrá visualizar los resultados que devuelve el servidor.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 40. RUC-A18

Identificador	RUC-A19
Nombre	<u>Ayuda</u>
Descripción	La aplicación dispondrá de un tutorial de ayuda para el usuario.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 41. RUC-A19

Identificador	RUC-A20
Nombre	<u>Generación de archivos de audios</u>
Descripción	Se generará un nuevo archivo de audio tras cumplirse un tiempo de espera sin detectar sonidos en una misma grabación.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 42. RUC-A20

Identificador	RUC-A21
Nombre	<u>Recibir archivos</u>
Descripción	La aplicación móvil podrá recibir archivos enviados desde el servidor.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 43. RUC-A21

Identificador	RUC-A22
Nombre	<u>Modificar SPL</u>
Descripción	El usuario podrá configurar tantas veces como quiera el valor del SPL a través de la detección automática.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 44. RUC-A22

Identificador	RUC-A23
Nombre	<u>Modificar tiempo de silencio entre grabaciones</u>
Descripción	El usuario podrá modificar el tiempo de silencio entre grabaciones.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 45. RUC-A24



Servicio Web

Identificador	RUC-W01
Nombre	<u>Insertar a un nuevo usuario</u>
Descripción	Insertar en el servicio web a un nuevo usuario.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 46. RUC-W01

Identificador	RUC-W02
Nombre	<u>Eliminar a un usuario registrado</u>
Descripción	Eliminar a un usuario del sistema, así como toda la información asociada a su cuenta de usuario.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 47. RUC-W02

Identificador	RUC-W03
Nombre	<u>Modificar usuario registrado</u>
Descripción	Modificar los datos de un usuario.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 48. RUC-W03

Identificador	RUC-W04
Nombre	<u>Almacenar archivos</u>
Descripción	La plataforma web podrá almacenar archivos recibidos o generados asociándolos a cada usuario.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 49. RUC-W04

Identificador	RUC-W05
Nombre	<u>Transcripción</u>
Descripción	La plataforma web será capaz de transcribir los archivos de audio a texto.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 50. RUC-W05



Identificador	RUC-W06
Nombre	<u>Análisis de emociones</u>
Descripción	El sistema se encargará de realizar un análisis emocional de los archivos de audio.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 51. RUC-W06

Identificador	RUC-W07
Nombre	<u>Comunicación</u>
Descripción	La plataforma web podrá comunicarse con el dispositivo móvil.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 52. RUC-W07

Identificador	RUC-W08
Nombre	<u>Procesar peticiones</u>
Descripción	El sistema se encargará de procesar las peticiones recibidas.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 53. RUC-W08

Identificador	RUC-W09
Nombre	<u>Enviar ficheros de audio</u>
Descripción	El servicio web podrá enviar audios al dispositivo móvil.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 54. RUC-W09

Identificador	RUC-W10
Nombre	<u>Enviar ficheros de texto</u>
Descripción	El servicio web podrá enviar archivos de texto al dispositivo móvil.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 55. RUC-W10

Identificador	RUC-W11
Nombre	<u>Crear jerarquía</u>
Descripción	El servicio web podrá crear jerarquías de directorios asociadas a un usuario para almacenar archivos.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 56. RUC-W11



Identificador	RUC-W12
Nombre	<u>Eliminar ruidos</u>
Descripción	El servicio web podrá eliminar los ruidos y ronquidos de las somniloquias.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 57. RUC-W12

Identificador	RUC-W13
Nombre	<u>Actualizar SPL</u>
Descripción	El servlet podrá actualizar el valor del SPL.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 58. RUC-W13

Identificador	RUC-W14
Nombre	<u>Actualizar tiempo de silencio entre grabaciones</u>
Descripción	El servlet podrá actualizar el valor del tiempo de silencio entre grabaciones.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 59. RUC-W14

Identificador	RUC-W15
Nombre	<u>Obtener características fundamentales de un audio</u>
Descripción	El servlet podrá realizar análisis acústico y obtener las características fundamentales de un audio.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 60. RUC-W15

Identificador	RUC-W16
Nombre	<u>Insertar características fundamental de un archivo neutro</u>
Descripción	El servlet podrá almacenar las características fundamentales de un archivo neutro.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 61. RUC-W16



Requisitos de Usuario de Restricción

Android

Identificador	RUR-A01
Nombre	<u>Crear una interfaz de usuario sencilla e intuitiva.</u>
Descripción	Crear una interfaz de usuario lo más sencilla e intuitiva posible.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 62. RUR-A01

Identificador	RUR-A02
Nombre	<u>Campos de login</u>
Descripción	Requerir un nombre de usuario único y una contraseña para iniciar sesión en el sistema.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 63. RUR-A02

Identificador	RUR-A03
Nombre	<u>Ocultación de contraseñas</u>
Descripción	La aplicación no mostrará el texto de la contraseña de un usuario por motivos de seguridad.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 64. RUR-A03

Identificador	RUR-A04
Nombre	<u>Enviar contraseñas cifradas</u>
Descripción	Por motivos de seguridad la contraseña se enviará cifrada al servidor.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 65. RUR-A04

Identificador	RUR-A05
Nombre	<u>Grabación inteligente.</u>
Descripción	La aplicación graba de forma inteligente, almacenando sólo aquellos fragmentos en los que el usuario hable.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 66. RUR-A05



Identificador	RUR-A06
Nombre	<u>Campos de crear usuario</u>
Descripción	Se pedirá un nombre de usuario único y dos veces una contraseña para crear un usuario en el sistema.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 67. RUR-A06

Identificador	RUR-A07
Nombre	<u>Valor por defecto el tiempo de espera entre audios</u>
Descripción	El tiempo que la aplicación espera en los momentos de silencio tendrá un valor por defecto de 5 minutos.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 68. RUR-A07

Identificador	RUR-A08
Nombre	<u>Rango de valores del tiempo de espera entre audios</u>
Descripción	El valor mínimo del tiempo en silencio entre audios será 0 segundos y el valor máximo será 60 minutos.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 69. RUR-A08

Identificador	RUR-A09
Nombre	<u>Dispositivo móvil</u>
Descripción	La aplicación se podrá ejecutar en dispositivos móviles con características similares al Samsung Galaxy S (I9000).
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 70. RUR-A09

Identificador	RUR-A10
Nombre	<u>Base de datos</u>
Descripción	La aplicación dispondrá de una base de datos (SQLite) para el almacenamiento de información.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 71. RUR-A10

Identificador	RUR-A11
Nombre	<u>Sistema operativo</u>
Descripción	La aplicación será compatible con el sistema operativo para móviles Android.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 72. RUR-A11

Identificador	RUR-A12
Nombre	<u>Versión de Android</u>
Descripción	La aplicación se podrá ejecutar en cualquier dispositivo móvil con versión de la plataforma Android 2.3.3 (API 10) y superiores.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 73. RUR-A12

Identificador	RUR-A13
Nombre	<u>Protocolo de comunicación</u>
Descripción	Las comunicaciones con el servidor se harán mediante el protocolo HTTP.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 74. RUR-A13

Identificador	RUR-A14
Nombre	<u>Formato de archivos de somniloquias originales</u>
Descripción	Los archivos de audio tendrán el siguiente formato: usuario-dd_mm_yy-hh_mm_ss, seguido de la extensión “.wav”.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 75. RUR-A14

Identificador	RUR-A15
Nombre	<u>Formato de archivos de somniloquias procesadas</u>
Descripción	Los archivos de audio tendrán el siguiente formato: usuario-dd_mm_yy-hh_mm_ss-SR, seguido de la extensión “.wav”.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 76. RUR-A15



Identificador	RUR-A16
Nombre	<u>Formato de archivos neutros</u>
Descripción	Los archivos de audio tendrán el siguiente formato: Neutro-usuario-dd_mm_yy-hh_mm_ss, seguido de la extensión “.wav”.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 77. RUR-A16

Identificador	RUR-A17
Nombre	<u>Formato de archivos de resultados</u>
Descripción	Los archivos de texto con los resultados tendrán el siguiente formato: usuario-dd_mm_yy-hh_mm_ss-R, seguido de la extensión “.txt”.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 78. RUR-A17

Identificador	RUR-A18
Nombre	<u>Formato de jerarquía</u>
Descripción	El formato .de la jerarquía contiene una carpeta por cada usuario y dentro los archivos de audio y texto organizados por tipo.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 79. RUR-A18

Identificador	RUR-A19
Nombre	<u>Idioma</u>
Descripción	La aplicación mostrará los textos en español.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 80. RUR-A19

Identificador	RUR-A20
Nombre	<u>Configuración inicial del valor del nivel de presión sonora</u>
Descripción	Una de las fases de la creación de usuario es la detección del valor del SPL.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 81. RUR-A20



Servicio Web

Identificador	RUR-W01
Nombre	<u>Emociones primarias</u>
Descripción	Las emociones que se clasificarán serán primarias, distinguiendo entre: alegría, enfado, tristeza y miedo.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 82. RUR-W01

Identificador	RUR-W02
Nombre	<u>Base de datos plataforma web</u>
Descripción	Se dispondrá de una base de datos MySQL.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 83. RUR-W02

Identificador	RUR-W03
Nombre	<u>Servidor</u>
Descripción	Se dispondrá de un servidor web Apache Tomcat.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 84. RUR-W03

Identificador	RUR-W04
Nombre	<u>Formato de archivos de somniloquias originales</u>
Descripción	Los archivos de audio tendrán el siguiente formato: usuario-dd_mm_yy-hh_mm_ss, seguido de la extensión “.wav”.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 85. RUR-W04

Identificador	RUR-W05
Nombre	<u>Formato de archivos de somniloquias procesadas</u>
Descripción	Los archivos de audio tendrán el siguiente formato: usuario-dd_mm_yy-hh_mm_ss-SR, seguido de la extensión “.wav”.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 86. RUR-W05



Identificador	RUR-W06
Nombre	<u>Formato de archivos neutros</u>
Descripción	Los archivos de audio tendrán el siguiente formato: Neutro-usuario-dd_mm_yy-hh_mm_ss-SR, seguido de la extensión “.wav”.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 87. RUR-W06

Identificador	RUR-W07
Nombre	<u>Formato de archivos de resultados</u>
Descripción	Los archivos de texto con los resultados tendrán el siguiente formato: usuario-dd_mm_yy-hh_mm_ss-R, seguido de la extensión “.txt”.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 88. RUR-W07

Identificador	RUR-W08
Nombre	<u>Formato de jerarquía de directorios</u>
Descripción	La jerarquía estará compuesta por cuatro carpetas principales y dentro de ellas se organizan los archivos por usuario.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 89. RUR-W08

Identificador	RUR-W09
Nombre	<u>Transcripción</u>
Descripción	La transcripción se realizará mediante peticiones a un servidor externo que procese el audio y devuelva un resultado. Este resultado se parseará para obtener únicamente el texto que interese.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 90. RUR-W09

Identificador	RUR-W10
Nombre	<u>Resultado del análisis emocional</u>
Descripción	El resultado del análisis de emociones de un archivo se almacenará en un archivo de texto.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 91. RUR-W10

Identificador	RUR-W11
Nombre	<u>Eliminar fragmentos de grabación</u>
Descripción	La aplicación eliminará fragmentos con ruidos como llantos o ronquidos mediante análisis acústico de las somniloquias.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 92. RUR-W11

Identificador	RUR-W12
Nombre	<u>Tipo de información</u>
Descripción	Cada tipo de información que se almacene en el sistema tendrá un formato adecuado.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 93. RUR-W12

Identificador	RUR-W13
Nombre	<u>Comunicación</u>
Descripción	La comunicación con el dispositivo móvil se hará por HTTP.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 94. RUR-W13

Identificador	RUR-W14
Nombre	<u>Frecuencias de la voz</u>
Descripción	El filtrado de audios se hará para las frecuencias de la voz humana.
Fuente	Equipo de desarrollo
Prioridad	Alta

Tabla 95. RUR-W14

4.4.2 Requisitos de Software

La plantilla que se utilizará para especificar los requisitos software consta de los siguientes campos que se pasan a describir a continuación:

- **Identificador:** Representa unívocamente a cada requisito. La nomenclatura utilizada será:
 - RSF-AXX: para los requisitos de software funcionales de Android.
 - RSF-WXX: para los requisitos de software funcionales del servicio web.
 - RSNF-AXX: para los requisitos de software no funcionales de Android.
 - RSNF-WXX: para los requisitos de software no funcionales del servicio web.



XX será un número desde el 01 hasta el 99.

- **Nombre:** Nombre representativo del requisito.
- **Descripción:** Fragmento en el que se explica brevemente en qué consiste el requisito.
- **Fuente:** Hace referencia a la procedencia del requisito.
- **Prioridad:** Grado de urgencia con el que debe cumplirse el requisito. Puede ser Alta, Media o Baja.

Requisitos de Software Funcionales

Android

Identificador	RSF-A01
Nombre	<u>Insertar un nuevo usuario</u>
Descripción	El sistema creará un nuevo usuario y almacenará sus datos: nombre de usuario, contraseña, SPL y tiempo de silencio entre grabaciones.
Fuente	RUC-A01
Prioridad	Alta

Tabla 96 RSF-A01

Identificador	RSF-A02
Nombre	<u>Borrar un usuario</u>
Descripción	El sistema podrá eliminar al usuario de la base de datos, además de toda su información asociada y su jerarquía de directorios.
Fuente	RUC-A02
Prioridad	Alta

Tabla 97 RSF-A02

Identificador	RSF-A03
Nombre	<u>Modificar los datos de usuario</u>
Descripción	El sistema permitirá al usuario modificar el valor del SPL y del tiempo de silencio entre grabaciones.
Fuente	RUC-A03
Prioridad	Alta

Tabla 98 RSF-A03



Identificador	RSF-A04
Nombre	<u>Login</u>
Descripción	El sistema contará con un servicio para que un usuario pueda loguearse a través de su nombre de usuario y contraseña. Para ello el sistema comprobará que los datos introducidos sean correctos antes de dar acceso.
Fuente	RUC-A04
Prioridad	Alta

Tabla 99 RSF-A04

Identificador	RSF-A05
Nombre	<u>Cierre de sesión</u>
Descripción	El sistema permitirá el cierre de sesión de un usuario que se encuentre actualmente logueado.
Fuente	RUC-A05
Prioridad	Alta

Tabla 100 RSF-A05

Identificador	RSF-A06
Nombre	<u>Grabar</u>
Descripción	El sistema será capaz de grabar distintas somniloquias a través de un mecanismo de detección del nivel de presión sonora.
Fuente	RUC-A06
Prioridad	Alta

Tabla 101 RSF- A06

Identificador	RSF-A07
Nombre	<u>Parar grabación</u>
Descripción	El sistema permitirá parar una grabación que previamente haya sido iniciada.
Fuente	RUC-A07
Prioridad	Alta

Tabla 102 RSF- A07

Identificador	RSF-A08
Nombre	<u>Reproducir grabación</u>
Descripción	El sistema proporcionará al usuario la opción de reproducir sus archivos de audio: somniloquias, neutros y post-procesados.
Fuente	RUC-A08
Prioridad	Alta

Tabla 103 RSF- A08



Identificador	RSF-A09
Nombre	<u>Pausar reproducción</u>
Descripción	El sistema proporcionará al usuario la opción de pausar la reproducción actual.
Fuente	RUC-A09
Prioridad	Alta

Tabla 104 RSF- A09

Identificador	RSF-A10
Nombre	<u>Parar reproducción</u>
Descripción	El sistema proporcionará al usuario la opción de parar la reproducción actual.
Fuente	RUC-A10
Prioridad	Alta

Tabla 105 RSF- A10

Identificador	RSF-A11
Nombre	<u>Acceder a una grabación</u>
Descripción	Se podrá acceder a las grabaciones realizadas a través de la interfaz de la aplicación.
Fuente	RUC-A11
Prioridad	Alta

Tabla 106 RSF- A11

Identificador	RSF-A12
Nombre	<u>Navegación</u>
Descripción	La aplicación contará con pantallas y pestañas para que el usuario pueda navegar entre ellas.
Fuente	RUC-A12
Prioridad	Alta

Tabla 107 RSF- A12

Identificador	RSF-A13
Nombre	<u>Envío de archivos</u>
Descripción	El envío de archivos se hará de manera transparente al usuario, produciéndose cada vez que se genere un nuevo audio. Si se ha terminado de grabar las somniloquias y quedan archivos sin enviar, se esperará hasta que se manden todos.
Fuente	RUC-A13
Prioridad	Alta

Tabla 108 RSF-A13



Identificador	RSF-A14
Nombre	<u>Creación de jerarquías</u>
Descripción	El sistema creará de manera transparente al usuario las jerarquías. Este proceso se asocia con la creación de un nuevo usuario.
Fuente	RUC-A14
Prioridad	Alta

Tabla 109 RSF-A14

Identificador	RSF-A15
Nombre	<u>Comunicación</u>
Descripción	La aplicación móvil se comunicará con el servicio web a través de un protocolo basado en peticiones y respuestas.
Fuente	RUC-A15
Prioridad	Alta

Tabla 110 RSF-A15

Identificador	RSF-A16
Nombre	<u>Almacenamiento de audio</u>
Descripción	La aplicación almacenará las somniloquias generadas en el sistema de ficheros del dispositivo.
Fuente	RUC-A16
Prioridad	Alta

Tabla 111 RSF-A16

Identificador	RSF-A17
Nombre	<u>Obtención del nivel de presión sonora de forma automática en silencio</u>
Descripción	El sistema calculará el nivel de presión sonora del ruido ambiental en el que se encuentra el usuario, para tomarla como referencia en el cálculo del valor del SPL.
Fuente	RUC-A17
Prioridad	Alta

Tabla 112 RSF-A17

Identificador	RSF-A18
Nombre	<u>Obtención del nivel de presión sonora con voz.</u>
Descripción	El sistema calculará el nivel de presión sonora de la voz del usuario, para tomarla como referencia en el cálculo del valor del SPL.
Fuente	RUC-A17
Prioridad	Alta

Tabla 113 RSF-A18



Identificador	RSF-A19
Nombre	<u>Cálculo del SPL</u>
Descripción	El sistema calculará el SPL a través de un algoritmo que tendrá como valores de entrada los resultados de la detección en silencio y con voz.
Fuente	RUC-A17
Prioridad	Alta

Tabla 114 RSF-A19

Identificador	RSF-A20
Nombre	<u>Ver resultados</u>
Descripción	La aplicación dispondrá de una opción para visualizar los resultados que devuelve el servidor.
Fuente	RUC-18
Prioridad	Alta

Tabla 115 RSF-A20

Identificador	RSF-A21
Nombre	<u>Ver ayuda</u>
Descripción	La aplicación dispondrá de un tutorial de ayuda para el usuario. Este manual explica la funcionalidad principal del sistema.
Fuente	RUC-19
Prioridad	Alta

Tabla 116 RSF-A21

Identificador	RSF-A22
Nombre	<u>Generación de archivos de audios</u>
Descripción	La aplicación generará un nuevo archivo de audio tras cumplirse un tiempo de espera sin detectar sonidos en una misma grabación.
Fuente	RUC-A20
Prioridad	Alta

Tabla 117 RSF-A22

Identificador	RSF-A23
Nombre	<u>Recibir archivos de audio</u>
Descripción	La aplicación móvil podrá recibir archivos de audio procesados en la plataforma web, además los almacena en la jerarquía de ficheros del usuario.
Fuente	RUC-A21, RUC-A16
Prioridad	Alta

Tabla 118. RSF-A23



Identificador	RSF-A24
Nombre	<u>Recibir archivos de texto</u>
Descripción	La aplicación móvil podrá recibir archivos de texto con los resultados generados en la plataforma web, además los almacena en la jerarquía de ficheros del usuario.
Fuente	RUC-A21, RUC-A16
Prioridad	Alta

Tabla 119. RSF-A24

Identificador	RSF-A25
Nombre	<u>Modificar SPL</u>
Descripción	El sistema ofrecerá la posibilidad de calcular el SPL tantas veces como desee el usuario. El nuevo valor se actualizará en la base de datos.
Fuente	RUC-A22
Prioridad	Alta

Tabla 120. RSF-A25

Identificador	RSF-A26
Nombre	<u>Modificar tiempo de silencio entre grabaciones</u>
Descripción	El sistema ofrecerá la posibilidad de modificar el tiempo de silencio entre grabaciones. El nuevo valor se actualizará en la base de datos.
Fuente	RUC-A23
Prioridad	Alta

Tabla 121. RSF-A26

Servicio Web

Identificador	RSF-W01
Nombre	<u>Insertar un nuevo usuario</u>
Descripción	El servlet insertará en la base de datos un nuevo usuario con estos campos: <ul style="list-style-type: none">• Nombre de usuario.• Contraseña.• SPL.• Tiempo de silencio entre grabaciones.
Fuente	RUC-W01
Prioridad	Alta

Tabla 122 RSF-W01



Identificador	RSF-W02
Nombre	<u>Dar de baja a un usuario registrado</u>
Descripción	El servlet eliminará a un usuario previamente creado de la base de datos, así como su jerarquía.
Fuente	RUC-W02
Prioridad	Alta

Tabla 123 RSF-W02

Identificador	RSF-W03
Nombre	<u>Modificar usuario</u>
Descripción	El servidor podrá actualizar los campos de un usuario en la base de datos.
Fuente	RUC-W03
Prioridad	Alta

Tabla 124 RSF-W02

Identificador	RSF-W04
Nombre	<u>Almacenamiento de audio</u>
Descripción	El sistema guardará automáticamente, en la jerarquía del usuario, todos sus archivos de audio recibidos o procesados. Cada archivo estará asociado únicamente a un usuario.
Fuente	RUC-W04
Prioridad	Alta

Tabla 125 RSF-W02

Identificador	RSF-W05
Nombre	<u>Almacenamiento de texto</u>
Descripción	El sistema guardará automáticamente, en la jerarquía del usuario, todos sus archivos de resultados generados. Cada archivo estará asociado únicamente a un usuario.
Fuente	RUC-W04
Prioridad	Alta

Tabla 126 RSF-W05

Identificador	RSF-W06
Nombre	<u>Transcripción</u>
Descripción	La aplicación será capaz de transcribir a texto un archivo de audio almacenado. Para ello es necesario que el idioma empleado sea el castellano.
Fuente	RUC-W05
Prioridad	Alta

Tabla 127 RSF-W06



Identificador	RSF-W07
Nombre	<u>Análisis de emociones</u>
Descripción	El servicio web contará con un módulo encargado del análisis emocional que consistirá en la detección de las siguientes emociones básicas: alegría, enfado, tristeza y miedo.
Fuente	RUC-W06
Prioridad	Alta

Tabla 128. RSF-W07

Identificador	RSF-W08
Nombre	<u>Recibir peticiones</u>
Descripción	El servlet podrá recibir peticiones del dispositivo móvil.
Fuente	RUC-W07
Prioridad	Alta

Tabla 129. RSF-W08

Identificador	RSF-W09
Nombre	<u>Procesar peticiones en servlet simple</u>
Descripción	El servlet simple procesará las peticiones a través de la lectura del valor de un parámetro.
Fuente	RUC-W08
Prioridad	Alta

Tabla 130. RSF-W09

Identificador	RSF-W10
Nombre	<u>Procesar peticiones en servlet fichero</u>
Descripción	El servlet fichero procesará una petición a través de la lectura de su cabecera, distinguiendo entre “neutro”, “archivo” y “procesado”.
Fuente	RUC-W08
Prioridad	Alta

Tabla 131. RSF-W10

Identificador	RSF-W11
Nombre	<u>Enviar ficheros de audio</u>
Descripción	El servlet podrá mandar ficheros de audio al dispositivo móvil. El envío se realizará en la respuesta de una petición.
Fuente	RUC-W09
Prioridad	Alta

Tabla 132. RSF-W11



Identificador	RSF-W12
Nombre	<u>Enviar ficheros de texto</u>
Descripción	El servlet podrá mandar ficheros de texto al dispositivo móvil. El envío se realizará en la respuesta de una petición.
Fuente	RUC-W10
Prioridad	Alta

Tabla 133. RSF-W12

Identificador	RSF- W13
Nombre	<u>Enviar cadenas de texto</u>
Descripción	El servlet podrá mandar cadenas de texto al dispositivo móvil. El envío se realizará en la respuesta de una petición.
Fuente	RUC-W07
Prioridad	Alta

Tabla 134. RSF-W13

Identificador	RSF-W14
Nombre	<u>Crear jerarquía</u>
Descripción	El servicio web podrá crear jerarquías de directorios asociadas a un usuario para almacenar archivos. La creación de la jerarquía es una fase de la creación de nuevos usuarios en el sistema.
Fuente	RUC-W11
Prioridad	Alta

Tabla 135. RSF-W14

Identificador	RSF-W15
Nombre	<u>Eliminar ruidos</u>
Descripción	El servicio web podrá procesar somniloquias y eliminar los ruidos y ronquidos, a través de un análisis acústico de las mismas.
Fuente	RUC-W12
Prioridad	Alta

Tabla 136. RSF-W15

Identificador	RSF-W16
Nombre	<u>Modificar SPL</u>
Descripción	El servlet podrá actualizar el valor del SPL de un usuario en la base de datos.
Fuente	RUC-W13
Prioridad	Alta

Tabla 137. RSF-W16



Identificador	RSF-W17
Nombre	<u>Modificar tiempo de silencio entre grabaciones</u>
Descripción	El servlet podrá actualizar el valor del tiempo de silencio entre grabaciones asociado a un usuario en la base de datos.
Fuente	RUC-W16
Prioridad	Alta

Tabla 138. RSF-W17

Identificador	RSF-W18
Nombre	<u>Obtener características fundamentales de un audio</u>
Descripción	El servlet podrá realizar análisis acústicos de los ficheros con emoción neutra y los archivos procesados para obtener sus características fundamentales.
Fuente	RUC-W17
Prioridad	Alta

Tabla 139. RSF-W18

Identificador	RSF-W19
Nombre	<u>Insertar características fundamental de un archivo neutro</u>
Descripción	El servlet podrá almacenar en la base de datos las características fundamentales de un archivo neutro.
Fuente	RUC-W18
Prioridad	Alta

Tabla 140. RSF-W19

Requisitos de Software No Funcionales

Android

Identificador	RSNF-A01
Nombre	<u>Crear una interfaz de usuario sencilla e intuitiva.</u>
Descripción	La interfaz de usuario del dispositivo móvil se adapta a las características de la pantalla. Será sencilla y poco sobrecargada, ofreciendo tanto comodidad como eficiencia. Utilizará colores contrastados para facilitar la lectura.
Fuente	RUR-A01
Prioridad	Alta

Tabla 141. RSNF-A01

Identificador	RSNF-A02
Nombre	<u>Pantalla de login</u>
Descripción	El sistema contará con una pantalla destinada al login que muestre los siguientes campos a rellenar: <ul style="list-style-type: none">• Nombre de usuario.• Contraseña
Fuente	RUR-A02
Prioridad	Alta

Tabla 142 RSNF-A02

Identificador	RSNF-A03
Nombre	<u>Formato campos de login</u>
Descripción	<ul style="list-style-type: none">• Nombre de usuario: Conjunto de caracteres alfanuméricos que definen unívocamente al usuario.• Contraseña: Conjunto de caracteres alfanuméricos.
Fuente	RUR-A02
Prioridad	Alta

Tabla 143 RSNF-A03

Identificador	RSNF-A04
Nombre	<u>Ocultación de contraseñas</u>
Descripción	La aplicación no mostrará el texto de la contraseña, en vez de eso aparecerá un asterisco “*” por cada letra.
Fuente	RUR-A03
Prioridad	Alta

Tabla 144. RSNF-A04

Identificador	RSNF-A05
Nombre	<u>Contraseñas cifradas</u>
Descripción	Por motivos de seguridad la contraseña se enviará cifrada al servidor. La contraseña se cifrará con el algoritmo DES.
Fuente	RUR-A04
Prioridad	Alta

Tabla 145. RSNF-A05

Identificador	RSNF-A06
Nombre	<u>Grabación inteligente mediante SPL</u>
Descripción	La aplicación móvil detectará cuando el usuario hable mediante el cálculo del nivel de presión sonora (SPL). El sistema almacenará aquellos fragmentos de audio que hayan sido detectados tras superar el SPL configurado por el usuario.
Fuente	RUR-A05
Prioridad	Alta

Tabla 146. RSNF-A06



Identificador	RSNF-A07
Nombre	<u>Pantalla para crear un nuevo usuario</u>
Descripción	La interfaz de la pantalla de creación de un nuevo usuario mostrará los siguientes campos: <ul style="list-style-type: none">• Nombre de usuario.• Contraseña.• Repetir contraseña.
Fuente	RUR-A06
Prioridad	Alta

Tabla 147 RSNF-A07

Identificador	RSNF-A08
Nombre	<u>Insertar un nuevo usuario</u>
Descripción	El sistema impedirá la creación de cuentas con nombre de usuario repetido. Para ello comprobará en el servidor que el nombre de usuario no exista.
Fuente	RUC-A01
Prioridad	Alta

Tabla 148 RSNF-A08

Identificador	RSNF-A09
Nombre	<u>Formato campos de nuevo usuario</u>
Descripción	<ul style="list-style-type: none">• Nombre de usuario: Conjunto de caracteres alfanuméricos que definen unívocamente al usuario.• Contraseña: Conjunto de caracteres alfanuméricos.• Repetir contraseña: mismo formato que contraseña.
Fuente	RUR-A06
Prioridad	Alta

Tabla 149 RSNF-A09

Identificador	RSNF-A10
Nombre	<u>Valor por defecto el tiempo de espera entre audios</u>
Descripción	El tiempo que la aplicación espera en los momentos de silencio tendrá un valor por defecto de 5 minutos.
Fuente	RUR-A07
Prioridad	Alta

Tabla 150. RSNF-A10

Identificador	RSNF-A11
Nombre	<u>Rango de valores del tiempo de espera entre audios</u>
Descripción	El valor mínimo del tiempo en silencio entre audios será 0 segundos y el valor máximo será 60 minutos.
Fuente	RUR-A08
Prioridad	Alta

Tabla 151. RSNF-A11

Identificador	RSNF-A12
Nombre	<u>Cierre de sesión</u>
Descripción	Se saldrá del sistema pulsando el botón “Cerrar Sesión” dentro del menú de Opciones. Para eso el usuario registrado tendrá que tener la sesión activa.
Fuente	RUC-A05
Prioridad	Alta

Tabla 152 RSNF-A12

Identificador	RSNF-A13
Nombre	<u>Iniciar grabación</u>
Descripción	Una grabación comienza cuando el usuario pulsa el botón “record” en la pestaña de “Grabar y Reproducir”. Cuando el usuario pulse el botón se creará una carpeta con el nombre correspondiente al día actual, en caso de que no estuviese creada con anterioridad.
Fuente	RUR-A05
Prioridad	Alta

Tabla 153 RSNF-A13

Identificador	RSNF-A14
Nombre	<u>Parar grabación</u>
Descripción	El sistema permitirá parar la grabación dentro de la pestaña “Grabar y Reproducir” pulsando sobre el botón “stop”. Previamente se tiene que haber iniciado la grabación.
Fuente	RUC-A07
Prioridad	Alta

Tabla 154 RSNF-A14

Identificador	RSNF-A15
Nombre	<u>Reproducir grabación</u>
Descripción	El sistema permitirá reproducir una grabación dentro de la pestaña “Grabar y Reproducir” pulsando sobre el botón “play”.
Fuente	RUC-A08
Prioridad	Alta

Tabla 155 RSNF-A15

Identificador	RSNF-A16
Nombre	<u>Pausar reproducción</u>
Descripción	El sistema permitirá pausar la reproducción de una grabación dentro de la pestaña “Grabar y Reproducir” pulsando sobre el botón “pause”. Previamente se tiene que haber iniciado la reproducción.
Fuente	RUC-A09
Prioridad	Alta

Tabla 156 RSNF-A16

Identificador	RSNF-A17
Nombre	<u>Parar reproducción</u>
Descripción	El sistema permitirá parar una reproducción dentro de la pestaña “Grabar y Reproducir” pulsando sobre el botón “stop”. Previamente se tiene que haber iniciado la reproducción.
Fuente	RUC-A10
Prioridad	Alta

Tabla 157 RSNF-A17

Identificador	RSNF-A18
Nombre	<u>Pantalla de detección del nivel de presión sonora en silencio.</u>
Descripción	Se proporcionará una pantalla para la detección automática del nivel de presión sonora en silencio.
Fuente	RUR-A20, RUC-A17
Prioridad	Alta

Tabla 158 RSNF-A18

Identificador	RSNF-A19
Nombre	<u>Pantalla de detección del nivel de presión sonora con voz.</u>
Descripción	Se proporcionará una pantalla para la detección automática del nivel de presión sonora con voz.
Fuente	RUR-A20, RUC-A17
Prioridad	Alta

Tabla 159 RSNF-A19

Identificador	RSNF-A20
Nombre	<u>Pantalla de ayuda</u>
Descripción	Se contará con una pantalla de ayuda donde se explican las principales funciones de la aplicación.
Fuente	RUC-A19
Prioridad	Alta

Tabla 160 RSNF-A20

Identificador	RSNF-A20
Nombre	<u>Pestaña de configuración</u>
Descripción	El sistema contará con una pestaña para la configuración de datos del usuario. Esta pestaña tendrá tres funcionalidades: <ul style="list-style-type: none">• Detección automática del SPL.• Tiempo en silencio.• Borrar usuario.
Fuente	RUC-A23
Prioridad	Alta

Tabla 161. RSNF-A20

Identificador	RSNF-A21
Nombre	<u>Eliminar un usuario registrado</u>
Descripción	En el apartado de Configuración existirá un botón para eliminar un usuario. Cuando el usuario lo pulse, y tras aceptar confirmación, se elimina de la base de datos así como su jerarquía de ficheros.
Fuente	RUC-A02
Prioridad	Alta

Tabla 162 RSNF-A21

Identificador	RSNF-A22
Nombre	<u>Tiempo de silencio entre audios</u>
Descripción	El tiempo de silencio entre audio tiene que seguir el siguiente formato: mm:ss. Además existirá el botón “Guardar” que permite almacenar en la base de datos el tiempo de silencio.
Fuente	RUR-A07, RUR-A08
Prioridad	Alta

Tabla 163. RSNF-A22

Identificador	RSNF-A23
Nombre	<u>Pestaña de resultados</u>
Descripción	El sistema contará con una pestaña para visualizar los resultados obtenidos en el servidor. Tendrá un botón para cambiar el archivo que se desee mostrar.
Fuente	RUC-A18
Prioridad	Alta

Tabla 164. RSNF-A23



Identificador	RSNF-A23
Nombre	<u>Dispositivo móvil</u>
Descripción	La aplicación se podrá ejecutar en dispositivos móviles con características similares al Samsung Galaxy S (I9000).
Fuente	RUR-A09
Prioridad	Alta

Tabla 165. RSNF-A23

Identificador	RSNF-A24
Nombre	<u>Base de datos</u>
Descripción	La aplicación dispondrá de una base de datos (SQLite) para el almacenamiento de información.
Fuente	RUR-A10
Prioridad	Alta

Tabla 166. RSNF-A24

Identificador	RSNF-A25
Nombre	<u>Sistema operativo</u>
Descripción	La aplicación será compatible con el sistema operativo para móviles Android.
Fuente	RUR-A11
Prioridad	Alta

Tabla 167. RSNF-A25

Identificador	RSNF-A26
Nombre	<u>Versión de Android</u>
Descripción	La aplicación se podrá ejecutar en cualquier dispositivo móvil con versión de la plataforma Android 2.3.3 (API 10) y superiores.
Fuente	RUR-A12
Prioridad	Alta

Tabla 168. RSNF-A26

Identificador	RSNF-A27
Nombre	<u>Protocolo de comunicación</u>
Descripción	Las comunicaciones con el servidor se harán mediante el protocolo HTTP.
Fuente	RUR-A13
Prioridad	Alta

Tabla 169. RSNF-A27



Identificador	RSNF-A28
Nombre	<u>Formato de archivos de somniloquias originales</u>
Descripción	Los archivos de audio tendrán el siguiente formato: usuario-dd_mm_yy-hh_mm_ss, seguido de la extensión “.wav”.
Fuente	RUR-A14
Prioridad	Alta

Tabla 170. RSNF-A28

Identificador	RSNF-A29
Nombre	<u>Formato de archivos de somniloquias procesadas</u>
Descripción	Los archivos de audio tendrán el siguiente formato: usuario-dd_mm_yy-hh_mm_ss-SR, seguido de la extensión “.wav”.
Fuente	RUR-A15
Prioridad	Alta

Tabla 171. RSNF-A29

Identificador	RSNF-A30
Nombre	<u>Formato de archivos neutros</u>
Descripción	Los archivos de audio tendrán el siguiente formato: Neutro-usuario-dd_mm_yy-hh_mm_ss, seguido de la extensión “.wav”.
Fuente	RUR-A16
Prioridad	Alta

Tabla 172. RSNF-A30

Identificador	RSNF-A31
Nombre	<u>Formato de archivos de resultados</u>
Descripción	Los archivos de texto con los resultados tendrán el siguiente formato: usuario-dd_mm_yy-hh_mm_ss-R, seguido de la extensión “.txt”.
Fuente	RUR-A17
Prioridad	Alta

Tabla 173 RSNF- A31



Identificador	RSNF-A32
Nombre	<u>Formato de jerarquía</u>
Descripción	<p>El formato .de la jerarquía contiene una carpeta por cada usuario identificada por su nombre de usuario. Dentro de esta carpeta principal existirán cuatro carpetas:</p> <ul style="list-style-type: none">• Neutro: incluye el archivo con la emoción neutra.• Somniloquias originales: se organiza por fecha y contiene las generadas.• Somniloquias sin ruidos: se organiza por fecha y contiene los archivos post-procesados.• Resultados: se organiza por fecha y contiene los archivos de resultados.
Fuente	RUR-A18
Prioridad	Alta

Tabla 174 RSNF- A32

Identificador	RSNF-A33
Nombre	<u>Idioma</u>
Descripción	La aplicación mostrará los textos en español.
Fuente	RUR-A19
Prioridad	Alta

Tabla 175. RSNF-A33

Identificador	RSNF-A34
Nombre	<u>Configuración inicial del valor del nivel de presión sonora</u>
Descripción	Una de las fases de la creación de usuario es la detección del valor del SPL.
Fuente	RUR-A20
Prioridad	Alta

Tabla 176. RSNF-A34

Servicio Web

Identificador	RSNF-W01
Nombre	<u>Emociones primarias</u>
Descripción	El sistema detectará las emociones primarias de alegría, enfado, tristeza y miedo en los archivos de audio a partir de las características fundamentales.
Fuente	RUR-W01
Prioridad	Alta

Tabla 177 RSNF-W01

Identificador	RSNF-W02
Nombre	<u>Base de datos plataforma web</u>
Descripción	El sistema contará con una base de datos MySQL.
Fuente	RUR-W02
Prioridad	Alta

Tabla 178 RSNF-W02

Identificador	RSNF-W03
Nombre	<u>Servidor Web</u>
Descripción	El sistema contará con un Servidor Web para responder las peticiones de los clientes. El servidor web empleado será un Apache Tomcat.
Fuente	RUR-W03
Prioridad	Alta

Tabla 179 RSNF-W03

Identificador	RSNF-W04
Nombre	<u>Formato de archivos de somniloquias originales</u>
Descripción	Los archivos de audio tendrán el siguiente formato: usuario-dd_mm_yy-hh_mm_ss, seguido de la extensión “.wav”.
Fuente	RUR-W04
Prioridad	Alta

Tabla 180 RSNF-W04

Identificador	RSNF-W05
Nombre	<u>Formato de archivos de somniloquias procesadas</u>
Descripción	Los archivos de audio tendrán el siguiente formato: usuario-dd_mm_yy-hh_mm_ss-SR, seguido de la extensión “.wav”.
Fuente	RUR-W05
Prioridad	Alta

Tabla 181 RSNF-W05

Identificador	RSNF-W06
Nombre	<u>Formato de archivos neutros</u>
Descripción	Los archivos de audio tendrán el siguiente formato: Neutro-usuario-dd_mm_yy-hh_mm_ss-SR, seguido de la extensión “.wav”.
Fuente	RUR-W06
Prioridad	Alta

Tabla 182 RSNF-W06



Identificador	RSNF-W07
Nombre	<u>Formato de archivos de resultados</u>
Descripción	Los archivos de texto con los resultados tendrán el siguiente formato: usuario-dd_mm_yy-hh_mm_ss-R, seguido de la extensión “.txt”.
Fuente	RUR-W07
Prioridad	Alta

Tabla 183 RSNF-W07

Identificador	RSNF-W08
Nombre	<u>Formato de jerarquía de directorios</u>
Descripción	La jerarquía estará compuesta por cuatro carpetas principales: “Neutro”, “Resultado”, “Somniloquias_originales” y “Somniloquias_sin_ruidos”. Cada usuario registrado posee dentro de ellas una carpeta con su nombre y los ficheros relacionados.
Fuente	RUR-W08
Prioridad	Alta

Tabla 184 RSNF-W08

Identificador	RSNF-W09
Nombre	<u>Transcripción</u>
Descripción	El servicio web realizará este proceso de manera automática. La transcripción se realiza mediante peticiones a una URL que procesa el audio y devuelve un resultado. Este resultado se parseará para obtener únicamente el texto que nos interesa.
Fuente	RUR-W09
Prioridad	Alta

Tabla 185 RSNF-W09



Identificador	RSNF-W10
Nombre	<u>Resultado del análisis emocional</u>
Descripción	<p>El resultado del análisis de emociones de un archivo se almacenará en un archivo de texto que contendrá los siguientes campos:</p> <ul style="list-style-type: none">• Emoción• Transcripción• Energía• Pitch• Rango del pitch• SPL• Rango del SPL• Media• Varianza• Velocidad
Fuente	RUR-W10
Prioridad	Alta

Tabla 186 RSNF-W10

Identificador	RSNF-W11
Nombre	<u>Post-procesado de la grabación</u>
Descripción	<p>La aplicación realizará un post-procesado de los archivos de audio para eliminar los fragmentos que no correspondan al habla del usuario, como ronquidos o llantos. Para ello hará un análisis acústico de la energía y los cruces por cero de la señal.</p>
Fuente	RUR-W11
Prioridad	Alta

Tabla 187 RSNF-W11

Identificador	RSNF-W12
Nombre	<u>Almacenamiento</u>
Descripción	<p>Toda la información de los usuarios y los valores neutros se almacenarán en la base de datos del servicio web. De los usuarios se guardará:</p> <ul style="list-style-type: none"> • Nombre de usuario • Contraseña • SPL • Tiempo de silencio entre somniloquias • De los archivos neutros se guardará: <ul style="list-style-type: none"> • Nombre de usuario • Pitch • Rango del pitch • Energía • SPL Rango del SPL • Media • Varianza • Velocidad
Fuente	RUR-W12
Prioridad	Alta

Tabla 188. RSNF-W12

Identificador	RSNF-W13
Nombre	<u>Tipo de información</u>
Descripción	<p>Cada tipo de información que se almacene en el sistema tendrá un formato adecuado. A continuación se especifica el formato:</p> <ul style="list-style-type: none"> • Nombre de usuario: cadena de caracteres. • Contraseña: cadena de caracteres • SPL: double. • Tiempo de silencio entre somniloquias: double. • Pitch: double. • Rango del pitch: double. • Energía: double. • SPL: double. • Rango del SPL: double. • Media: double. • Varianza: double. • Velocidad: double.
Fuente	RUR-W12
Prioridad	Alta

Tabla 189. RSNF-W13

Identificador	RSNF-W14
Nombre	<u>Comunicación</u>
Descripción	La comunicación con el dispositivo móvil se hará por HTTP. El servlet contestará a las peticiones que reciba.
Fuente	RUR-W13
Prioridad	Alta

Tabla 190. RSNF-W14

Identificador	RSNF-W15
Nombre	<u>Filtrar por frecuencias</u>
Descripción	El filtrado de audios por frecuencias se hará para frecuencias mayores de 82 Hz y menores a 1056 Hz, debido a que este es el rango de la voz humana.
Fuente	RUR-W14
Prioridad	Alta

Tabla 191. RSNF-W15

4.5 Arquitectura

La arquitectura cliente-servidor es un modelo de aplicación distribuida entre múltiples procesadores donde hay clientes que solicitan servicios y servidores que los proporcionan. Normalmente se utiliza para separar los servicios, situando cada uno en su plataforma más adecuada.

Nuestro sistema se ajusta a esta arquitectura, debido a que el dispositivo móvil se comporta como el cliente, realizando peticiones al servicio web que es el que las procesa y emite las respuestas.

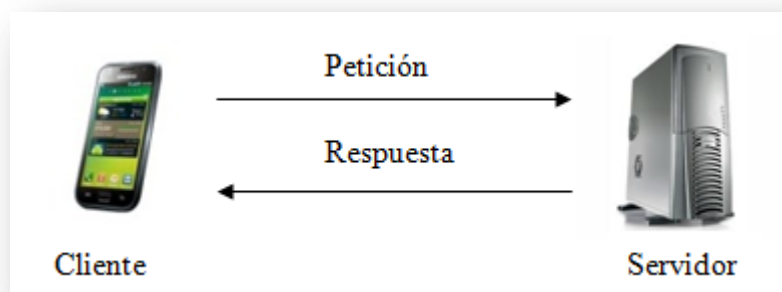


Ilustración 8. Arquitectura cliente-servidor

4.6 Diagrama de clases

En esta sección se muestra la estructura del sistema a partir de sus clases. La vista que proponemos no es la clásica, con todos los atributos y métodos por cada clase, debido a que nos extenderíamos en exceso con ello, al contar con una gran cantidad de clases en el proyecto. Los diagramas únicamente incluyen las clases y la relación que poseen con el resto de módulos del sistema, es decir, si una clase del módulo que se explica se asocia con otra clase de otro módulo, habrá una relación en el diagrama clase que lo represente. La descripción de los módulos se describe a continuación en dos bloques separados: plataforma móvil y servicio web.

Plataforma Móvil

4.6.1 Módulo Core Functionality

En el siguiente diagrama de clases se muestra el módulo encargado de la funcionalidad principal de la aplicación móvil.

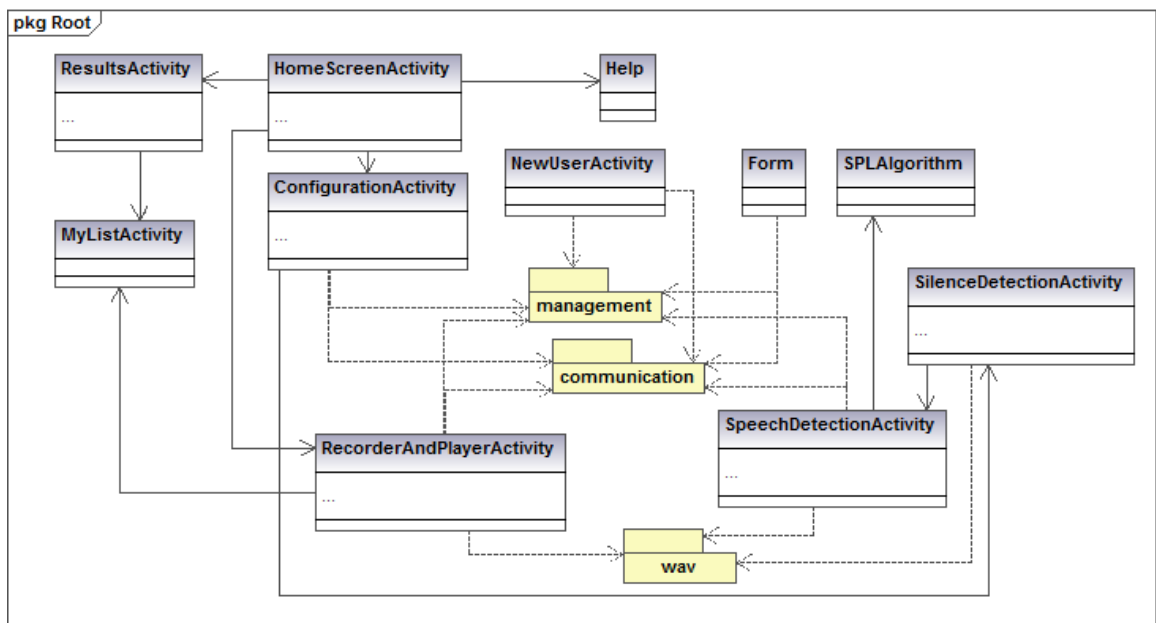


Ilustración 9. Diagrama de clases de Core Functionality

Descripción de las clases:

- **RecorderAndPlayerActivity:** Esta clase implementa la funcionalidad relacionada con la grabación y reproducción de archivos de audio mediante una interfaz de usuario.
 - Asociación externa: Se comunica con el módulo “wav” a través de las clases “WavAudioPlayer” y “WavAudioRecorder”. Se comunica con el módulo “management” a través de las clases “FileManagement”, “User” y “UserDataManagement”. Por último se relaciona con el módulo “communication” a través de la clase “Communication”:

- **ConfigurationActivity:** Esta clase contiene la interfaz dedicada a la ventana de configuración de usuario. En ella se implementan las funcionalidades siguientes: detección automática del SPL, modificar el tiempo en silencio y borrar usuario.
 - Asociación externa: Se comunica con el módulo “management” a través de las clases “FileManagement” y “UserDataManagement”, y con el módulo “communication”, a través de la clase “Communication”.

- **NewUserActivity:** Esta clase implementa la funcionalidad dedicada a la creación de un nuevo usuario en la aplicación mediante la introducción de un usuario y contraseña.
 - Asociación externa: Se comunica con el módulo “management” a través de “FileManagement”, “UserDataManagement” y “StringEncrypter” y con el módulo “communication”, a través de “Communication”.

- **Form:** Esta clase contiene la interfaz que permite realizar el *login* de un usuario mediante un usuario y contraseña.
 - Asociación externa: Se comunica con el módulo “communication” a través de la clase “Communication” y con el módulo “management” a través de “FileManagement”, “User”, “UserDataManagement” y “StringEncrypter”.



- **SilenceDetectionActivity:** Esta clase implementa la funcionalidad relacionada con la detección del SPL del ruido ambiental. El resultado se almacena para después tenerlo en cuenta en el cálculo del SPL de activación del usuario.
 - Asociación externa: Se comunica con el módulo “wav” a través de la clase “WavAudioRecorderWithoutFile”.
- **SpeechDetectionActivity:** Esta clase implementa la funcionalidad relacionada con la detección del SPL que se registra durante una locución del usuario en voz alta. Además se realiza la grabación del habla del usuario en un archivo de audio.
 - Asociación externa: Se comunica con el módulo “wav” a través de “WavAudioRecorder”. Se comunica con el módulo “management” a través de “FileManagement”, y “UserDataManagement”, y con el módulo “communication” a través de la clase “Communication”.
- **SPLAlgorithm:** Esta clase implementa un algoritmo que calcula el SPL de activación del usuario a partir de las intensidades captadas por el micrófono en la fase de silencio y de detección del habla.
- **Help:** Esta clase contiene la interfaz dedicada a mostrar la información de ayuda al usuario.
- **MyListActivity:** Esta clase contiene la interfaz dedicada a mostrar un listado de archivos y directorios. Permite seleccionar un archivo para su futura lectura.
- **HomeScreenActivity:** Esta clase implementa la interfaz de la pantalla principal de usuario, que contiene las pestañas de navegación.
- **ResultsActivity:** Esta clase contiene la interfaz que muestra los resultados de un archivo de audio generados en el servidor.

4.6.2 Módulo Wav

En el siguiente diagrama de clases se muestra el módulo encargado del tratamiento de los archivos de audio con formato wav.

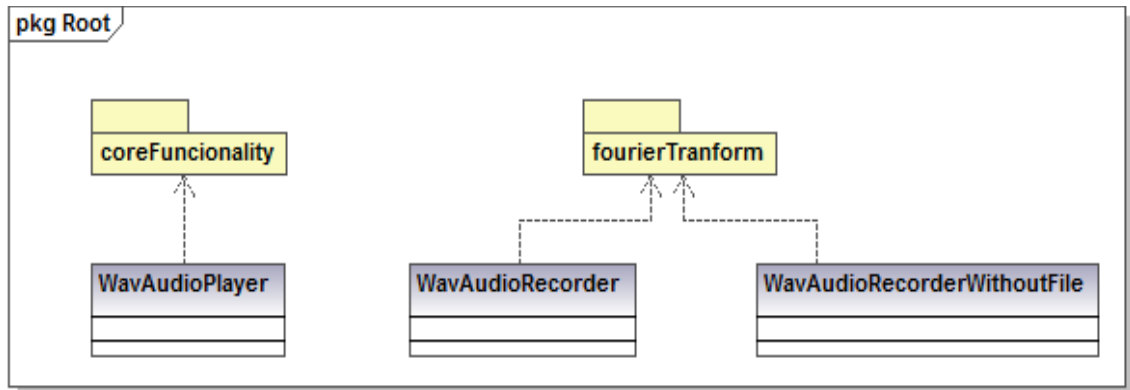


Ilustración 10. Diagrama de clases de Wav

Descripción de las clases:

- **WavAudioRecorder:** Esta clase se encarga de grabar los sonidos que el micrófono del dispositivo capta en archivos de audio que siguen el formato wav. Además prepara y escribe los archivos de este formato. Calcula los campos relacionados con la cabecera antes de crear archivos de audio wav.
 - Asociación externa: Se comunica con el módulo “fourierTransform” a través de las clases “FFT” y “Complex”.
- **WavAudioRecorderWithoutFile:** Esta clase realiza una funcionalidad parecida a la clase “WavAudioRecorder”, pero no realiza la escritura del fichero de audio. Se utiliza para analizar los sonidos captados por el micrófono del dispositivo móvil.
 - Asociación externa: Se comunica con el módulo “fourierTransform” a través de las clases “FFT” y “Complex”.
 -
- **WavAudioPlayer:** Esta clase se encarga de reproducir archivos wav. En primer lugar lee los valores de la cabecera. Después instancia un objeto *AudioTrack* para reproducir los datos que va leyendo de fichero de entrada.
 - Asociación externa: Se comunica con el módulo “coreFuncionalidad” a través de la clase “RecorderAndPlayerActivity”.

4.6.3 Módulo Fourier Transform

En el siguiente diagrama de clases se muestra el módulo encargado del cálculo de la transformada de Fourier.

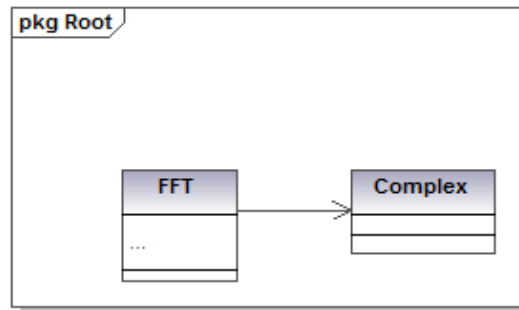


Ilustración 11. Diagrama de clases de Fourier Transform

Descripción de las clases:

- **FFT**: Esta clase se encarga de calcular la transformada rápida de Fourier para pasar al dominio de la frecuencia.
- **Complex**: Esta clase representa un número complejo, necesario para realizar los cálculos de la transformada de Fourier.

4.6.4 Módulo Management

En el siguiente diagrama de clases se muestra el módulo encargado de la gestión de los datos de usuario en la base de datos y la jerarquía de ficheros.

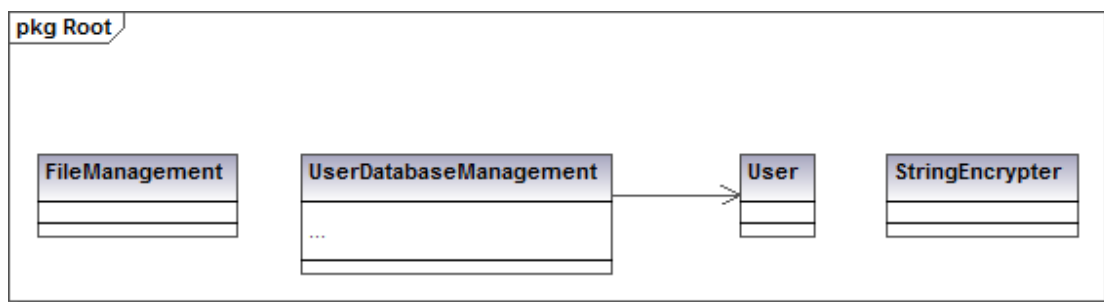


Ilustración 12. Diagrama de clases de Management

Descripción de las clases:

- **FileManagement:** Esta clase se encarga de gestionar la jerarquía de ficheros asociada a un usuario.
- **UserDataManagement:** Esta clase se encarga de gestionar la base de datos con los datos del usuario en el dispositivo móvil.
- **User:** Esta clase gestiona los datos del usuario actual.
- **StringEncrypter:** Esta clase se encarga de encriptar y desencriptar cadenas de texto. Se utiliza para cifrar la contraseña.

4.6.5 Módulo Communication

En el siguiente diagrama de clases se muestra la relación del módulo encargado de la comunicación con el servidor.

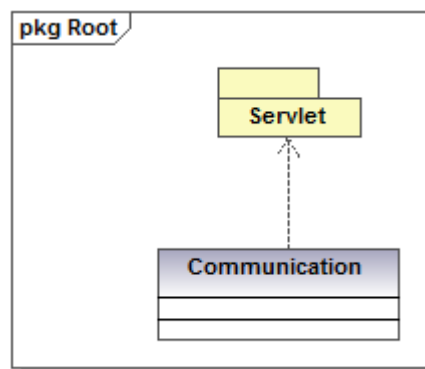


Ilustración 13. Diagrama de clases de Communication

Descripción de las clases:

- **Communication:** esta clase es la encargada de gestionar las peticiones que se hacen al servlet, además de recibir las respuestas.
 - Asociación externa: Se comunica con el módulo “Servlet” a través de las clases “FileServlet” y “SimpleServlet”.

Servicio Web

4.6.6 Módulo Servlet

El siguiente diagrama UML representa la interacción de las clases del módulo “Servlet”.

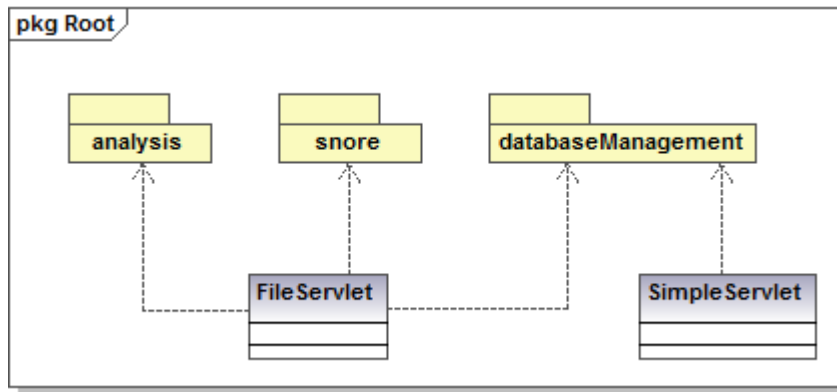


Ilustración 14. Diagrama de clases de Servlet

Descripción de las clases:

- **FileServlet:** es la clase que gestiona todas las peticiones relacionadas con el procesado o envío de ficheros.
 - Asociación externa: Se comunica con el módulo “Analysis” a través de las clases “NeutralAnalyzer” y “EmotionAnalyzer”. Utiliza la clase “PostProcessing” del módulo “Snore”.
- **SimpleServlet:** es la clase que gestiona todas las peticiones relacionadas con el procesado de cadenas de texto.
 - Asociación externa: Se comunica con el módulo “Database Management” mediante las clases “Helper” y “User”.

4.6.7 Módulo Analysis

El siguiente diagrama UML representa la interacción de las clases del módulo “Analysis” entre ellas y con otros módulos del sistema.

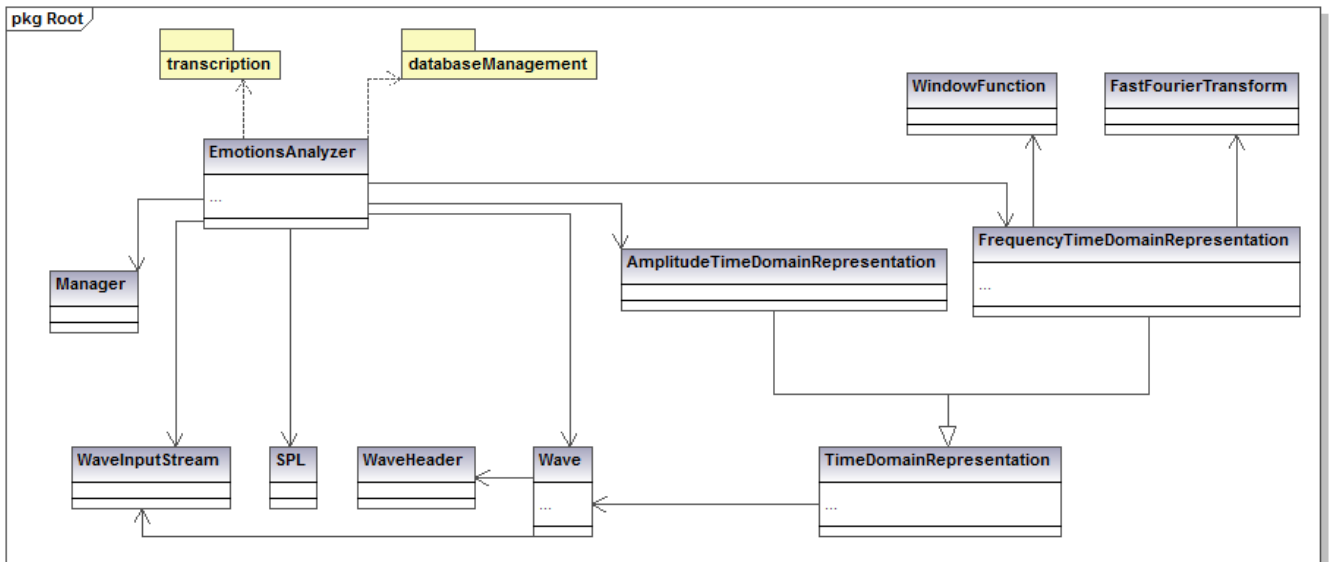


Ilustración 15. Diagrama de clases de Analysis

Descripción de las clases:

- **EmotionsAnalyzer:** es la clase principal del módulo, cuyo propósito es la gestión del proceso de análisis acústico de los ficheros.
 - Asociación externa: Se comunica con el módulo “Transcription” a través de las clases “Transcription” y “Parser”. También accede a la clase “Emotion” del paquete “DatabaseManagement”.
- **Manager:** es la clase que gestiona el reconocimiento de emociones.
- **WaveHeader:** es la clase encargada de determinar si la cabecera del fichero “wav” es correcta o no.
- **WaveInputStream:** es la clase encargada de leer los datos del InputStream de entrada, a excepción de los primeros 44 bytes de los datos, que son los pertenecientes a la cabecera del archivo “wav”.
- **Wave:** Es la clase que gestiona la lectura de los ficheros de tipo “wav”



- **TimeDomainRepresentation**: es la clase encargada de la representación en el dominio del tiempo y ejerce de clase padre de la cual heredan “FrequencyTimeDomainRepresentation” y “AmplitudeTimeDomainRepresentation”.
- **AmplitudeTimeDomainRepresentation**: Esta clase permite manejar los datos de la señal en el dominio de amplitud-tiempo. Su principal función es la de normalizar las amplitudes y devolvérselas al usuario para que pueda trabajar con ellas.
- **FrequencyTimeDomainRepresentation**: es la clase encargada de manejar los datos del audio en el dominio de la frecuencia y de construir el espectrograma de la señal.
- **WindowsFuction**: esta clase permite generar y usar diferentes tipos de ventanas: rectangular, Bartlett, Hanning, Hamming y Blackman.
- **FastFourierTransform**: es la clase encargada de la realización de la Transformada Rápida de Fourier
- **SPL**: es la clase encargada de obtener el SPL y el rango del SPL.

4.6.8 Módulo Transcription

El siguiente diagrama UML representa la interacción de las clases del módulo “Transcription”.

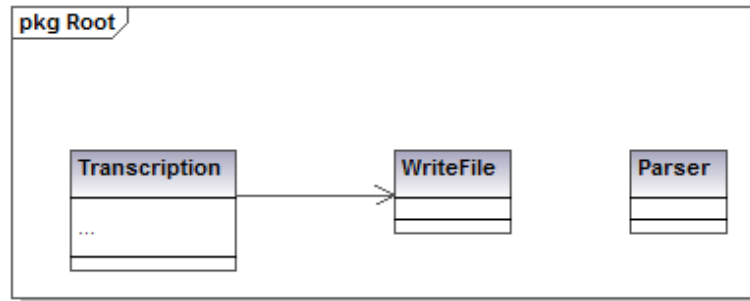


Ilustración 16. Diagrama de clases de Transcription

Descripción de las clases:

- **Transcription:** es la clase principal del módulo, cuyo propósito es la transcripción a texto de un fichero de audio.
- **WriteFile:** es la clase encargada de la escritura de ficheros.
- **Parser:** es la clase encargada de analizar el texto de entrada de un fichero para obtener las partes importantes.

4.6.9 Módulo Snore

El siguiente diagrama UML representa el módulo encargado de la eliminación de ruidos y ronquidos.

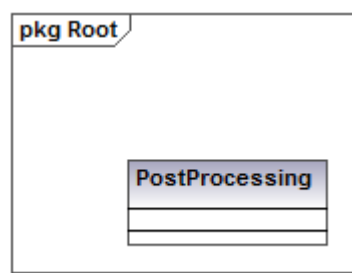


Ilustración 17. Diagrama de clases de Snore

Descripción de las clases:

- **PostProcessing:** Esta clase se encarga de analizar acústicamente las somniloquias del usuario, detecta los fragmentos de ruidos y ronquido y los elimina, generando un nuevo archivo de audio.

4.6.10 Módulo Database Management

El siguiente diagrama UML representa la interacción de las clases del módulo “Database Management”.

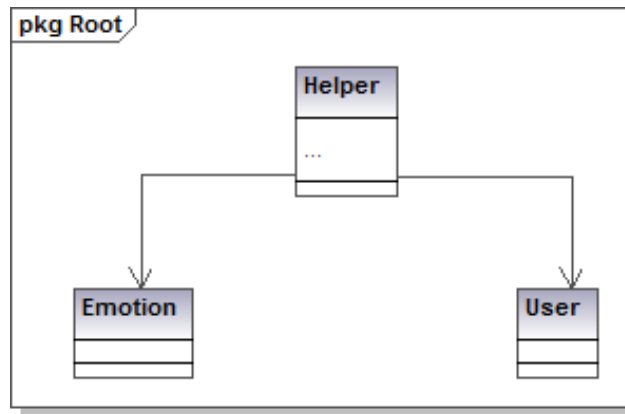


Ilustración 18. Diagrama de clases de Database Management

Descripción de las clases:

- **Helper:** es la clase encargada de gestionar la base de datos y los usuarios del sistema.
- **Emotion:** esta clase representa a una emoción y posee todas las características fundamentales necesarias para el reconocimiento.
- **User:** esta clase representa a un usuario del sistema.





Capítulo 5

Implementación

"Crear es tan difícil como ser libre."

Elsa Triolet (1896-1970)

5.1 Dispositivo móvil

5.1.1 Módulo de funcionalidad principal

Este módulo recoge las principales funciones de la aplicación Android: módulo de Grabación y módulo de detección del SPL. Ha sido desarrollado por Jorge Pérez Muñoz y la información referente a esta sección puede encontrarla en su memoria [24].

5.1.2 Módulo de gestión

Este módulo se encarga de la gestión de usuarios, de ficheros y de la base de datos del dispositivo móvil. Ha sido desarrollado por Jorge Pérez Muñoz y la información referente a esta sección puede encontrarla en su memoria [24].

5.1.3 Módulo de comunicación

Este módulo se encarga de interactuar con el servlet, funcionando según la arquitectura cliente-servidor, debido a que el dispositivo móvil realiza las peticiones al servidor, que le da las respuestas. La comunicación se realiza por HTTP utilizando un cliente HTTP de Apache, el cual permite al dispositivo móvil ejecutar peticiones POST a la URL del servlet.

A continuación se explica mediante un esquema las peticiones que se hacen, las peticiones en color morado son las que interactúan con el servlet que se encarga de tratar las cadenas de texto, mientras que las de color azul se dirigen al servlet encargado de los ficheros.



Ilustración 19. Peticiones al servlet



- **Llamada a servicio Login:** se envía al servidor un nombre de usuario y una contraseña, para que compruebe la validez de los datos y si ha sido dado de alta en el sistema. La contraseña se envía cifrada por motivos de seguridad.
- **Llamada a servicio Nuevo Usuario:** se envía al servidor un nombre de usuario, una contraseña, un valor de SPL y un valor de tiempo entre grabaciones. Estos datos componen los valores de un usuario y le sirven al servlet para poder registrarlo en el sistema.
- **Llamada a servicio Devolver Usuario:** petición de un cliente dado su nombre de usuario.
- **Llamada a servicio Borrar Usuario:** petición que envía el nombre de usuario para que el servlet borre las tablas y directorios asociados a ese cliente.
- **Llamada a servicio SPL:** petición para que se actualice en el servidor los registros de un usuario con el nuevo valor de SPL que se le envía.
- **Llamada a servicio Tiempo:** petición para que se actualice en el servidor los registros de un usuario con el nuevo valor de Tiempo que se le envía.
- **Llamada a servicio Neutro:** se envía al servidor el archivo con la emoción neutra de un usuario para que realice un análisis acústico sobre él y almacene los valores para usos posteriores.
- **Llamada a servicio Archivo:** esta petición se ejecuta una vez que se ha grabado una somniloquia en el móvil, remitiéndola al servlet para que elimine los ruidos del archivo y obteniendo como respuesta la somniloquia procesada.
- **Llamada a servicio Procesado:** al recibir la somniloquia sin ruidos, se abre una conexión con el servidor y se le pasa el nombre del fichero recibido. La tarea del servlet es hacer una comparación de ese fichero con los valores guardados para la emoción neutra y generar un fichero de resultados.

5.2 Servidor

5.2.1 Módulo Servlet

Era necesario contar en el sistema con un servicio web que se encargase de recibir las peticiones del dispositivo móvil, las procesase y devolviese los resultados. El servlet es el que realiza estas tareas, cooperando con el módulo de comunicación del dispositivo móvil. Ambos trabajan mediante una arquitectura cliente-servidor.

Para poder trabajar con los servlet se ha necesitado un Eclipse con soporte J2EE y el servidor Apache Tomcat, debido a que soporta servlets y JSPs. Además se tuvo que configurar Apache Tomcat para poder trabajar con él desde Eclipse lo cual resultó sencillo ya que se encontró mucha información al respecto. Sin embargo para poder trabajar en modo debug de manera remota, el proceso fue más complicado ya que se tuvieron que asignar permisos especiales para puertos y realizar un script para lanzar la aplicación.

Respecto a la implementación, se ha separado la funcionalidad en dos servlets distintos, el primero de ellos se ocupa de peticiones simples, las cuales únicamente poseen "String", el otro servlet se encarga de procesar las peticiones relativas a ficheros.

A continuación se explicará mediante esquemas las tareas de cada servlet, acompañándolo de un resumen.

Servlet Simple

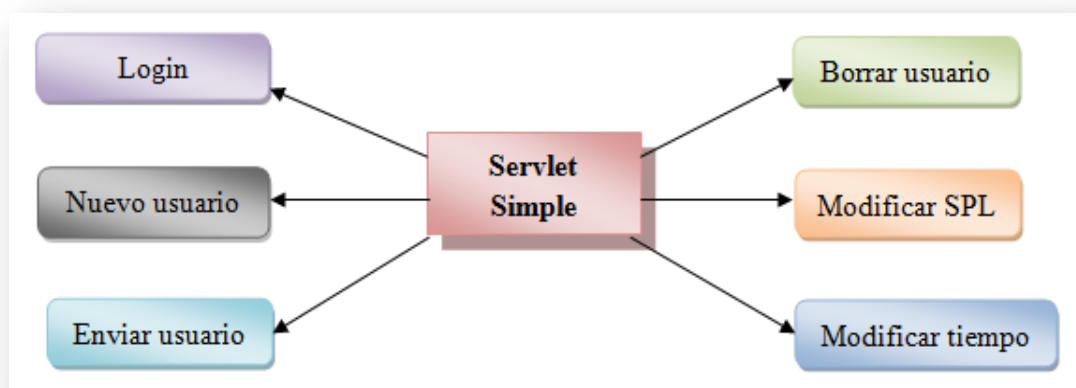


Ilustración 20. Funcionalidad del Servlet Simple

- **Login:** se comprueba que el usuario recibido exista en la base de datos, en ese caso se envía una respuesta afirmativa, en caso contrario se envía una respuesta de error.
- **Nuevo usuario:** se inserta el nuevo usuario en la base de datos y se crea la jerarquía de directorios, si no se han producido errores se envía una respuesta afirmativa, en caso contrario se envía una respuesta de error.
- **Enviar usuario:** se devuelve el usuario acorde con el nombre de usuario que se busca, si no existe se devuelve un mensaje de error, en caso contrario una respuesta afirmativa.
- **Borrar usuario:** se busca el nombre de usuario recibido en la base de datos y se eliminan los registros y la jerarquía de carpetas de ese usuario. Si todo ha ido bien se envía una respuesta afirmativa, en caso contrario un mensaje de error.
- **Modificar SPL:** el servlet recibe el nuevo valor del SPL de un usuario, busca en la base de datos su registro y lo actualiza con la nueva información. Si todo ha ido bien se envía una respuesta afirmativa, en caso contrario un mensaje de error.
- **Modificar tiempo:** el servlet recibe el nuevo valor del tiempo de espera de grabación entre ficheros de un usuario, busca en la base de datos su registro y lo actualiza con la nueva información. Si todo ha ido bien se envía una respuesta afirmativa, en caso contrario un mensaje de error.

Servlet Ficheros

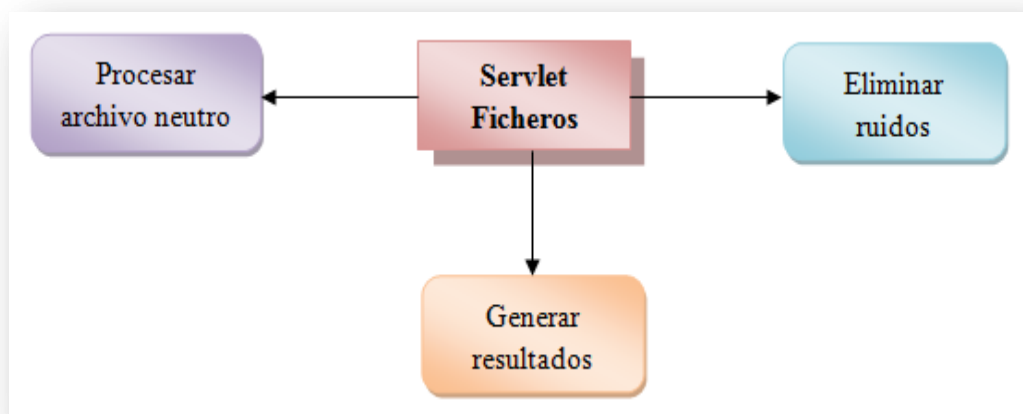


Ilustración 21. Funcionalidad del Servlet Ficheros

- **Procesar archivo neutro:** el servlet recibe un nombre de usuario y el archivo de audio con la emoción neutra para ese usuario. Se procesa acústicamente el fichero para obtener los valores de: pitch, rango del pitch, energía, SPL, rango del SPL, media, varianza y velocidad. Una vez obtenidos se almacenan en la base de datos, si no se ha producido ningún error se devuelve una respuesta afirmativa, en caso contrario se devuelve un mensaje de error.
- **Eliminar ruidos:** cuando en el dispositivo móvil se ha terminado de grabar una somniloquia, ésta se envía al servlet para que la analice y elimine los ruidos que contiene, entendiendo por ruidos: ronquidos, llantos, etc. Cuando el servlet termina de quitar los ruidos, se genera un nuevo archivo que se devuelve al dispositivo móvil. Tanto el fichero con la somniloquia original, como la somniloquia sin ruidos, son almacenados en el servidor en la carpeta correspondiente.
- **Generar resultados:** el servlet recibe el nombre de una somniloquia generada anteriormente. Se busca el fichero que coincide con ese nombre y se ejecuta una serie de algoritmos con el fin de generar un fichero de texto con los resultados. Estos procesos consisten en un análisis acústico para las características fundamentales, seguido de una transcripción a texto del audio y un reconocimiento de la emoción predominante en el fichero. Los resultados obtenidos son enviados al dispositivo móvil para informar al usuario de los valores.


5.2.2 Módulo de gestión de la base de datos

Este módulo se encarga de la gestión de la base de datos MySQL del servidor, proporcionando métodos para: la inserción, borrado y modificación de las diferentes tablas. También ofrece los datos de las tablas para poder trabajar con ellos.

A continuación se muestra el esquema de la base de datos:

Usuario (**user**, password, spl, tiempo)

Neutro (**user**, pitch, rango_pitch, energia, spl, rango_spl, valor_medio, varianza, velocidad)



5.2.3 Módulo de eliminación de ruidos

El módulo de “postprocesado” de los archivos de audio es un módulo encargado de eliminar aquellos fragmentos de audio en los que no exista habla humana. Ha sido desarrollado por Jorge Pérez Muñoz y la información referente a esta sección puede encontrarla en su memoria [24].

5.2.4 Modulo de Transcripción

En esta sección se explicarán los pasos que se han seguido para conseguir realizar la transcripción de audio a texto por medio de un sistema de reconocimiento del habla.

Lo primero que se hizo fue buscar información sobre las herramientas de reconocimiento de voz de código libre existentes en el mercado y se descubrió que los programas más importantes en Open Source son CMU Sphinx y Julius. Sphinx cuenta con tres versiones de sistemas ASR: Sphinx2 y Sphinx3 son librerías de reconocimiento escritas en C, mientras que Sphinx4 es una versión escrita en Java más completa que las anteriores según explican los propios autores.

Julius opera normalmente sobre sistemas Linux y Unix aunque también contiene versiones para trabajar sobre Windows y el lenguaje de programación en el que está creado es C.

Al final nos decantamos por trabajar con sphinx4 ya que el lenguaje de programación sobre el que trabajaba era Java, el cual ya conocíamos previamente y módulos posteriores iban a ser realizados también sobre este lenguaje por lo que de esta manera la integración sería más fácil.

CMU Sphinx

Una vez seleccionado el motor de reconocimiento de voz, se investigó acerca de su funcionamiento. Para ello, la página web del [proyecto](#) resultó bastante útil ya que contiene tutoriales, ejemplos y una wiki con bastante documentación, los cuales ayudaron para aprender a utilizar la herramienta y el proceso de ejecución, así como los archivos implicados en la misma. Los ejemplos de la página están orientados al modelo acústico Wall Street Journal por lo que se empezará explicando esta parte.

Pruebas con el modelo acústico Wall Street Journal

Este modelo ha sido creado con el corpus acústico de la lectura de una serie de artículos del reconocido periódico Wall Street Journal, junto con las transcripciones, es

decir, la versión en texto de los mismos. El corpus como tal tiene una licencia comercial, pero los modelos son de código abierto.

Sphinx cuenta con varios paquetes con los que probar las distintas funcionalidades del programa. A nosotros en concreto nos interesa el paquete “edu.cmu.sphinx.demo.transcriber” el cual contiene una clase llamada Transcriber.java que permite realizar el proceso de transcripción de audio a texto de una manera sencilla, para lo cual se apoya en llamadas a Recognizer, Result y ConfigurationManager.

La clase ConfigurationManager se encarga de leer un archivo interno de configuración que contiene toda la información necesaria como el diccionario, el modelo acústico o la configuración lingüística con la que trabajar. Recognizer se encarga del reconocimiento de las palabras en el audio y envía la transcripción a la clase Result para que le muestre la información al usuario.

Tras probar con ejemplos sencillos de audios en inglés incluidos en la distribución, como un “hello world” o una secuencia de números, se pudo comprobar que las salidas que se obtenían eran correctas.

Adecuación a un modelo acústico en español

Cmu-Sphinx permite al usuario la creación de propios modelos acústicos, así como la descarga de algunos que ya están creados. Para la creación de un modelo acústico es necesario disponer de una buena base de datos de voces. En la página explican que dependiendo del tipo de sistema de reconocimiento que queramos se necesitan más o menos horas de audios para entrenar. Por ejemplo se necesitaría una hora de grabación para reconocimiento de comandos simples de un único hablante y cinco horas de grabación de 200 personas para el reconocimiento de comandos de varios hablantes. Para un sistema de dictado de un lenguaje concreto serían necesarias 10 horas para una sola persona, y si lo queremos para varias personas necesitamos 50 horas de grabación de 200 hablantes.

Además de las grabaciones también son necesarias las transcripciones en texto de cada audio, para que en el entrenamiento consiga asociar las señales de voz con la secuencia de palabras escritas en el fichero. Además también debe disponer de un diccionario que contenga la lista de palabras que puede reconocer, para que pueda mapear los sonidos reconocidos.

La estructura debe ser de esta manera:

- /etc
 - db.dic – *Diccionario fonético*
 - db.phone – *Fichero fonético*
 - db.lm.DMP – *Modelo del lenguaje*
 - db.filler – *List of fillers*
 - db_train.fileids – *Lista de los ficheros para entrenamiento*
 - db_train.transcription – *Transcripciones para entrenamiento*
 - db_test.fileids – *Lista de los ficheros para testeo*
 - db_test.transcription – *Transcripciones para testeo*
- /wav
 - ponente_1
 - file_1.wav
 - ponente_2
 - file_2.wav

Debido a que la complejidad de la creación de un modelo acústico es elevada, se optó por la obtención de un modelo acústico ya existente. En un primer momento se encontró información sobre un modelo acústico creado por los investigadores Varela, Cuayáhuitl y Nolazco-Flores [25]. Este modelo fue obtenido tras lograr ajustar el motor de reconocimiento CMU Sphinx-3 para su uso en el reconocimiento de español dialecto mexicano. Este modelo se encuentra albergado en los documentos de la propia página de Sphinx con el nombre “Mexican Spanish Broadcast News”.

Tras descargar el modelo acústico, se tuvo que realizar una serie de ajustes en el Cmu-sphinx para que el programa en vez de trabajar con los modelos por defecto del Wall Street cogiese el de reconocimiento para español de México. Los cambios más significativos fueron la inserción del modelo acústico dentro de los propios recursos del proyecto, así como la modificación del archivo “config.xml”. El fichero “config.xml” está dividido en secciones y no es necesario cambiar todas ellas pero sí las referentes a la configuración del diccionario, el modelo del lenguaje y el modelo acústico, porque las referencias deben actualizarse a los nuevos valores.

```
<!-- ***** -->
<!-- The Dictionary configuration -->
<!-- ***** -->
<component name="dictionary"
  type="edu.cmu.sphinx.linguist.dictionary.FastDictionary">
  <property name="dictionaryPath"
    value="./ es_MX_broadcast_cont_2500/etc/h4.dic"/>
  <property name="fillerPath"
    value="./ es_MX_broadcast_cont_2500/etc/filler.dic"/>
  <property name="addSilEndingPronunciation" value="false"/>
  <property name="wordReplacement" value="&lt;sil&gt;"/>
  <property name="unitManager" value="unitManager"/>
</component>
<!-- ***** -->
<!-- The Language Model configuration -->
<!-- ***** -->
<component name="trigramModel"
  type="edu.cmu.sphinx.linguist.language.ngram.SimpleNGramModel">
  <property name="location"
    value="./ es_MX_broadcast_cont_2500/etc/transcription.lm"/>
  <property name="logMath" value="logMath"/>

  <property name="dictionary" value="dictionary"/>
  <property name="maxDepth" value="3"/>
  <property name="unigramWeight" value=".7"/>
</component>

<!-- ***** -->
<!-- The acoustic model configuration -->
<!-- ***** -->
-->
<component name="sphinx3"
  type="edu.cmu.sphinx.linguist.acoustic.tiedstate.TiedStateAcousticModel">
  <property name="loader" value="sphinx3Loader"/>
  <property name="unitManager" value="unitManager"/>
  <property name="vectorLength" value="29"/>
  </component> <component name="sphinx3Loader"
  type="edu.cmu.sphinx.linguist.acoustic.tiedstate.Sphinx3Loader">
  <property name="logMath" value="logMath"/>
  <property name="unitManager" value="unitManager"/>
  <property
    name="location" value="./
es_MX_broadcast_cont_2500/model_parameters/ hub4_spanish_itesm.cd_cont_2500"/>
</component>
```

El reconocimiento que se obtuvo no fue para nada el esperado ya que no conseguía reconocer casi ninguna palabra. El gran problema residía en que la fonética del español de México es diferente a la de España, fijándonos en el diccionario observamos que la traducción a fonemas que se realizaba no era la adecuada para el castellano, ya que por ejemplo el sonido “ce-” o “ci-” no se contemplaba:

CERVEZA	S E R V E S A
CESAN	S E S A N
CESAR	S E S A R

Debido a los problemas fonéticos se decidió desistir con este modelo y buscar otro que fuese para el castellano. No tardamos en averiguar la existencia de un proyecto llamado [Voxforge](#) que tiene como objetivo el recoger voz que la gente dona para ser usada con herramientas de reconocimiento de voz Código Abierto (Open Source). Este proyecto ha conseguido obtener una amplia base de datos acústica y permite descargarnos grabaciones que han sido donadas por personas.



Name	Last modified	Size
Parent Directory		-
 AGOs-20100831-jdm.tgz	14-Sep-2010 06:42	432K
 AGOs-20100831-tmi.tgz	14-Sep-2010 06:42	443K
 ARI-20100216-hds.tgz	16-Feb-2010 22:54	360K
 Adrian-20120203-lit.tgz	04-Jun-2012 16:02	508K
 AdrianRamirezEscalante-20111202-pgl.tgz	04-Jun-2012 16:02	520K

Ilustración 22. Voxforge

Estas grabaciones contienen información sobre la persona que la ha realizado, como su sexo, edad, lenguaje y pronunciación dialéctica, además de las frases que ha leído. Es muy útil en el caso de que se decida crear un modelo acústico, pero el propio proyecto ha creado uno para el castellano, aunque estaba todavía en una versión beta y en la página siguen pidiendo donaciones.

Tras descargar el modelo acústico voxforge-es.01 se creó un nuevo proyecto con las clases y librerías necesarias del Sphinx y se incluyó en él. Para probar el funcionamiento se grabaron varios archivos de audios con el micrófono del ordenador en formato “.wav”, que es el tipo de formato que se utilizan en los ejemplos de Sphinx. No se obtuvo ninguna transcripción ya que el programa no conseguía mapear las palabras que se decían con el diccionario. Al principio no se consiguió entender dónde podía residir el problema, hasta que se descubrió que si en las grabaciones se separaba muy bien una palabra de otra, marcando bien el silencio entre ellas se lograba obtener transcripciones de los audios. Las pruebas que se realizaron se detallan a continuación, para ellas se utilizaron audios grabados a 8000Hz.

Frase 1	Resultado
Hola esto es una prueba	cola
	unas
	seda

Tabla 192. Sphinx - transcripción primera prueba



Frase 2	Resultado
Ayer compré un kilo de tomates	frío de camareros

Tabla 193. Sphinx - transcripción segunda prueba

Frase 3	Resultado
España tiene un buen clima	puedes quién a

Tabla 194. Sphinx - transcripción tercera prueba

Frase 4	Resultado
Todo estaba lo mismo que una hora antes	vino mismo té una corta antes

Tabla 195. Sphinx - transcripción cuarta prueba

Frase 5	Resultado
Las personas cercanas a él fumaban silenciosas y seguían sus conversaciones con lentitud	plantas a té fumaban silenciosas y seguían sus conversaciones con y lentitud

Tabla 196. Sphinx - transcripción quinta prueba

Frase 6	Resultado
Del fondo del salón llegaban confundidos con risas de mujeres y choques de bandejas	salón estaban confundidos con derecha tú mujeres y choque de bandejas

Tabla 197. Sphinx - transcripción sexta prueba

Frase 7	Resultado
Unas líneas negras y oblicuas semejantes a cuerdas	y oblicuas semejantes a cuerdas

Tabla 198. Sphinx - transcripción séptima prueba

Los resultados obtenidos no son del todo malos ya que algunas frases consigue transcribirlas de una manera bastante correcta. Sin embargo, el principal problema reside en que la manera en la que se debe hablar para que funcione, no es la manera natural. Esto resulta un gran inconveniente para otras partes futuras de la implementación, como es el reconocimiento de emociones de la voz, debido a que la velocidad en la que hablamos es diferente cuando estamos tristes o enfadados. En consecuencia, con esta herramienta de transcripción perderíamos este factor al tener que separar todas las palabras y no poder pronunciarlas a diferentes velocidades, por lo que se debió encontrar otra solución factible.

Reconocedor de Google

Tras descartar a Sphinx, se tuvo que volver a investigar sobre otros métodos para el reconocimiento automático de voz. Se observó que Youtube había incorporado una nueva funcionalidad para la transcripción de audio de los videos y generar subtítulos, gracias a la tecnología de Google Automatic Speech Recognition (ASR) incluida en Google Voice y la generación de subtítulos ya existente en Youtube. Los subtítulos se despliegan automáticamente cuando escuchamos el audio, por lo que el timing es casi perfecto.



Ilustración 23. Subtítulos en YouTube

El botón “CC” (Closed Captions) permite activar y desactivar los subtítulos, indicándonos el idioma. Normalmente suele estar desactivados ya que actualmente no hay muchos videos con subtítulos. Además permite traducir el texto a otros idiomas, lo cual puede resultar muy útil si queremos aprender nuevos lenguajes o enterarnos de lo que dicen en el video. El objetivo principal por el que se desarrolló esta herramienta, es para ayudar a los de millones de personas que, en todo el mundo, padecen algún tipo de deficiencia auditiva, aunque tiene otras ventajas añadidas.

Google Voice es un servicio de telecomunicaciones que salió en el 2009 y permite realizar llamadas entre ordenadores o entre PCs y teléfonos y envío de SMS. La parte que nos interesa de este servicio, es una herramienta que lanzaron para transcribir automáticamente mails de voz a texto, por lo que permite leer tus mensajes además de escucharlos y así poder enterarte aunque estés en un ambiente ruidoso. Esta funcionalidad también está presente en otros servicios de Google como Google Search, en el cual se pueden realizar búsquedas conectando un micrófono al ordenador y expresando lo que se desea encontrar. Las transcripciones y búsquedas que se pueden realizar con esta tecnología son muy potentes tal y como se ha visto con YouTube y Google Search, por lo que se decidió obtener información y ver cómo funciona a más bajo nivel.

Google posee unos servidores que se encargan de esta tarea, en ellos se realiza un análisis complejo de reconocimiento de voz de un archivo, y se identifica tanto la frase más probable acorde al audio que se ha grabado, como una serie de alternativas posibles. Para obtener una transcripción únicamente debemos contactar con el servidor y enviarle el audio en formato “.flac”. Tras unos instantes nos devuelve el resultado, que

tras el análisis resulta ser el más probable, junto con un porcentaje de la fiabilidad de la respuesta obtenida.

Si se trabaja en Linux resulta bastante fácil enviar y recibir datos del servicio de Google, ya que simplemente con una llamada con el comando `wget` al servidor encargado de esta labor podemos realizar el proceso. En nuestro caso, la aplicación iba a ser realizada para que funcionase en Windows sobre el lenguaje de programación Java y esto complicó un poco más la tarea.

Lo primero que se buscó fue una forma para pasar al formato “.flac” los archivos de audios. FLAC (Free Lossless Audio Codec) es un codec, que permite que el audio digital se comprima sin que se produzcan pérdidas, de tal manera que el tamaño del archivo se reduce sin riesgos. Los audios que generamos en el dispositivo móvil son “.wav”, por lo que era indispensable una conversión entre formatos. Para ello, se utilizó SoX (Sound eXchange), una utilidad multiplataforma que funciona a través de línea de comandos y puede convertir varios formatos de archivos de audio de ordenador a otros formatos. También se puede utilizar para aplicar varios efectos a estos archivos de sonido, y, como ventaja añadida, SoX puede reproducir y grabar archivos de audio en la mayoría de las plataformas. La elección de esta utilidad resultó ser un gran acierto ya que la versatilidad que posee es enorme y fue de gran ayuda para la realización de las pruebas, debido a que permite cambiar la frecuencia de las grabaciones o pasar de mono a estéreo con un simple comando. Además, la propia herramienta trae incorporado un archivo ejecutable para la transferencia y descarga de ficheros a través del comando `wget`, que es justo lo que necesitábamos para la comunicación con el servidor encargado de la transcripción automática.

Una vez solucionado el tema de la conversión a FLAC, se pensó en la manera de conectar el programa en Java con la herramienta SoX y las llamadas al servidor de Google y se decidió utilizar ficheros batch para esta labor. Un fichero batch es un archivo de procesamiento por lotes: se trata de archivos de texto sin formato, guardados con la extensión *.bat que contienen un conjunto de comandos DOS. Cuando se ejecuta este archivo bat, los comandos contenidos son ejecutados en grupo, de forma secuencial, permitiendo automatizar diversas tareas. Además, al tener que hacerse sobre un programa Java y dado que los comandos del fichero bat son propios del sistema operativo, fue necesario utilizar clases específicas que permitiesen acceder al propio sistema. El fichero creado empieza realizando llamadas a SoX para la conversión de formatos, y una vez que se tiene el archivo en .flac se transfiere al servidor de Google que a su vez devuelve la respuesta de la transcripción.

Para probar inicialmente el funcionamiento se grabaron una serie de audios con el micrófono del ordenador a 8000Hz, pero no se obtuvieron los resultados esperados ya que en muchos casos la respuesta estaba vacía y al tratarse de un servicio de Google,

utilizado en páginas tan importantes como YouTube, era más inesperado aún. Debido a que el error no podía estar en la parte de los servidores, se decidió probar ajustando las grabaciones con otros parámetros y ver si de ésta manera el producto obtenido era diferente. Los ensayos consistieron en probar diferentes combinaciones de frecuencias de muestreo, y que los bits por muestra fuesen 8 o 16. Los mejores resultados se obtuvieron con archivos grabados a una frecuencia de 16000Hz y a 16 bits por muestra.

A continuación se exponen una serie de pruebas que muestran las transcripciones obtenidas, para ello se utilizaron las mismas frases que en CmuSphinx pero expresadas de manera natural, es decir, sin espacios marcados entre cada una de las palabras. La respuesta que se obtiene del servidor de Google posee este formato:

```
{"status":0,"id":"8d12654d9e8a7062ac59a8696c355d43-1","hypotheses":[{"utterance":"unas líneas negras y oblicua semejantes a cuerda","confidence":0.6634928}]}
```

Para simplificarlo, únicamente nos quedamos con la parte correspondiente al campo utterance porque es lo que puede resultar útil al usuario final.

Frase 1	Resultado
Hola esto es una prueba	hola esto es una prueba

Tabla 199. ASR Google - transcripción primera prueba

Frase 2	Resultado
Ayer compré un kilo de tomates	ayer compre un kilo de tomates

Tabla 200. ASR Google - transcripción segunda prueba

Frase 3	Resultado
España tiene un buen clima	españa tiene un buen clima

Tabla 201. ASR Google - transcripción tercera prueba

Frase 4	Resultado
Todo estaba lo mismo que una hora antes	todo estaba lo mismo que una hora antes

Tabla 202. ASR Google - transcripción cuarta prueba

Frase 5	Resultado
Las personas cercanas a él fumaban silenciosas y seguían sus conversaciones con lentitud	las personas cercanas a él fumaban silenciosas poseía sus conversaciones con lentitud

Tabla 203. ASR Google - transcripción quinta prueba

Frase 6	Resultado
Del fondo del salón llegaban confundidos con risas de mujeres y choques de bandejas	el fondo del salón llegaban confundidos con risas de mujeres y choque de bandejas

Tabla 204. ASR Google - transcripción sexta prueba

Frase 7	Resultado
Unas líneas negras y oblicuas semejantes a cuerdas	unas líneas negras y oblicua semejantes a cuerda

Tabla 205. ASR Google - transcripción séptima prueba

Como se puede observar, se han obtenido unas transcripciones muy fidedignas al audio original, con unos resultados mejores que los alcanzados con CmuSphinx, es por ello que nos decantamos por este sistema de transcripción.

5.2.5 Módulo de análisis de emociones

En este apartado, se va a explicar el proceso que se ha seguido para construir el módulo software encargado de la obtención de las emociones presentes en archivos de audios.

Lo primero que se tuvo que hacer fue una intensa documentación sobre investigaciones y proyectos ya realizados [26-31] y los métodos y técnicas utilizadas para este proceso. Para realizar un análisis acústico se debe extraer ciertas características que puedan ser estudiadas posteriormente. Este conjunto no es preestablecido y cada trabajo utiliza un conjunto diferente, por lo que hemos decidido quedarnos con las características que nos han parecido más importantes. A continuación se explica de una manera breve cada una de ellas, para aportar al lector toda la información necesaria.

Características fundamentales

1. Pitch

El pitch es la frecuencia fundamental a la que las cuerdas vocales vibran, también llamada frecuencia fundamental o F0. Se considera que las características de la frecuencia fundamental son una de las principales portadoras de la información sobre las emociones. El valor medio del pitch expresa el nivel de excitación del locutor, una media elevada de F0 indica un mayor grado de excitación.

2. Rango del pitch

El rango del pitch es la distancia entre el valor máximo y mínimo de la frecuencia fundamental. Suele utilizarse en análisis acústico porque refleja el grado de exaltación del locutor dependiendo de lo extenso que sea.

3. SPL

El nivel de presión sonora determina la intensidad del sonido que genera una presión sonora instantánea, es decir, del sonido que alcanza a una persona en un

momento dado. Su unidad es el decibelio (dB) y varía entre los 0 dB del umbral de audición y los 140 dB del umbral de dolor.

4. Rango del spl

El rango del SPL mide la amplitud de los valores de la muestra y es la distancia entre el valor máximo y mínimo de la intensidad.

5. Velocidad

Se calcula a través del número de palabras expresadas por minuto. Una persona exaltada aumenta su velocidad de locución mientras que una persona disgustada disminuye el ritmo.

6. Energía

La función de energía de una señal representa la energía disipada por una resistencia de 1 Ohm cuando se le aplica un voltaje $x(t)$. En una señal continua, la energía total en el intervalo de tiempo t_1 a t_2 está definida como:

$$E = \int_{t_1}^{t_2} |x(t)|^2 dt$$

Para el caso de las señales discretas donde N es el número de muestras de la señal, la energía se define por:

$$E = \sum_{m=0}^{N-1} x(m)^2$$

7. Media

La media aritmética de un conjunto finito de números es el valor característico de una serie de datos cuantitativos objeto de estudio que parte del principio de la esperanza matemática o valor esperado, se obtiene a partir de la suma de todos sus valores dividida entre el número de sumandos

8. Varianza

Mide la distancia existente entre los valores de la serie y la media. Se calcula como sumatorio de las diferencias al cuadrado entre cada valor y la media, multiplicadas por el número de veces que se ha repetido cada valor. El sumatorio obtenido se divide por el tamaño de la muestra.

$$S_x^2 = (\sum (x_i - x_m)^2 * n_i) / n$$

La varianza siempre será mayor que cero. Mientras más se aproxima a cero, más concentrados están los valores de la serie alrededor de la media. Por el contrario, mientras mayor sea la varianza, más dispersos están.

Numerosas investigaciones realizan la clasificación de emociones a partir de la emoción neutra, ya que de esta manera puedes realizar un sistema independiente del usuario que lo utilice. Nosotros nos hemos centrado en realizar el módulo del sistema con estas características, ya que creemos que es más flexible que realizarlo sin comparar con el archivo neutro, debido a que cada persona habla de una manera distinta con unas características de la voz diferentes y resulta más sencillo obtener un conjunto de reglas cuando partes de unos valores personales.

Extracción de las características fundamentales

Para la extracción de los parámetros acústicos se utilizó una librería de código libre llamada Musicg, creada por Jacquet Wong, además de creación de código propio. La librería Musicg se puede encontrar en GoogleCode y para su utilización tuvimos que instalarnos en la herramienta Eclipse el sistema de control de versiones Subversion y bajar el código a través de él.



Ilustración 24. Musicg en svn

Musicg es una biblioteca dedicada al análisis acústico de audios, con el fin de obtener, tanto a alto nivel como a bajo nivel, las características fundamentales de los audios. Esta API permite operar con los datos del audio ya que se puede leer y cortar fácilmente desde un `InputStream`. También proporciona herramientas para el procesamiento digital de señales, y dispone de utilidades que permite renderizar la señal en el tiempo o mostrar su espectrograma. El código está escrito en Java y es compatible con la plataforma Android.

Las principales funcionalidades de la aplicación son:

- Silbato Api - Detectar si el audio de entrada es un silbato.
- Leer encabezados PCM-WAVE.

- Leer los datos de un archivo de audio.
- Recortar los datos de un archivo de audio.
- Guardar un archivo de audio editado.
- Trabajar con los datos de audio en el dominio del tiempo.
- Trabajar con los datos de audio en el dominio de la frecuencia.
- Representar la imagen de forma de onda de audio.
- Renderizar la imagen del espectrograma de audio.

La librería está compuesta por 23 clases, de las cuales utilizaremos en nuestro proyecto 9 de ellas, que son las que vamos a detallar.

MainTest

Clase principal de la aplicación, es la encargada de recibir el fichero de audio que se va a analizar y gestionar la comunicación entre el resto de clases para que se pueda realizar el procesamiento del archivo.

WaveHeader

Clase encargada de determinar si la cabecera del fichero wav es correcta o no.

Un fichero WAVE se compone de una cabecera formada por el descriptor RIFF seguido de un bloque de datos que consiste en dos subcampos, el campo "fmt" que especifica el formato de los datos y el campo de datos ("data"). A este formato se le conoce como Forma Canónica y puede verse en la ilustración siguiente.

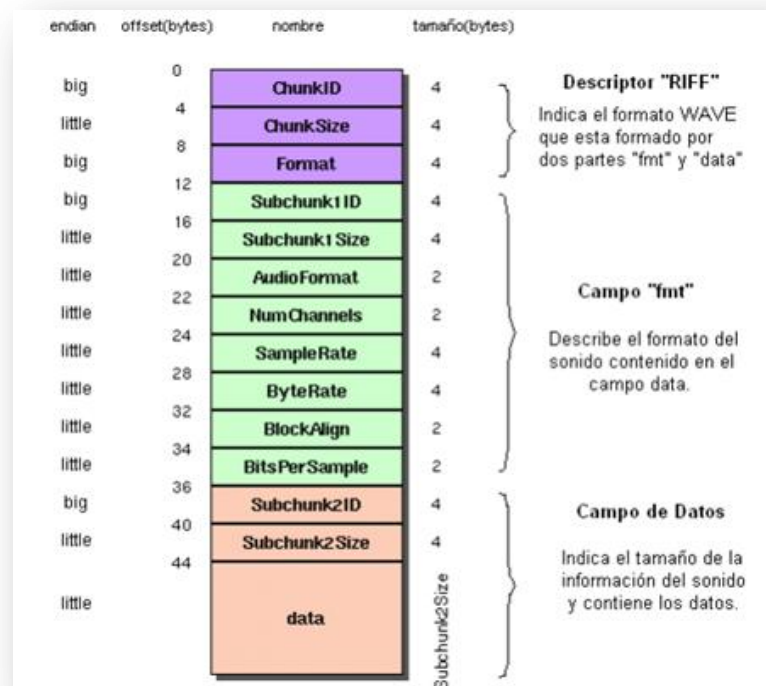


Ilustración 25. Forma canónica del formato de fichero WAVE

WaveInputStream

Es la clase encargada de leer los datos del InputStream de entrada, a excepción de los primeros 44 bytes de los datos, que son los pertenecientes a la cabecera del archivo wav.

Wave

Es la clase que gestiona la lectura de los ficheros WAVE, para ello hace uso de las clase WaveHeader y WaveInputStrem. Proporciona información relevante del audio como:

- **ChunkID:** Contiene las letras “RIFF” en formato ASCII.
- **ChunkSize:** Tamaño del fichero entero en bytes menos los 8 bytes de los campos ChunkID y ChunkSize.
- **Format:** Contiene el string “WAVE”. El formato “WAVE” consiste en dos subporciones de datos: “fmt” y “data”: la parte “fmt” describe el formato de los datos que forman el sonido.
- **SubChunk1ID:** Contiene el string “fmt”.
- **SubChunk1Size:** 16 para PCM.
- **AudioFormat:** Define cómo se encuentra codificado el audio en el fichero. Su valor debe ser siempre uno, que significa Pulse Code Modulation (PCM).
- **NumChannels:** Número de canales de audio presentes en el fichero WAV. Para sonidos mono hay un canal, para sonidos estéreo hay dos canales.
- **SampleRate:** El ratio de muestreo se expresa en muestras por segundo o Hertzios. Por ejemplo: 44100 Hz. nos indica que en un segundo se tomaron 44100 muestras de la señal analógica de audio para crear el audio digital correspondiente. Un audio tendrá más calidad cuanto mayor sea su tasa de muestreo.
- **ByteRate:** El número medio de bytes por segundo que un reproductor debe procesar para reproducir audio en tiempo real. Se puede calcular como $\text{SampleRate} * \text{NumChannels} * \text{BitsPerSample}/8$.
- **BlockAlign:** Este valor indica cuantos bytes se obtienen en la salida por unidad de tiempo. En PCM es el número de bytes por muestra multiplicado por el número de canales.
- **BitsPerSample:** Define el número de bits por amplitud de audio muestreada. Si se usa PCM no hay espacio para parámetros extras.
- **SubChunk2ID:** Contiene el string “data”.
- **SubChunk2Size** Número de bytes de los datos. Se calcula como $\text{NumSamples} * \text{NumChannels} * \text{BitsPeSample}/8$.
- **Data:** Datos de sonido
- Además, posee funciones que permiten recortar los datos de un archivo de audio y guardar el archivo de audio una vez editado.



TimeDomainRepresentation

Se encarga de la representación en el dominio del tiempo y ejerce de clase padre de `FrequencyTimeDomainRepresentation` y de `AmplitudeTimeDomainRepresentation`. Su principal función es obtener las amplitudes de la señal que se le pase.

FrequencyTimeDomainRepresentation

Esta clase es la encargada de manejar los datos del audio en el dominio de la frecuencia y de construir el espectrograma de la señal.

AmplitudeTimeDomainRepresentation

Esta clase permite manejar los datos de la señal en el dominio de amplitud-tiempo. Su principal función es la de normalizar las amplitudes y devolvérselas al usuario para que pueda trabajar con ellas.

FastFouriourTransform

Clase encargada de la realización de la Transformada Rápida de Fourier.

WindowsFunction

Esta clase permite generar y usar diferentes tipos de ventanas: rectangular, Bartlett, Hanning, Hamming y Blackman.

Aparte de la utilización de las clases anteriormente explicadas, tuvimos que implementar una serie de funcionalidades extras que nos permitiesen obtener el resto de las características fundamentales, además del algoritmo de clasificación de emociones y la gestión del proceso completo.

Una de las características que necesitábamos era el SPL, tanto el cálculo del mismo, como la obtención del valor máximo y mínimo, ya que eran necesarios para el cálculo del rango del SPL. Para ello, se realizó un algoritmo que recorriese cada una de las muestras de la señal obteniendo la frecuencia y la amplitud, necesarias para hallar las presiones intermedias mediante las fórmulas:

$$\begin{aligned} \text{Intensidad} &= 2 * P^2 * \text{frecuencia} * \text{amplitud} \\ P1 &= \sqrt{2 * Z * \text{Intensidad}} \\ \text{Spl} &= 20 * \log(P1 / P); \end{aligned}$$

Z es la impedancia acústica y su valor es 406^2 (N·s/m³). P es la presión de referencia del umbral de audición y su valor es de 20 micro Pascales.

El valor de SPL que aparece anteriormente se almacena para hallar el SPL total de la señal. Una vez que se ha terminado de recorrer las muestras hallamos el SPL medio dividiendo el SPL total entre el cantidad de valores recorridos (muestras * n°

amplitudes). Los valores mínimos y máximos se obtienen durante el recorrido comparando el valor actual del nivel de presión sonora con sus valores anteriores.

Análisis acústico

En este apartado vamos a describir el proceso seguido para implementar el reconocedor de emociones. Para poder realizar el estudio tuvimos que crear un corpus de voces con el que poder realizar análisis y obtener resultados.

Para la construcción del corpus del habla contamos con las grabaciones de once actores, siete hombres y cuatro mujeres. Los actores tuvieron que interpretar la frase “esta es la prueba número uno” con las siguientes emociones primarias: alegría, enfado, tristeza, miedo y neutro. Decidimos quedarnos con todas las grabaciones para no perder ninguna muestra, ascendido el número de muestras a un total de 86.

Para la grabación de los audios se utilizó el teléfono móvil Samsung Galaxy S1 i9000 y se realizó en un aula de la escuela de cine Septima Ars sin insonorización acústica.

Estudio de las características prosódicas

Una vez creado el corpus, se realizó un estudio de las características prosódicas de los audios para intentar encontrar un patrón que nos indicase cómo diferenciar unas emociones de otras. La prosodia es una rama de la lingüística que analiza y representa formalmente aquellos elementos de la expresión oral, tales como el acento, los tonos y la entonación.

La técnica que se ha seguido en nuestro sistema, ya ha sido utilizada por algunos investigadores con mucho éxito para el reconocimiento de emociones a partir de la información vocal no verbal del habla. Esta técnica consiste en la evaluación de las emociones primarias de una persona con su propia emoción neutra. Normalmente se aplica para que un oyente escuche los audios y diferencie unas e mociones de otras. En nuestro sistema, en vez de disponer de un oyente se realizará automáticamente mediante algoritmos creados con este propósito.

Para ello, nuestro programa crea un fichero de texto por cada uno de los audios, en los que almacenará los valores obtenidos para cada una de las variables que se estudian. Tras crear los ficheros, se realizó un estudio, para cada una de las personas, de los resultados medios obtenidos para la emoción neutra con el resto de emociones. Estos ficheros, junto con los obtenidos por medio de nuestro programa, pueden observarse en el capítulo 11 “ANEXOS”, en el apartado “Anexo 1”, de la presente memoria.

En las siguientes tablas mostraremos un resumen con los valores de los once actores, obtenidos tras comparar los valores medios neutro con los valores medios del resto de emociones.

Nota al lector: Los valores finales están calculados con todos los decimales de los números, sin embargo en las tablas se ha utilizado únicamente los tres primeros decimales redondeados para poder presentar toda la información, por lo que si algún decimal es diferente es por esta razón.

ALEGRÍA	F0	Rango pitch	SPL	Rango SPL	Velocidad	Energía	Media	Varianza
Actor 1	1	1,109	1,000	1,028	0,519	1,093	1,088	1,094
Actor 2	0,829	1,026	1,002	1,041	2,379	0,948	1,046	1,314
Actor 3	0,942	1,084	1,004	0,955	1,166	1,0928	1,088	1,094
Actor 4	4,816	1,099	0,998	1,161	0,552	0,767	0,955	1,333
Actor 5	1	1,212	0,994	1,050	0,715	1,111	0,886	1,667
Actor 6	0,599	0,908	1,002	0,958	1,013	1,073	1,039	0,778
Actor 7	0,783	0,913	1,001	0,959	1,674	1,188	1,0145	1
Actor 8	0,837	1,046	0,998	1,040	0,445	1,031	0,957	1,52
Actor 9	2,904	0,838	0,999	0,953	0,355	2,761	0,993	0,571
Actor 10	1,430	1,1373	0,994	1,0198	0,515	0,312	0,886	2
Actor 11	0,872	1,116	0,998	1,0299	2,281	1,265	0,970	1,538

Tabla 206. Comparativa de los valores neutro y alegría

ENFADO	F0	Rango pitch	SPL	Rango SPL	Velocidad	Energía	Media	Varianza
Actor 1	1,029	1,112	1,001	1,025	0,202	1,007	1,027	1,591
Actor 2	1	1,11	1,001	0,999	0,653	0,882	1,026	1,343
Actor 3	1	1,153	1,005	0,968	1,336	1,074	1,109	1,097
Actor 4	4,559	0,955	0,999	0,994	1,107	0,806	0,989	1,071
Actor 5	0,934	1,207	0,995	1,041	1	0,811	0,892	1,889
Actor 6	0,62	0,908	1,001	0,958	0,874	1,243	1,0256	0,75
Actor 7	0,783	0,908	1,002	0,939	2,013	0,8956	1,042	1,056
Actor 8	0,783	0,978	1,000	1,006	1,009	1,195	1,011	1,16
Actor 9	2,503	0,867	0,997	0,983	1,529	1,168	0,942	1,071
Actor 10	1	1,119	0,998	1,015	0,699	0,527	0,954	1,25
Actor 11	1,248	0,956	1,004	1,007	3,199	1,469	1,075	1,692

Tabla 207. Comparativa de los valores neutro y enfado

TRISTEZA	F0	Rango pitch	SPL	Rango SPL	Velocidad	Energía	Media	Varianza
Actor 1	0,852	1,227	0,998	1,025	0,653	1,199	0,954	1,273
Actor 2	0,536	0,931	1,004	1,041	0,796	1,186	1,083	0,829

Actor 3	0,471	1,288	0,999	0,984	0,497	0,982	0,989	1,067
Actor 4	1,779	1,159	0,996	1,032	0,346	1,006	0,912	0,893
Actor 5	0,868	1,090	0,999	1,024	0,972	1,001	0,986	1,111
Actor 6	0,84	0,894	1,002	0,948	1,283	1,835	1,037	0,556
Actor 7	0,456	0,920	1,002	0,954	0,440	1,222	1,038	0,833
Actor 8	1,189	0,934	1,001	0,997	0,381	2,044	1,024	1
Actor 9	1,5	0,916	1,000	0,939	0,597	1,677	1	0,786
Actor 10	1,071	1,058	0,998	1,080	0,628	1,453	0,952	1
Actor 11	1,5	0,835	1,004	1,007	0,459	2,564	1,098	0,538

Tabla 208. Comparativa de los valores neutro y tristeza

MIEDO	F0	Rango pitch	SPL	Rango SPL	Velocidad	Energía	Media	Varianza
Actor 1	1,118	1,227	0,999	1,030	0,482	1,112	0,995	1,273
Actor 2	1,206	1,190	0,995	1,052	0,468	1,106	0,89	1,25
Actor 3	1	0,969	1,006	0,956	0,497	1,165	1,137	0,867
Actor 4	3,331	0,966	0,998	1,011	0,588	0,775	0,915	1,643
Actor 5	0,267	1,019	0,999	1,000	1,668	2,234	0,969	0,778
Actor 6	0,94	0,835	1,004	0,954	0,913	1,305	1,074	0,444
Actor 7	0,304	0,82	1,003	0,934	0,471	1,217	1,077	0,722
Actor 8	1,054	0,903	1,001	0,998	1,321	1,193	1,036	1,2
Actor 9	1,603	0,922	0,997	0,959	1,083	0,817	0,936	0,929
Actor 10	1,035	1,0363	1,001	1,001	0,456	1,348	1,025	1,188
Actor 11	1	0,759	1,008	0,961	0,410	3,337	1,185	0,385

Tabla 209. Comparativa de los valores neutro y miedo

Medias

	Alegría	Enfado	Tristeza	Miedo
F0	1,45571058	1,40535376	1,00561387	1,16888879
Rango del pitch	1,04447784	1,02480929	1,02298221	0,96788352
SPL	0,99914438	1,00029663	1,00031379	1,00099609
Rango SPL	1,01778103	0,9941832	1,0029844	0,98690632
Velocidad	1,05565891	1,2384114	0,64101828	0,75972957
Energía	1,14934556	1,00700469	1,47000459	1,41901408
Valor medio	0,99301804	1,00836691	1,00656429	1,02176788
Varianza	1,26449578	1,27001361	0,89863621	0,97067112

Tabla 210. Medias de todos los actores por emociones

De estos análisis podemos sacar las siguientes conclusiones:

- Alegría: posee la frecuencia fundamental más alta, además de una gran varianza.
- Enfado: su energía es la menor de las cuatro emociones. Tiene un pitch muy alto y la varianza y la velocidad son las mayores.
- Tristeza: se caracteriza por un bajo pitch, su velocidad es lenta y tiene poca varianza. Su energía es la mayor de las cuatro emociones.
- Miedo: tiene el menor rango de pitch y de SPL, y su energía es elevada.

Si pensamos en la manera en la que expresamos una frase estando enfadados o estando tristes, lo primero que puede venirnos a la mente es que cuando estamos enfadados se produce un aumento en el nivel de excitación en la locución, y el pitch o frecuencia fundamental es un gran indicador de esta medida. Observando la tabla anterior puede verse que en media el pitch es mucho mayor en la emoción de enfado que en la de tristeza, lo cual es razonable y concuerda con la realidad. También es normal que si dialogamos enfadados la velocidad crezca respecto al habla normal, mientras que cuando estamos tristes hablamos más pausados y despacio. En la tabla esta medida describe un aumento del 23% para el enfado en comparación con una locución en estado neutro, mientras que la tristeza posee un 36% menos de velocidad.

Al igual que el enfado, la alegría se caracteriza por un aumento en el nivel de excitación del locutor, un aumento en la velocidad de locución y una gran fluctuación en sus valores, representados por la varianza de la señal.

La emoción que más dista de los resultados que a priori se pensaba que se podían obtener es el miedo. En un principio se pensó que el miedo debería tener uno de los mayores tonos medios, un gran rango del pitch y una locución de velocidad rápida. Sin embargo, escuchando las grabaciones de una manera detenida, los actores expresaban el miedo con un tono cercano al de la tristeza y un final de frase similar a la entonación de una pregunta interrogativa. Es por ello, que los valores que se obtienen son contradictorios a los esperados ya que el tono es medio y el rango estrecho.

Una vez analizados los resultados, el siguiente paso fue programar un algoritmo en Java cuya finalidad sería realizar un análisis acústico comparando las características de un archivo de audio con emoción neutra y otro con cualquiera de las otras 4 emociones primarias, a través de reglas basándonos en las conclusiones obtenidas anteriormente. Para ello fue necesario construir clases que obtuviesen los atributos de ambos archivos, y en base a unas reglas preestablecidas obtener la emoción primaria que más se ajuste. A continuación se explicará el proceso para cotejar las características fundamentales:

Obtención de la emoción predominante

Esta tarea se encarga de obtener la emoción primaria en base a los datos del archivo de audio considerado neutro y del archivo de audio que se desea evaluar, realizando una evaluación en base a unas reglas preestablecidas para obtener la emoción primaria que más se ajuste.

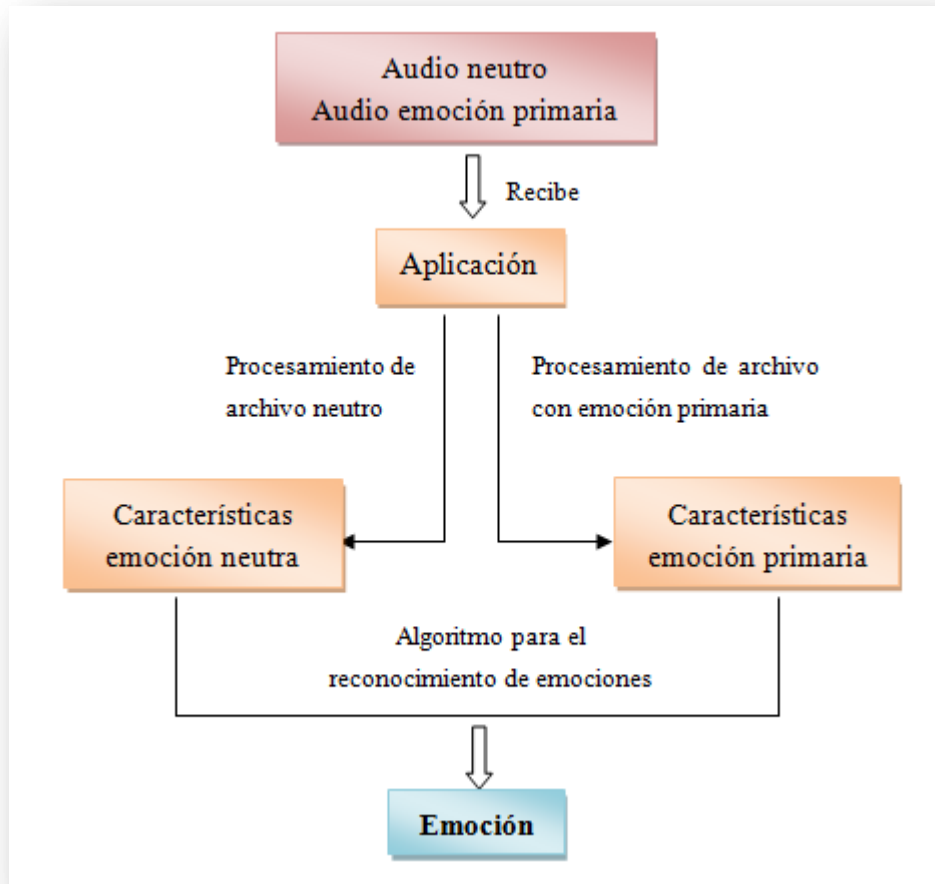


Ilustración 26. Flujo del reconocimiento de emociones

Tal y como aparece en la anterior ilustración, lo primero que se reciben son los ficheros WAVE por parámetro. Antes de analizarlos se comprueba que sus cabeceras sean correctas, por si ha ocurrido algún problema en la grabación, y a continuación se obtienen los datos que nos interesan, primero en el dominio del tiempo para obtener la duración del audio y después en el dominio de la frecuencia se obtienen el resto de valores.

Dominio del tiempo

La duración del audio es necesaria para hallar la velocidad a la que habla una persona, debido a que se estima como el número de palabras que expresa por minuto. Se computa como:

$$\text{duración} = (n^{\circ} \text{ muestras}) / (n^{\circ} \text{ de muestras por segundo})$$

Dominio de la frecuencia

Una señal puede representarse o ser descrita por los valores que toma en cada instante de tiempo (dominio del tiempo). Pero también puede representarse o ser descrita por una suma de senoides de distintas amplitudes, fases y frecuencias (dominio de la frecuencia). El conjunto de herramientas matemáticas que nos permiten pasar de un dominio a otro son las Transformadas de Fourier y es lo que aplicamos a una señal para trabajar en los distintos dominios.

Una vez que tenemos la señal en frecuencia, se realiza un filtrado de la misma en base a los rangos de la frecuencia de la voz humana, es decir entre 82 y 1056 Hz.

Gama de frecuencias fundamentales		
Voz	Extensión (Hz)	Tesitura
Soprano	247-1056	SI ₃ -DO ₆
Mezzosoprano	220-900	LA ₃ -SI ₅
Contralto	176-840	FA ₃ -LA ₅
Tenor	132-528	DO ₃ -DO ₅
Baritono	110-440	LA ₂ -LA ₄
Bajo	82-396	MI ₂ -SOL ₄

Ilustración 27. Frecuencias de la voz humana

Se decidió filtrar los archivos para quedarnos únicamente con los valores que están dentro de ese rango de frecuencias, ya que de esta manera reducimos el número de muestras y agilizamos los procesos, además así nos aseguramos un poco más que los datos se corresponden con valores pertenecientes a la voz humana. Para realizar esta tarea lo primero que se necesita tener es el espectrograma de la señal, ya que es justamente a lo que vamos a aplicar nuestro filtro. El espectro de frecuencia de una señal, caracteriza qué distribución de amplitudes presenta cada una de las frecuencias existentes en la señal. Como únicamente queremos quedarnos con las que estén entre 82 y 1056 Hz, se tuvo que hallar la posición del array del espectrograma correspondiente a cada una de estas dos frecuencias para desechar las que se estén por debajo de la inferior y por encima de la superior. Las posiciones las obtuvimos así:

$$\text{Posición inferior} = 82 * (\text{frecuencia de muestreo}) / (2 * \text{frecuencia de grabación})$$

$$\text{Posición superior} = 1056 * (\text{frecuencia de muestreo}) / (2 * \text{frecuencia de grabación})$$

Una vez que se tiene el espectrograma filtrado, se obtienen las características fundamentales en el dominio de la frecuencia para ambos archivos de audios y se aplica el algoritmo de reconocimiento de emociones, obteniendo como resultado la emoción. La emoción resultante, junto con las características fundamentales y la transcripción, se almacenará en un archivo de resultados que posteriormente será enviado al usuario para que pueda conocer los datos.

Para probar el algoritmo se creó una base de datos de test, compuesta por los archivos de audios de otras personas distintas a los de entrenamiento. Esta base de datos es menor que la de entrenamiento, ya que no se dispone de una gran cantidad de muestras. A continuación se exponen los resultados que se obtuvieron.

Resultados

Para la visualización de los resultados se utilizó una matriz de confusión, de cada una de las emociones respecto a cada uno de los audios con emoción neutra, para cada persona. La matriz de confusión consta de cuatro estados representados por:

- TP (True Positive): es la emoción deseada y ha acertado en el resultado obtenido.
- FP (False Positive): no es la emoción deseada y devuelve un resultado verdadero de que es esa emoción.
- FN (False Negative): debería salir la emoción deseada pero el resultado es otra emoción distinta.
- TN (True Negative): no es la emoción deseada y ha acertado en el resultado obtenido, ya que devuelve otra emoción.

Las matrices de confusión pueden verse en el capítulo 11 “ANEXOS”, en el apartado “Anexo 2”, de la memoria.

A continuación se presenta un resumen de los resultados, divididos en porcentajes de aciertos totales y aciertos por emociones.

Resultados totales	
Porcentaje de aciertos	39 de 71 muestras = 54,93%
Porcentajes de fallos	32 de 71 muestras = 45,07%

Tabla 211. Porcentaje de éxito de las pruebas

Resultados para la emoción Alegría	
Porcentaje de aciertos	6 de 17 muestras = 35,29%
Porcentajes de fallos	11 de 17 muestras = 64,71%

Tabla 212. Porcentaje de éxito de la emoción alegría

Resultados para la emoción Enfado	
Porcentaje de aciertos	11 de 17 muestras = 64,71%
Porcentajes de fallos	6 de 17 muestras = 35,29%

Tabla 213. Porcentaje de éxito de la emoción alegría

Resultados para la emoción Miedo	
Porcentaje de aciertos	6 de 17 muestras = 35,29%
Porcentajes de fallos	11 de 17 muestras = 64,71%

Tabla 214. Porcentaje de éxito de la emoción alegría

Resultados para la emoción Tristeza	
Porcentaje de aciertos	16 de 20 muestras = 80%
Porcentajes de fallos	4 de 20 muestras = 20%

Tabla 215. Porcentaje de éxito de la emoción tristeza

Como puede observarse, el sistema es capaz de reconocer la tristeza con un alto porcentaje de aciertos. Esto puede deberse a que las características que posee son más significativas que la del resto de emociones, ya que la baja frecuencia y la baja velocidad permite la diferenciación. El enfado es la siguiente emoción mejor reconocida con un 64,71% de aciertos, seguida por la alegría y el miedo. La alegría no ha obtenido un buen resultado debido a que algunos archivos que debían ser clasificados como alegría, fueron clasificados como enfado. Del mismo modo ocurre con el miedo y la tristeza, ya que la mayoría de los errores de miedo son porque la emoción reconocida es tristeza.

La causa de esta mala clasificación puede deberse a que las emociones con las que se confunde poseen un nivel equitativo de actividad. Algunos estudios clasifican las emociones como activas y pasivas:

- Pasivas: Se caracterizan por una velocidad de locución lenta, un volumen bajo, un tono bajo y un timbre más resonante.
- Activas: Caracterizadas por una velocidad de locución rápida, alto volumen, alto tono y un timbre fuerte.

La alegría y el enfado son emociones activas, mientras que el miedo y la tristeza son emociones pasivas. Resulta interesante que el grueso de los errores caiga en emociones con el mismo nivel de actividad.





Capítulo 6

Presupuesto

"No estimes el dinero en más ni en menos de lo que vale, porque es un buen siervo y un mal amo."

Alejandro Dumas (1803-1870)

6.1 Diagrama de tareas

Este proyecto ha sido realizado en un tiempo aproximado de un año, durante el cual los proyectantes lo han compaginado con estudios y becas. La elaboración del proyecto se llevó a cabo en diferentes fases, empezando con un planteamiento de la idea principal del proyecto, seguido de la realización del diseño del sistema, para continuar con el desarrollo. Una vez terminada la implementación, se probó el sistema y se realizó la documentación pertinente.

Se muestra a continuación una lista detallada de tareas y su correspondiente diagrama de Gantt.

	Nombre de tarea	Duración
1	Análisis del problema	40 días
2	Recopilación de información sobre somniloquias	7 días
3	Estudio de aplicaciones Android	10 días
4	Estudio de herramientas de transcripción	13 días
5	Recopilación de información sobre análisis de emociones	10 días
6	Definición del alcance	2 días
7	Diseño	14 días
8	Diseño arquitectónico	7 días
9	Diseño detallado	7 días
10	Implementación	177 días
11	Aprender Android	7 días
12	Módulo aplicación	170 días
13	Módulo detección SPL	63 días
14	Módulo gestión bases de datos Android	10 días
15	Módulo eliminación ruidos	40 días
16	Módulo de transcripción	44 días
17	Módulo de análisis de emociones	83 días
18	Aprender J2EE	9 días
19	Módulo de comunicación	22 días
20	Módulo Servlet	30 días
21	Módulo de gestión de bases de datos Servidor	7 días
22	Pruebas	2 días
23	Documentación	28 días

Ilustración 28. Tareas del proyecto

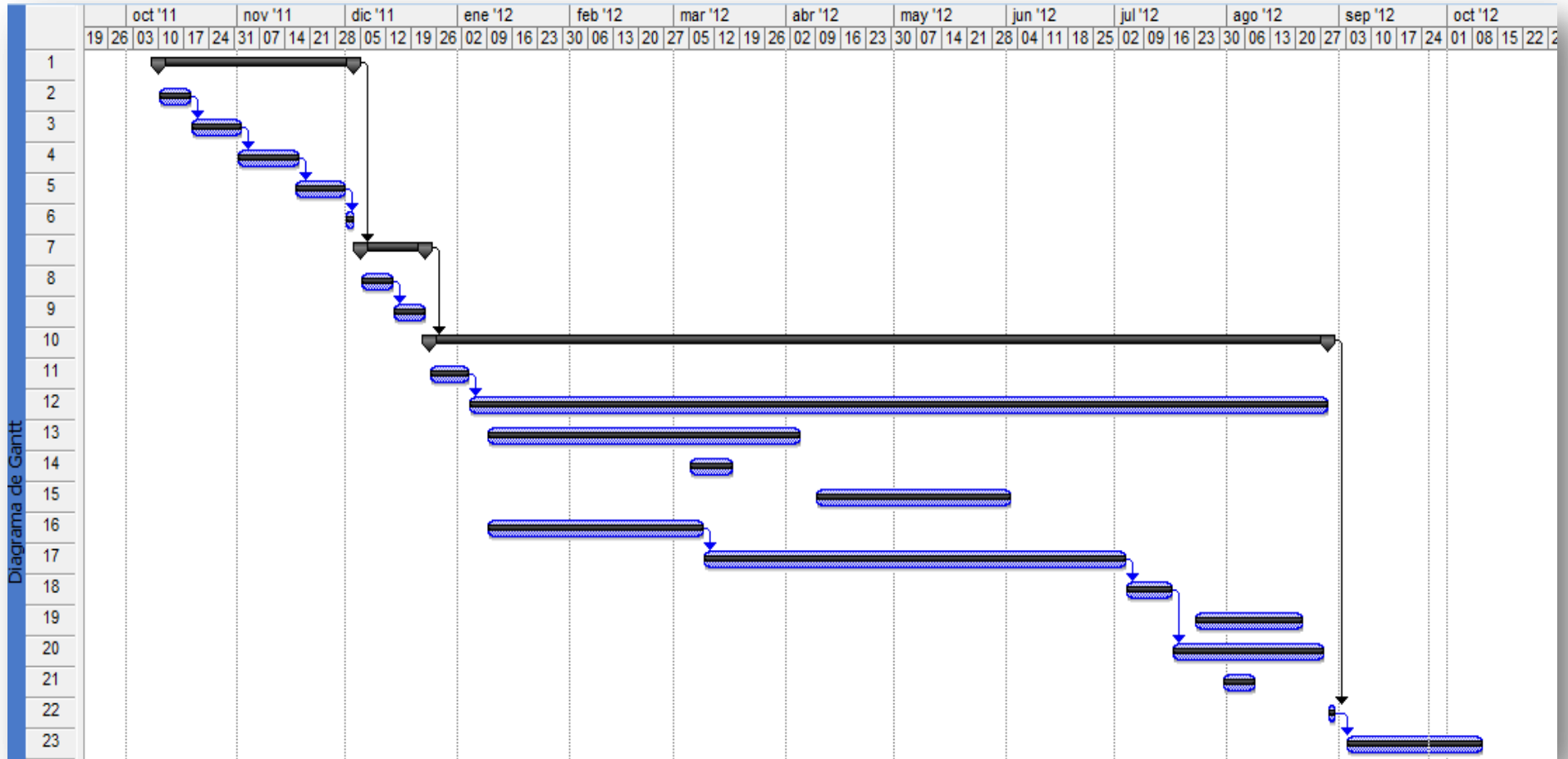


Ilustración 29. Diagrama de Gantt

6.2 Desglose de costes

Como se comentó anteriormente, la duración del proyecto ha sido de aproximadamente 12 meses, con una media de trabajo de cuatro horas diarias. Si contamos los días laborales de este periodo nos salen 251 días que si se multiplican por las horas de trabajo al día y el número de proyectantes nos sale un total de **2008 horas** de trabajo.

6.2.1 Personal

Los costes del personal para el proyecto ascienden a **24.096 €** (veinticuatro mil noventa y seis euros), calculados a partir del coste de la hora para un ingeniero en informática sin experiencia, situado en 12 €/h (doce euros/hora), multiplicado por las 2008 h. (dos mil ocho horas) dedicadas a la elaboración del proyecto. Para el cálculo anterior no se han tenido en cuenta los gastos derivados de la contratación de personal como puede ser las cantidades a pagar al sistema de Seguridad Social español.

6.2.2 Material

En este apartado se detalla el coste de los materiales empleados para la realización del proyecto. Estos costes se pueden dividir en equipos y licencias.

Equipos

Equipo	Precio
Ordenador portátil Toshiba Satellite L655	599,99 €
Ordenador portátil Toshiba Satellite A300	699,99 €
Teléfono móvil Samsun Galaxy i9000	292,72 €
Micrófono ordenador	5,70 €
Total	1598,40 €

Tabla 216. Coste de los equipos

El total gastado en equipos asciende a la cantidad de 1598,40 € (mil quinientos noventa y ocho euros y cuarenta céntimos).

Licencias

Concepto	Unidades	Precio/unidad	Coste
Windows 7 Professional (*)	1	0,00 €	0,00 €
Windows 7 Home Premium	1	199,99 €	199,99 €
Microsoft Office Word 2007	2	229 €	458 €
Microsoft Excel 2007 (*)	2	0,00 €	0,00 €
Microsoft Office Project 2007 (*)	1	0,00 €	0,00 €
Microsoft Office PowerPoint 2007 (*)	2	0,00 €	0,00 €
SoX	1	0,00 €	0,00 €
Eclipse Helios	2	0,00 €	0,00 €
Eclipse Índigo IDE	2	0,00 €	0,00 €
Android SDK	2	0,00 €	0,00 €
MySQL 5.5	1	0,00 €	0,00 €
SQLite	1	0,00 €	0,00 €
Apache Tomcat	2	0,00 €	0,00 €
Java Platform 7	2	0,00 €	0,00 €
Total			657,99 €

Tabla 217. Coste de las licencias

(*) Las licencias de estos productos *Microsoft* se obtuvieron gracias al acuerdo MSDNAA entre *Microsoft Corp.* y la *Universidad Carlos III de Madrid*.

El total gastado en licencias asciende a la cantidad de **657,99 €** (seiscientos cincuenta y siete euros con noventa y nueve céntimos).

6.2.3 Transporte

Para acudir al lugar de trabajo se empleó el transporte público, para ello se adquirieron abonos de transporte mensuales tipo B1. El coste del abono ha variado en el transcurso del proyecto pasando de 55,50 € (cincuenta y cinco euros y cincuenta céntimos) a 60,90 € (sesenta euros y noventa céntimos) el 1 de septiembre de 2012, por lo que contabilizaremos once meses con el primer precio y uno con el segundo. El total asciende a **1342,8 €** (mil trescientos cuarenta y dos euros y ochenta céntimos).

6.2.4 Resumen de costes

Calculando la suma económica necesaria para cubrir los costes derivados de los apartados anteriores, se necesitan **27.695,19 €** (veintisiete mil seiscientos noventa y euros y diecinueve céntimos) para realizar el proyecto. Sin embargo, esta cantidad no incluye el impuesto sobre valor añadido ni los márgenes de beneficio y riesgo asociados a este tipo de proyectos.

Concepto	Cantidad
Personal	24.096 €
Equipos	1.598,40 €
Licencias	657,99 €
Transporte	1.342,8 €
Total	27.695,19 €

Tabla 218. Resumen de los costes

6.2.5 Totales

Para terminar se muestra el precio final del proyecto.

Concepto	Cantidad
Coste total	27.695,19 €
Riesgo (20%)	5.539,04 €
Beneficio (20%)	5.539,04 €
Total sin I.V.A.	38.773,27 €
I.V.A. (21%)	8.142,39 €
Total	46.915,66 €

Tabla 219. Coste total



Teniendo en cuenta los costes desglosados en los apartados anteriores, el presupuesto total de este proyecto asciende a la cantidad de **46.915,66 € (cuarenta y seis mil novecientos quince euros y sesenta y seis céntimos)**.

Leganés, Octubre de 2012

Los ingenieros proyectistas Virginia Aparicio Paniagua y Jorge Pérez Muñoz.

Fdo.

Virginia Aparicio Paniagua

Jorge Pérez Muñoz





Capítulo 7

Conclusiones

"La llave del éxito es el conocimiento del valor de las cosas."

John Boyle O'Reilly (1844-1890)



En este punto nos gustaría comentar una serie de conclusiones que hemos obtenido con la realización del presente proyecto:

- El sistema ha sido desarrollado de manera modular, de forma que cada una de las funcionalidades cuenta con un paquete propio. Esto favorece a la implantación de nuevas funcionalidades y a la eliminación de algún módulo ya existente.
- Para la grabación de somniloquias se ha creado un modelo de grabación inteligente que obtiene únicamente las palabras que una determinada persona haya dicho durante el proceso del sueño. Esto aporta mejoras frente a la grabación tradicional, ya que permite ahorrar en tamaño de almacenamiento, debido a que se descartan los periodos de silencio.
- Se han conseguido eliminar los ruidos que no corresponden a palabras. Junto a la grabación inteligente, esto permite reducir la duración de los audios, por lo que el usuario que desee oír sus grabaciones perderá menos tiempo en este proceso.
- El sistema dispone de un módulo de análisis acústico que funciona de manera eficiente, debido a que agiliza el proceso, eliminando aquellas frecuencias que no corresponden a la voz humana.
- Los resultados obtenidos en el reconocimiento automático de emociones sitúan a la emoción primaria tristeza como la emoción que mejor se diferencia con un resultado de un 80% de aciertos, seguido del enfado con un 64,71%, terminando con la alegría y el miedo con un 35,29%.
- Para utilizar el sistema se debe disponer de un dispositivo móvil con el sistema operativo Android. El problema es que en estos momentos el servidor es local, por lo que solamente funciona si el móvil está conectado a la misma red que el servidor.
- Si se desea implantar la parte del servicio web en un servidor externo, se pide como requisito que funcione con Apache Tomcat y que cuente con una base de datos MySQL.
- La transcripción funciona en estos momentos para sistemas operativos Windows, siendo muy fácil su implantación en Linux, debido a que únicamente deberíamos instalar el software de SoX.



- Para el reconocimiento de emociones no se dispone de gran cantidad de herramientas que se dediquen a esta labor. Además las herramientas existentes no son de código libre por lo que la implementación en esta área está muy acotada.
- El sistema puede servir a profesionales del área de la psicología como herramienta de ayuda para diagnosticar ciertas patologías, como la ansiedad o el estrés.





Capítulo 8

Líneas futuras

*"Sólo cerrando las puertas detrás de uno se abren
ventanas hacia el porvenir."*

Françoise Sagan (1935-2004)

A continuación se expondrán las ideas principales que podrían ser aplicadas en un futuro al sistema desarrollado en este proyecto.

En primer lugar, es importante resaltar que la funcionalidad que ofrece el sistema propuesto en este proyecto es abundante, siendo receptor de un sinfín de funcionalidades más. Las opciones que ofrece la aplicación móvil se ven enormemente potenciadas por las que presenta la plataforma web.

El diseño de la arquitectura cliente-servidor requiere una comunicación constante y necesaria entre ambas partes. Esto implica que, actualmente, ambos componentes sean dependientes uno del otro, por lo que si el servidor no se encuentra activo, la aplicación móvil no podrá realizar funcionalidad alguna. Como línea futura, se propone la opción de trabajar “*offline*”, de forma que la aplicación móvil grabe las somniloquias, y que la recepción de los archivos procesados por el servidor se realice una vez éste se encuentre activo.

Unido a lo anterior, se podría ofrecer al usuario la posibilidad de decidir qué archivos desea que sean procesados en el servidor y cuáles no. Tal y como se ha comentado en algunos puntos del documento, el envío y posterior procesado de las somniloquias se realiza automáticamente tras la creación de un nuevo archivo de audio. En lugar de esto, sería interesante ofrecer al usuario la elección de los archivos a procesar en el servidor.

También podrían mejorarse algunas opciones que el sistema ya presenta:

- Se podría dar la posibilidad al usuario de elegir el formato de sus archivos de audio.
- La reproducción de los ficheros de audio podría presentar una línea de tiempo para que el usuario pudiese trasladarse a cualquier punto dentro del archivo.

Por otra parte, sería conveniente que la visualización de los resultados generados por el servidor se pudiese realizar desde una página web. Es decir, como funcionalidad añadida a la existente, se podría crear una página web donde cada usuario pudiese entrar en ella con su nombre de usuario y contraseña. Dentro de su perfil podría visualizar los resultados de una forma más limpia, ordenada y con opciones adicionales, como por ejemplo:

- Descargar las somniloquias y los archivos post-procesados a su disco duro
- Descargar los archivos de resultados que genera el servidor.
- Acceder a datos avanzados resultantes del análisis acústico de sus archivos de audio.



Por último, en una versión posterior del sistema, se podría crear un nuevo perfil de usuario asociado a un profesional de la psicología. Teniendo en cuenta que el sistema propuesto puede ser de gran ayuda a un psicólogo en su labor de ayuda a sus pacientes, este perfil podría acceder a toda la información de éstos.





Capítulo 9

Glosario de términos y acrónimos

Darí­a todo lo que sé, por la mitad de lo que ignoro.

René Descartes (1596-1650)



A continuación se explican términos y acrónimos que aparecen o se utilizan en el presente proyecto.

9.1 Glosario

Término	Definición
Android	Sistema operativo basado en Linux para dispositivos móviles tales como los teléfonos inteligentes o tablets.
Base de datos	Una base de datos es un conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso.
Contenedor de servlet	Entorno de ejecución para componentes web que incluye seguridad, concurrencia, gestión del ciclo de vida, procesamiento de transacciones, despliegue y otros servicios.
Eclipse	Entorno de desarrollo integrado de código abierto multiplataforma que cuenta con una plataforma de programación, desarrollo y compilación de elementos.
Framework	Estructura conceptual y tecnológica de soporte definida, normalmente con artefactos o módulos de software concretos, con base en la cual otro proyecto.
Java	Lenguaje de programación orientado a objetos, desarrollado por Sun Microsystems.
Parasomnia	Trastorno de la conducta durante el sueño asociado con episodios breves o parciales de despertar, sin que se produzca una interrupción importante del sueño ni una alteración del nivel de vigilia diurno.
Prosodia	La prosodia es una rama de la lingüística que analiza y representa formalmente aquellos elementos de la expresión oral, tales como el acento, los tonos y la entonación.
Servidor	Es una computadora que, formando parte de una red, provee servicios a otras computadoras denominadas clientes.
Servlet	Los servlets son objetos que corren dentro y fuera del contexto de un contenedor de servlets y extienden su funcionalidad.

Sistema Operativo	Es un conjunto de programas que en un sistema informático gestiona los recursos de hardware y provee servicios a los programas de aplicación.
Somniloquia	Parasomnia referida al hablar en voz alta durante el sueño. Puede ser bastante fuerte, variar entre simples sonidos hasta largos discursos, y ocurrir una o varias veces durante el sueño.
Transcripción	En sentido lingüístico, es la representación sistemática de la lengua en forma escrita

Tabla 220. Glosario

9.2 Acrónimos

Acrónimo	Significado
API	Application Programming Interface. Conjunto de funciones y procedimientos que ofrece cierta biblioteca para ser utilizado por otro software como una capa de abstracción.
ASR	Automatic Speech Recognition. Es una parte de la Inteligencia Artificial que tiene como objetivo permitir la comunicación hablada entre seres humanos y computadoras.
AWT	Abstract Windowing Toolkit. Es un kit de herramientas de gráficos, interfaz de usuario, y sistema de ventanas independiente de la plataforma original de Java.
CGI	Common Gateway Interface. Es una importante tecnología de la World Wide Web que permite a un cliente solicitar datos de un programa ejecutado en un servidor web.
DDR2	Double Data Rate. Es un tipo de memoria RAM.
GB	Un gigabit es una unidad de medida de almacenamiento informático normalmente abreviada como Gb, que equivale a 10^9 bits.
GHZ	El Gigahercio es un múltiplo de la unidad de medida de frecuencia hercio y equivale a 10^9 hercios.
HMM	Hidden Markov Model. Es un modelo estadístico en el que se asume que el sistema a modelar es un proceso de Márkov de parámetros desconocidos.



HNR	Harmonics to Noise Ratio. Razón-armónico-ruido
HTTP	Hypertext Transfer Protocol. Es el protocolo usado en cada transacción de la World Wide Web.
IDE	Integrated Development Environment. Es un programa informático compuesto por un conjunto de herramientas de programación.
IOS	(Anteriormente denominado iPhone OS) es un sistema operativo móvil de Apple.
J2EE	Java 2 Platform, Enterprise Edition. La plataforma proporciona una API y entorno de ejecución para el desarrollo y ejecución de software empresarial.
JDBC	Java Database Connectivity. Es una API que permite la ejecución de operaciones sobre bases de datos desde el lenguaje de programación Java.
JDK	Java Development Kit. Es un software que provee herramientas de desarrollo para la creación de programas en Java. Puede instalarse en una computadora local o en una unidad de red.
JVM	Java Virtual Machine. Es una máquina virtual de proceso nativo, capaz de interpretar y ejecutar instrucciones expresadas en bytecode de Java.
JRE	Java Runtime Environment. Es un conjunto de utilidades que permite la ejecución de programas Java.
NHR	Noise-to-Harmonics-Ratio. Razón-ruido-armónico
NREM	Non-Rapid Eye Movement. Engloba las fases 1-4 del sueño y es conocido como sueño lento.
RAH	Reconocimiento Automático del Habla. Es una parte de la Inteligencia Artificial que tiene como objetivo permitir la comunicación hablada entre seres humanos y computadoras.
RAM	Random-Access Memory. se utiliza como memoria de trabajo para el sistema operativo, los programas y la mayoría del software.



REM	Rapid Eye Movements. Es la quinta fase del sueño y se caracteriza por un rápido y movimiento de los ojos.
RMI	Remote Method Invocation. Mecanismo ofrecido por Java para invocar métodos de manera remota. Se utiliza para la comunicación entre servidores y aplicaciones distribuidas basadas en Java.
SDK	Software Development Kit. Es un conjunto de herramientas de desarrollo de software que le permite al programador crear aplicaciones.
SPL	Sound Pressure Level. Se define como la diferencia de presión instantánea y la presión atmosférica estática.
UML	Lenguaje Unificado de Modelado. Es el lenguaje de modelado de sistemas software más conocido y utilizado en la actualidad.
XML	eXtensible Markup Language. Es un lenguaje de marcas y se propone como un estándar para el intercambio de información estructurada entre diferentes plataformas.
WAV	WAVEform audio file format. Formato de fichero de audio con forma de onda.
ZCR	Zero Crossing Rate. Tasa de cruces por cero.

Tabla 221. Acrónimos





Capítulo 10

Bibliografía y referencias

"De razones vive el hombre, de sueños sobrevive."

Miguel de Unamuno (1864-1936)



[1] AASM. *The International Classification of Sleep Disorders. Diagnostic and coding manual*. 2. AASM, Westchester, 2005.

[2] A.M. Arkin, M.F. Toth, J. Barker and J.M. Haste, *The frequency of sleeptalking in the laboratory among chronic sleepwalkers and good dream recallers*. J. Nerv. Ment. 1970.

[3] A. Rechtschaffen, D.R. Goodnough and A. Shapiro, *Patterns of sleeptalking*. Arch. Gen. Psychiatr., 1962.

[4] G. Andriani, *Fisiología psicológica del somniloquio*. Ann. Neurol, 1892.

[5] C. Darwin, *The Expression of the Emotions in Man and Animals*, 1872.

[6] Klaus R. Scherer, What are emotions? And how can they be measured? Social Science Information December 2005 vol. 44 no. 4 695-729

[7] Ribot, Th., *The psychology of the emotions* (2nd ed.). The contemporary science series. London, England: Walter Scott Publishing Co. (1911)

[8] Plutchik, R., *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion*, 1, New York: Academic (1980)

[9] Cortes, Corinna; and Vapnik, Vladimir N.; "Support-Vector Networks", Machine Learning, 20, 1995.

[10] Lawrence R. Rabiner, *A tutorial on Hidden Markov Models and selected applications in speech recognition*. Proceedings of the IEEE 77 (2): 257–286. (February 1989).

[11] H. Pérez Espinosa, L. Villaseñor Pineda, C.A. Reyes García. *EmoSpeech*. Disponible en: <http://wayra.org/es> (En línea. Accedido el 8 de Octubre de 2012)

[12] eXaudios Technologies. *eXaudios – Decoding de Human intonation*. Disponible en: <http://www.exaudios.com/default.aspx> (En línea. Accedido el 8 de Octubre de 2012)

[13] Nuance. *Dragon Naturally Speaking: Reconocimiento de voz Dragon*. Disponible en: <http://www.nuance.es/particulares/producto/dragon-para-pc/index.htm> (En línea. Accedido el 8 de Octubre de 2012)



[14] CMU Sphinx. *CMU Sphinx. Open Source Toolkit For Speech Recognition*. Project by Carnegie Mellon University. Disponible en: <http://cmusphinx.sourceforge.net/> (En línea. Accedido el 8 de Octubre de 2012)

[15] Julius. *Julius*. Disponible en: http://julius.sourceforge.jp/en_index.php (En línea. Accedido el 8 de Octubre de 2012)

[16] Microsoft. *Windows speech recognition*. Disponible en: <http://www.microsoft.com/enable/products/windowsvista/speech.aspx> (En línea. Accedido el 8 de Octubre de 2012)

[17] Apple. *Siri*. Disponible en: <http://www.apple.com/ios/siri/> (En línea. Accedido el 8 de Octubre de 2012)

[18] Google. *Voice action for Android*. <http://www.google.com/mobile/voice-actions/> (En línea. Accedido el 8 de Octubre de 2012)

[19] Oracle. *Página de la documentación de Java*. Disponible en: <http://www.oracle.com/technetwork/java/javase/documentation/index.html> (En línea. Accedido el 8 de Octubre de 2012)

[20] Oracle. *Página de la documentación de Java EE*. Disponible en: <http://www.oracle.com/technetwork/java/javaee/documentation/index.html> (En línea. Accedido el 8 de Octubre de 2012)

[21] Apache Foundation. *Apache Software Foundation site*. Disponible en: <http://apachefoundation.wikispaces.com/> (En línea. Accedido el 8 de Octubre de 2012)

[22] MySQL. *MySQL Documentation: MySQL Reference Manuals*. Disponible en: <http://dev.mysql.com/doc/> (En línea. accedido el 8 de Octubre de 2012)

[23] SoX. *SoX - Sound eXchange | Documentation*. Disponible en: <http://sox.sourceforge.net/Docs/Documentation> (En línea. Accedido el 8 de Octubre de 2012)

[24] J. Pérez Muñoz. *Memoria*. Leganés, Octubre 2012.

[25] A. Varela, H. Cuayáhuatl, J.A. Nolzco-Flores. *Creating a Mexican Spanish Version of the CMU Sphinx-III Speech Recognition System*. Universidad Autónoma de Tlaxcala, Department of Engineering and Technology.



[26] G. Fairbanks, *Recent experimental investigations of vocal pitch in speech*. Journal of the Acoustical Society of America, 1940

[27] C.E. Williams, K.N. Stevens, *Vocal correlates of emotional states*, J.K. Darby (ed.): *Speech Evaluation in Psychiatry*, New York, 1981.

[28] H. Levin, W. Lord, *Speech Pitch Frequency as an Emotional State Indicator*, 1975.

[29] R. van Bezooijen, *The characteristics and recognizability of vocal expression of emotions*. Dordrecht, The Netherlands, 1984.

[30] A. Rodríguez, P. Lázaro, N. Montoya, J.M. Blanco, D. Bernardas, J.M. Oliver, L. Longhi, *Modelización acústica de la expresión emocional en el español*. U.A.B. Dto. de Comunicación Audiovisual y Publicidad, 1999

[31] V. Francisco, P. Gervás, R. Hervás, *Análisis y síntesis de expresión emocional en cuentos leídos en voz alta*, Procesamiento del Lenguaje Natural, núm. 35, pp. 293-300, 2005





Capítulo 11

Anexos

"Allí donde el agua alcanza su mayor profundidad, se mantiene más en calma."

William Shakespeare (1564-1616)

11.1 Anexo 1

Esta sección muestra los ficheros de cada uno de los actores con los resultados medios obtenidos entre la emoción neutra y el resto de emociones.

RESULTADOS

1ª PERSONA

- **Alegría**

Emoción	Neutro 1	Neutro 2	Alegría 1	Alegría 2
F0	437	625	437	625
Rango del pitch	0.353-1.0	0.398-1.0	0.34-1.0	0.274-1.0
SPL	33.74	33.812	33.813	33.762
Rango SPL	24.707-35.608	24.71-35.654	24.664-35.715	24.261-35.667
Velocidad (palabras/min)	61.475	126.404	64.378	33.186
Energía	1791.674	1452.863	1921.567	1763.074
Valor medio	0.681	0.717	0.72	0.697
Varianza	0.01	0.012	0.016	0.017

Tabla 222. Actor 1, valores neutro y alegría

Emoción	Neutro Media	Alegría Media	Comparación	Variación
F0	531	531	100	1
Rango del pitch	0,6245	0,693	110,968775	1,10968775
SPL	33,776	33,7875	100,0340478	1,00034048
Rango SPL	10,9225	11,2285	102,8015564	1,02801556
Velocidad (palabras/min)	93,9395	48,782	51,92916718	0,51929167
Energía	1622,2685	1842,3205	113,5644624	1,13564462
Valor medio	0,699	0,7085	101,3590844	1,01359084
Varianza	0,011	0,0165	150	1,5

Tabla 223. Actor 1, comparaciones entre neutro y alegría

- **Enfado**

Emoción	Neutro 1	Neutro 2	Enfado 1	Enfado 2
F0	437	625	656	437
Rango del pitch	0.353-1.0	0.398-1.0	0.276-1.0	0.335-1.0
SPL	33.74	33.812	33.751	33.857
Rango SPL	24.707-35.608	24.71-35.654	24.481-35.665	24.47-35.684
Velocidad (palabras/min)	61.475	126.404	18.029	19.947
Energía	1791.674	1452.863	1618.004	1648.514
Valor medio	0.681	0.717	0.694	0.742
Varianza	0.01	0.012	0.02	0.015

Tabla 224. Actor 1, valores neutro y enfado

Emoción	Neutro Media	Enfado Media	Comparación	Variación
F0	531	546,5	102,9190207	1,02919021
Rango del pitch	0,6245	0,6945	111,2089672	1,11208967
SPL	33,776	33,804	100,0828991	1,00082899
Rango SPL	10,9225	11,199	102,5314717	1,02531472
Velocidad (palabras/min)	93,9395	18,988	20,21300944	0,20213009
Energía	1622,2685	1633,259	100,6774772	1,00677477
Valor medio	0,699	0,718	102,7181688	1,02718169
Varianza	0,011	0,0175	159,0909091	1,59090909

Tabla 225. Actor 1, comparaciones entre neutro y enfado

- **Tristeza**

Emoción	Neutro 1	Neutro 2	Tristeza 1	Tristeza 2	Tristeza3
F0	437	625	343	562	437
Rango del pitch	0.353-1.0	0.398-1.0	0.157-1.0	0.311-1.0	0.313-1.0
SPL	33.74	33.812	33.701	33.699	33.727
Rango SPL	24.707-35.608	24.71-35.654	24.212-35.582	24.598-35.579	24.34-35.582
Velocidad (palabras/min)	61.475	126.404	63.345	59.289	64.879
Energía	1791.674	1452.863	2104.338	1785.291	2102.548
Valor medio	0.681	0.717	0.667	0.666	0.677
Varianza	0.01	0.012	0.014	0.014	0.013

Tabla 226. Actor 1, valores neutro y enfado

Emoción	Neutro Media	Tristeza Media	Comparación	Variación
F0	531	452,5	85,2165725	0,85216573
Rango del pitch	0,6245	0,766	122,6581265	1,22658127
SPL	33,776	33,7	99,77498816	0,99774988
Rango SPL	10,9225	11,199	102,5314717	1,02531472
Velocidad (palabras/min)	93,9395	61,317	65,27286179	0,65272862
Energía	1622,2685	1944,8145	119,8824054	1,19882405
Valor medio	0,699	0,6665	95,35050072	0,95350501
Varianza	0,011	0,014	127,2727273	1,27272727

Tabla 227. Actor 1, comparaciones entre neutro y enfado

- Miedo

Emoción	Neutro 1	Neutro 2	Miedo 2	Miedo 3
F0	437	625	562	625
Rango del pitch	0.353-1.0	0.398-1.0	0.264-1.0	0.39-1.0
SPL	33.74	33.812	33.763	33.765
Rango SPL	24.707-35.608	24.71-35.654	24.362-35.69	24.484-35.666
Velocidad (palabras/min)	61.475	126.404	17.689	72.957
Energía	1791.674	1452.863	1641.255	1968.132
Valor medio	0.681	0.717	0.696	0.695
Varianza	0.01	0.012	0.015	0.013

Tabla 228. Actor 1, valores neutro y miedo

Emoción	Neutro Media	Miedo Media	Comparación	Variación
F0	531	593,5	111,7702448	1,11770245
Rango del pitch	0,6245	0,766	122,6581265	1,22658127
SPL	33,776	33,764	99,96447181	0,99964472
Rango SPL	10,9225	11,255	103,0441749	1,03044175
Velocidad (palabras/min)	93,9395	45,323	48,24701004	0,4824701
Energía	1622,2685	1804,6935	111,2450559	1,11245056
Valor medio	0,699	0,6955	99,49928469	0,99499285
Varianza	0,011	0,014	127,2727273	1,27272727

Tabla 229. Actor 1, comparaciones entre neutro y miedo

2ª PERSONA

- **Alegría**

Emoción	Neutro 1	Neutro 2	Alegría 1	Alegría 2
F0	625	656	281	531
Rango del pitch	0.262-1.0	0.325-1.0	0.314-1.0	0.275-1.0
SPL	33.608	33.582	33.629	33.654
Rango SPL	24.273-35.661	24.055-35.653	24.256-35.722	23.71-35.678
Velocidad (palabras/min)	43.605	73.892	95.663	139.752
Energía	1111.86	1242.098	1298.476	1115.941
Valor medio	0.631	0.617	0.639	0.653
Varianza	0.019	0.016	0.019	0.023

Tabla 230. Actor 2, valores neutro y alegría

Emoción	Neutro Media	Alegría Media	Comparación	Variación
F0	640,5	531	82,9039813	0,82903981
Rango del pitch	0,7065	0,725	102,618542	1,02618542
SPL	33,595	33,654	100,175621	1,00175621
Rango SPL	11,493	11,968	104,13295	1,0413295
Velocidad (palabras/min)	58,7485	139,752	237,881818	2,37881818
Energía	1176,979	1115,941	94,8140111	0,94814011
Valor medio	0,624	0,653	104,647436	1,04647436
Varianza	0,0175	0,023	131,428571	1,31428571

Tabla 231. Actor 2, comparaciones entre neutro y alegría

- **Enfado**

Emoción	Neutro 1	Neutro 2	Enfado 1	Enfado 2
F0	625	656	656	625
Rango del pitch	0.262-1.0	0.325-1.0	0.261-1.0	0.17-1.0
SPL	33.608	33.582	33.623	33.62
Rango SPL	24.273-35.661	24.055-35.653	24.567-35.813	23.961-35.687
Velocidad (palabras/min)	43.605	73.892	21.552	55.147
Energía	1111.86	1242.098	1147.6	927.939
Valor medio	0.631	0.617	0.637	0.643
Varianza	0.019	0.016	0.02	0.027

Tabla 232. Actor 2, valores neutro y enfado

Emoción	Neutro Media	Enfado Media	Comparación	Variación
F0	640,5	640,5	100	1
Rango del pitch	0,707	0,7845	110,96181	1,1096181
SPL	33,595	33,6215	100,078881	1,00078881
Rango SPL	11,493	11,486	99,9390934	0,99939093
Velocidad (palabras/min)	58,7485	38,3495	65,2774113	0,65277411
Energía	1176,979	1037,7695	88,1723038	0,88172304
Valor medio	0,624	0,64	102,564103	1,02564103
Varianza	0,0175	0,0235	134,285714	1,34285714

Tabla 233. Actor 2, comparaciones entre neutro y enfado

- Miedo

Emoción	Neutro 1	Neutro 2	Miedo 1	Miedo 2
F0	625	656	531	531
Rango del pitch	0.262-1.0	0.325-1.0	0.358-1.0	0.403-1.0
SPL	33.608	33.582	33.724	33.828
Rango SPL	24.273-35.661	24.055-35.653	24.564-35.639	24.898-35.674
Velocidad (palabras/min)	43.605	73.892	26.408	28.626
Energía	1111.86	1242.098	1043.886	1085.103
Valor medio	0.631	0.617	0.678	0.725
Varianza	0.019	0.016	0.015	0.013

Tabla 234. Actor 2, valores neutro y miedo

Emoción	Neutro Media	Miedo Media	Comparación	Variación
F0	640,5	531	82,9039813	0,82903981
Rango del pitch	0,738	0,62	84,0108401	0,8401084
SPL	33,595	33,776	100,538771	1,00538771
Rango SPL	11,493	10,9255	95,0622118	0,95062212
Velocidad (palabras/min)	58,7485	27,517	46,8386427	0,46838643
Energía	1176,979	1064,4945	90,4429476	0,90442948
Valor medio	0,624	0,7015	112,419872	1,12419872
Varianza	0,0175	0,014	80	0,8

Tabla 235. Actor 2, comparaciones entre neutro y miedo

- **Tristeza**

Emoción	Neutro 1	Neutro 2	Tristeza 1	Tristeza 2
F0	625	656	156	531
Rango del pitch	0.262-1.0	0.325-1.0	0.345-1.0	0.34-1.0
SPL	33.608	33.582	33.741	33.699
Rango SPL	24.273-35.661	24.055-35.653	24.829-35.668	24.444-35.732
Velocidad (palabras/min)	43.605	73.892	72.115	21.429
Energía	1111.86	1242.098	1543.135	1248.716
Valor medio	0.631	0.617	0.683	0.668
Varianza	0.019	0.016	0.012	0.017

Tabla 236. Actor 2, valores neutro y tristeza

Emoción	Neutro Media	Tristeza Media	Comparación	Variación
F0	640,5	343,5	53,62997658	0,53629977
Rango del pitch	0,7065	0,658	93,13517339	0,93135173
SPL	33,595	33,72	100,3720792	1,00372079
Rango SPL	11,493	11,968	104,1329505	1,0413295
Velocidad (palabras/min)	58,7485	46,772	79,61394759	0,79613948
Energía	1176,979	1395,9255	118,6024135	1,18602413
Valor medio	0,624	0,6755	108,2532051	1,08253205
Varianza	0,0175	0,0145	82,85714286	0,82857143

Tabla 237. Actor 2, comparaciones entre neutro y miedo

3ª PERSONA

- **Alegría**

Emoción	Neutro 1	Alegría 1	Alegría 2
F0	531	500	500
Rango del pitch	0.214-1.0	0.358-1.0	0.348-1.0
SPL	33.611	33.705	33.765
Rango SPL	24.104-35.675	24.614-35.751	24.796-35.754
Velocidad (palabras/min)	81.522	107.759	112.952
Energía	1171.082	1256.292	1303.176
Valor medio	0.629	0.67	0.699
Varianza	0.015	0.017	0.018

Tabla 238. Actor 3, valores neutro y alegría

Emoción	Neutro Media	Alegría Media	Comparación	Variación
F0	531	500	94,1619586	0,94161959
Rango del pitch	0,607	0,658	108,401977	1,08401977
SPL	33,611	33,735	100,368927	1,00368927
Rango SPL	11,571	11,0485	95,4844007	0,95484401
Velocidad (palabras/min)	94,6405	110,3555	116,604942	1,16604942
Energía	1171,082	1279,734	109,277916	1,09277916
Valor medio	0,629	0,6845	108,823529	1,08823529
Varianza	0,016	0,0175	109,375	1,09375

Tabla 239. Actor 3, comparaciones entre neutro y alegría

- **Enfado**

Emoción	Neutro 1	Enfado 1	Enfado 2
F0	531	531	531
Rango del pitch	0.214-1.0	0.303-1.0	0.297-1.0
SPL	33.611	33.778	33.747
Rango SPL	24.104-35.675	24.707-35.78	24.415-35.732
Velocidad (palabras/min)	81.522	161.043	163.043
Energía	1171.082	1285.024	1230.813
Valor medio	0.629	0.704	0.691
Varianza	0.015	0.016	0.018

Tabla 240. Actor 3, valores neutro y enfado

Emoción	Neutro Media	Enfado Media	Comparación	Variación
F0	531	531	100	1
Rango del pitch	0,607	0,7	115,321252	1,15321252
SPL	33,611	33,7625	100,450745	1,00450745
Rango SPL	11,571	11,195	96,7504969	0,96750497
Velocidad (palabras/min)	121,2825	162,043	133,607899	1,33607899
Energía	1171,082	1257,9185	107,415066	1,07415066
Valor medio	0,629	0,6975	110,890302	1,10890302
Varianza	0,0155	0,017	109,677419	1,09677419

Tabla 241. Actor 3, comparaciones entre neutro y enfado

- **Miedo**

Emoción	Neutro 1	Miedo 1
F0	531	531
Rango del pitch	0.214-1.0	0.412-1.0
SPL	33.611	33.805
Rango SPL	24.104-35.675	24.648-35.706
Velocidad (palabras/min)	81.522	110.947
Energía	1171.082	1364.16
Valor medio	0.629	0.715
Varianza	0.015	0.013

Tabla 242. Actor 3, valores neutro y miedo

Emoción	Neutro Media	Miedo Media	Comparación	Variación
F0	531	531	100	1
Rango del pitch	0,607	0,588	96,8698517	0,968698517
SPL	33,611	33,805	100,577192	1,00577192
Rango SPL	11,571	11,058	95,5665025	0,955665025
Velocidad (palabras/min)	81,522	110,947	136,094551	1,360945512
Energía	1171,082	1364,16	116,487146	1,164871461
Valor medio	0,629	0,715	113,672496	1,13672496
Varianza	0,015	0,013	86,6666667	0,866666667

Tabla 243. Actor 3, comparaciones entre neutro y miedo

- **Tristeza**

Emoción	Neutro 1	Tristeza 1
F0	531	250
Rango del pitch	0.214-1.0	0.218-1.0
SPL	33.611	33.594
Rango SPL	24.104-35.675	24.316-35.7
Velocidad (palabras/min)	81.522	40,541
Energía	1171.082	1150.562
Valor medio	0.629	0.622
Varianza	0.015	0.016

Tabla 244. Actor 3, valores neutro y tristeza

Emoción	Neutro Media	Tristeza Media	Comparación	Variación
F0	531	250	47,08097928	0,47080979
Rango del pitch	0,607	0,782	114,9917628	1,14991763
SPL	33,611	33,594	100,1606617	1,00160662
Rango SPL	11,571	11,384	95,98997494	0,95989975
Velocidad (palabras/min)	81,522	40,541	41,29560119	0,41295601
Energía	1171,082	1150,562	128,0616558	1,28061656
Valor medio	0,629	0,622	103,1796502	1,0317965
Varianza	0,015	0,016	73,33333333	0,73333333

Tabla 245. Actor 3, comparaciones entre neutro y tristeza

4ª PERSONA

- Alegría

Emoción	Neutro 1	Neutro 2	Alegría 1	Alegría 2	Alegría 3
F0	156	125	750	687	593
Rango del pitch	0.38-1.0	0.328-1.0	0.283-1.0	0.334-1.0	0.254-1.0
SPL	33.786	33.822	33.717	33.756	33.705
Rango SPL	24.804-35.752	24.522- 35.755	24.317- 35.792	24.476- 35.747	19.877- 35.781
Velocidad (palabras/min)	154.412	117.188	118.11	47.17	59.524
Energía	1344.161	1334.024	935.291	1219.096	927.994
Valor medio	0.705	0.723	0.678	0.694	0.674
Varianza	0.013	0.015	0.019	0.017	0.02

Tabla 246. Actor 4, valores neutro y alegría

Emoción	Neutro Media	Alegría Media	Comparación	Variación
F0	140,5	676,666667	481,6132859	4,81613286
Rango del pitch	0,646	0,71	109,9071207	1,09907121
SPL	33,804	33,726	99,76925808	0,99769258
Rango SPL	11,0905	12,882	116,1534647	1,16153465
Velocidad (palabras/min)	135,8	74,9346667	55,18016691	0,55180167
Energía	1339,0925	1027,46033	76,72810753	0,76728108
Valor medio	0,714	0,682	95,51820728	0,95518207
Varianza	0,014	0,01866667	133,3333333	1,33333333

Tabla 247. Actor 4, comparaciones entre neutro y alegría

- **Enfado**

Emoción	Neutro 1	Neutro 2	Enfado 1	Enfado 2
F0	156	125	625	656
Rango del pitch	0.38-1.0	0.328-1.0	0.335-1.0	0.431-1.0
SPL	33.786	33.822	33.744	33.827
Rango SPL	24.804-35.752	24.522- 35.755	24.554- 35.649	24.792- 35.742
Velocidad (palabras/min)	154.412	117.188	117.188	183.566
Energía	1344.161	1334.024	972.586	1186.162
Valor medio	0.705	0.723	0.688	0.725
Varianza	0.013	0.015	0.017	0.013

Tabla 248. Actor 4, valores neutro y enfado

Emoción	Neutro Media	Enfado Media	Comparación	Variación
F0	140,5	640,5	455,871886	4,55871886
Rango del pitch	0,646	0,617	95,5108359	0,95510836
SPL	33,804	33,7855	99,9452727	0,99945273
Rango SPL	11,0905	11,0225	99,3868626	0,99386863
Velocidad (palabras/min)	135,8	150,377	110,734168	1,10734168
Energía	1339,0925	1079,374	80,6048873	0,80604887
Valor medio	0,714	0,7065	98,9495798	0,9894958
Varianza	0,014	0,015	107,142857	1,07142857

Tabla 249. Actor 4, comparaciones entre neutro y enfado

- **Miedo**

Emoción	Neutro 1	Neutro 2	Miedo 1
F0	156	125	468
Rango del pitch	0.38-1.0	0.328-1.0	0.376-1.0
SPL	33.786	33.822	33.729
Rango SPL	24.804-35.752	24.522-35.755	24.505-35.72
Velocidad (palabras/min)	154.412	117.188	79.787
Energía	1344.161	1334.024	1038.347
Valor medio	0.705	0.723	0.68
Varianza	0.013	0.015	0.016

Tabla 250. Actor 4, valores neutro y miedo

Emoción	Neutro Media	Miedo Media	Comparación	Variación
F0	140,5	468	333,0960854	3,330960854
Rango del pitch	0,646	0,624	96,59442724	0,965944272
SPL	33,804	33,729	99,77813277	0,997781328
Rango SPL	11,0905	11,215	101,1225824	1,011225824
Velocidad (palabras/min)	135,8	79,787	58,7533137	0,587533137
Energía	1339,0925	1038,347	77,54109593	0,775410959
Valor medio	0,714	0,653	91,45658263	0,914565826
Varianza	0,014	0,023	164,2857143	1,642857143

Tabla 251. Actor 4, comparaciones entre neutro y miedo

- Tristeza

Emoción	Neutro 1	Neutro 2	Tristeza 1	Tristeza 2
F0	156	125	250	250
Rango del pitch	0.38-1.0	0.328-1.0	0.267-1.0	0.236-1.0
SPL	33.786	33.822	33.701	33.636
Rango SPL	24.804-35.752	24.522-35.755	24.199-35.695	24.255-35.654
Velocidad (palabras/min)	154.412	117.188	56.533	37.313
Energía	1344.161	1334.024	1401.157	1294.056
Valor medio	0.705	0.723	0.666	0.637
Varianza	0.013	0.015	0.013	0.012

Tabla 252. Actor 4, valores neutro y tristeza

Emoción	Neutro Media	Tristeza Media	Comparación	Variación
F0	140,5	250	177,935943	1,77935943
Rango del pitch	0,646	0,7485	115,866873	1,15866873
SPL	33804	33668,5	99,5991599	0,9959916
Rango SPL	11,0905	11,4475	103,218971	1,03218971
Velocidad (palabras/min)	135,8	46,923	34,5530191	0,34553019
Energía	1339,0925	1347,6065	100,635804	1,00635804
Valor medio	0,714	0,6515	91,2464986	0,91246499
Varianza	0,014	0,0125	89,2857143	0,89285714

Tabla 253. Actor 4, comparaciones entre neutro y tristeza

5ª PERSONA

- **Alegría**

Emoción	Neutro 1	Alegre 1
F0	468	468
Rango del pitch	0.425-1.0	0.303-1.0
SPL	33.822	33.634
Rango SPL	24.862-35.704	24.324-35.709
Velocidad (palabras/min)	27.174	19.43
Energía	1127.501	1253.049
Valor medio	0.719	0.637
Varianza	0.009	0.015

Tabla 254. Actor 5, valores neutro y alegría

Emoción	Neutro Media	Alegría Media	Comparación	Variación
F0	468	468	100	1
Rango del pitch	0,575	0,697	121,217391	1,21217391
SPL	33,822	33,634	99,4441488	0,99444149
Rango SPL	10,842	11,385	105,008301	1,05008301
Velocidad (palabras/min)	27,174	19,43	71,5021712	0,71502171
Energía	1127,501	1253,049	111,135068	1,11135068
Valor medio	0,719	0,637	88,5952712	0,88595271
Varianza	0,009	0,015	166,666667	1,66666667

Tabla 255. Actor 5, comparaciones entre neutro y alegría

- **Enfado**

Emoción	Neutro 1	Enfado 1
F0	468	437
Rango del pitch	0.425-1.0	0.306-1.0
SPL	33.822	33.638
Rango SPL	24.862-35.704	24.409-35.695
Velocidad (palabras/min)	27.174	27.174
Energía	1127.501	914.551
Valor medio	0.719	0.641
Varianza	0.009	0.017

Tabla 256. Actor 5, valores neutro y enfado

Emoción	Neutro Media	Enfado Media	Comparación	Variación
F0	468	437	93,3760684	0,93376068
Rango del pitch	0,575	0,694	120,695652	1,20695652
SPL	33,822	33,638	99,4559754	0,99455975
Rango SPL	10,842	11,286	104,095185	1,04095185
Velocidad (palabras/min)	27,174	27,174	100	1
Energía	1127,501	914,551	81,1130988	0,81113099
Valor medio	0,719	0,641	89,1515994	0,89151599
Varianza	0,009	0,017	188,888889	1,88888889

Tabla 257. Actor 5, comparaciones entre neutro y enfado

- Miedo

Emoción	Neutro 1	Miedo 1
F0	468	125
Rango del pitch	0.425-1.0	0.414-1.0
SPL	33.822	33.779
Rango SPL	24.862-35.704	24.909-35.753
Velocidad (palabras/min)	27.174	45.317
Energía	1127.501	2518.671
Valor medio	0.719	0.697
Varianza	0.009	0.007

Tabla 258. Actor 5, valores neutro y miedo

Emoción	Neutro Media	Miedo Media	Comparación	Variación
F0	468	125	26,70940171	0,267094017
Rango del pitch	0,575	0,586	101,9130435	1,019130435
SPL	33822	33779	99,87286382	0,998728638
Rango SPL	10,842	10,844	100,0184468	1,000184468
Velocidad (palabras/min)	27174	45317	166,7660263	1,667660263
Energía	1127,501	2518,671	223,3852564	2,233852564
Valor medio	0,719	0,697	96,94019471	0,969401947
Varianza	0,009	0,007	77,77777778	0,777777778

Tabla 259. Actor 5, comparaciones entre neutro y miedo

- Tristeza

Emoción	Neutro 1	Tristeza 1
F0	468	406

Rango del pitch	0.425-1.0	0.373-1.0
SPL	33.822	33.8
Rango SPL	24.862-35.704	24.445-35.546
Velocidad (palabras/min)	27.174	26.408
Energía	1127.501	1128.482
Valor medio	0.719	0.709
Varianza	0.009	0.01

Tabla 260. Actor 5, valores neutro y tristeza

Emoción	Neutro Media	Tristeza Media	Comparación	Variación
F0	468	406	86,75213675	0,86752137
Rango del pitch	0,575	0,627	109,0434783	1,09043478
SPL	33,822	33,8	99,93495358	0,99934954
Rango SPL	10,842	11,101	102,3888581	1,02388858
Velocidad (palabras/min)	27,174	26408	97181,12902	971,81129
Energía	1127,501	1128,482	100,0870066	1,00087007
Valor medio	0,719	0,709	98,60917942	0,98609179
Varianza	0,009	0,01	111,1111111	1,11111111

Tabla 261. Actor 5, comparaciones entre neutro y tristeza

6ª PERSONA

- **Alegría**

Emoción	Neutro 1	Alegre 1	Alegre 2
F0	781	500	437
Rango del pitch	0.283-1.0	0.379-1.0	0.319-1.0
SPL	33.65	33.751	33.676
Rango SPL	24.052-35.732	24.554-35.724	24.616-35.835
Velocidad (palabras/min)	21.186	24.351	18.564
Energía	1188.824	1166.375	1386.128
Valor medio	0.647	0.689	0.655
Varianza	0.018	0.014	0.014

Tabla 262. Actor 6, valores neutro y alegría

Emoción	Neutro Media	Alegría Media	Comparación	Variación
F0	781	468,5	59,9871959	0,59987196
Rango del pitch	0,717	0,651	90,7949791	0,90794979
SPL	33,65	33,7135	100,188707	1,00188707
Rango SPL	11,68	11,1945	95,8433219	0,95843322
Velocidad (palabras/min)	21,186	21,4575	101,281507	1,01281507
Energía	1188,824	1276,2515	107,354116	1,07354116
Valor medio	0,647	0,672	103,863988	1,03863988
Varianza	0,018	0,014	77,7777778	0,77777778

Tabla 263. Actor 6, comparaciones entre neutro y alegría

- **Enfado**

Emoción	Neutro 1	Enfado 1	Enfado 2
F0	781	437	531
Rango del pitch	0.283-1.0	0.301-1.0	0.291-1.0
SPL	33.65	33.691	33.701
Rango SPL	24.052-35.732	24.743-35.783	24.43-35.739
Velocidad (palabras/min)	21.186	14.706	22.321
Energía	1188.824	1758.822	1196.309
Valor medio	0.647	0.66	0.667
Varianza	0.018	0.012	0.015

Tabla 264. Actor 6, valores neutro y enfado

Emoción	Neutro Media	Enfado Media	Comparación	Variación
F0	781	484	61,971831	0,61971831
Rango del pitch	0,717	0,651	90,7949791	0,90794979
SPL	33,65	33,696	100,136701	1,00136701
Rango SPL	11,68	11,1945	95,8433219	0,95843322
Velocidad (palabras/min)	21,186	18,5135	87,3855376	0,87385538
Energía	1188,824	1477,5655	124,287994	1,24287994
Valor medio	0,647	0,6635	102,550232	1,02550232
Varianza	0,018	0,0135	75	0,75

Tabla 265. Actor 6, comparaciones entre neutro y enfado

- Miedo

Emoción	Neutro 1	Miedo 1	Miedo 2
F0	781	750	718
Rango del pitch	0.283-1.0	0.31-1.0	0.493-1.0
SPL	33.65	33.676	33.862
Rango SPL	24.052-35.732	24.175-35.695	25.033-35.8
Velocidad (palabras/min)	21.186	13.838	24.834
Energía	1188.824	1823.716	1279.446
Valor medio	0.647	0.652	0.738
Varianza	0.018	0.01	0.006

Tabla 266. Actor 6, valores neutro y miedo

Emoción	Neutro Media	Miedo Media	Comparación	Variación
F0	781	734	93,98207426	0,93982074
Rango del pitch	0,717	0,599	83,54253835	0,83542538
SPL	33,65	33,769	100,3536404	1,0035364
Rango SPL	11,68	11,1435	95,40667808	0,95406678
Velocidad (palabras/min)	21,186	19,336	91,26781837	0,91267818
Energía	1188,824	1551,581	130,5139365	1,30513936
Valor medio	0,647	0,695	107,4188563	1,07418856
Varianza	0,018	0,008	44,44444444	0,44444444

Tabla 267. Actor 6, comparaciones entre neutro y miedo

- Tristeza

Emoción	Neutro 1	Tristeza 1
F0	781	656
Rango del pitch	0.283-1.0	0.359-1.0
SPL	33.65	33.719
Rango SPL	24.052-35.732	24.654-35.726
Velocidad (palabras/min)	21.186	27.174
Energía	1188.824	2181.228
Valor medio	0.647	0.671
Varianza	0.018	0.01

Tabla 268. Actor 6, valores neutro y tristeza

Emoción	Neutro Media	Tristeza Media	Comparación	Variación
F0	781	656	83,99487836	0,83994878
Rango del pitch	0,717	0,641	89,40027894	0,89400279
SPL	33,65	33,719	100,205052	1,00205052
Rango SPL	11,68	11,072	94,79452055	0,94794521
Velocidad (palabras/min)	21,186	27,174	128,2639479	1,28263948
Energía	1188,824	2181,228	183,4777898	1,8347779
Valor medio	0,647	0,671	103,7094281	1,03709428
Varianza	0,018	0,01	55,55555556	0,55555556

Tabla 269. Actor 6, comparaciones entre neutro y tristeza

7ª PERSONA

- Alegría

Emoción	Neutro 1	Alegre 1	Alegre 2
F0	718	593	531
Rango del pitch	0.296-1.0	0.363-1.0	0.351-1.0
SPL	33.678	33.715	33.688
Rango SPL	24.397-35.685	24.434-35.76	24.701-35.77
Velocidad (palabras/min)	33.186	86.806	24.272
Energía	1526.352	1531.995	2094.771
Valor medio	0.653	0.669	0.656
Varianza	0.009	0.01	0.008

Tabla 270. Actor 7, valores neutro y alegría

Emoción	Neutro Media	Alegría Media	Comparación	Variación
F0	718	562	78,2729805	0,78272981
Rango del pitch	0,704	0,643	91,3352273	0,91335227
SPL	33,678	33,7015	100,069778	1,00069778
Rango SPL	11,68	11,198	95,8732877	0,95873288
Velocidad (palabras/min)	33,186	55,539	167,356717	1,67356717
Energía	1526,352	1813,383	118,805033	1,18805033
Valor medio	0,653	0,6625	101,454824	1,01454824
Varianza	0,009	0,009	100	1

Tabla 271. Actor 7, comparaciones entre neutro y alegría

- **Enfado**

Emoción	Neutro 1	Enfado 1	Enfado 2
F0	718	625	500
Rango del pitch	0.296-1.0	0.404-1.0	0.318-1.0
SPL	33.678	33.752	33.727
Rango SPL	24.397-35.685	24.746-35.623	24.707-35.762
Velocidad (palabras/min)	33.186	81.878	51.724
Energía	1526.352	1690.122	1043.818
Valor medio	0.653	0.685	0.676
Varianza	0.009	0.008	0.011

Tabla 272. Actor 7, valores neutro y enfado

Emoción	Neutro Media	Enfado Media	Comparación	Variación
F0	718	562,5	78,3426184	0,78342618
Rango del pitch	0,704	0,639	90,7670455	0,90767045
SPL	33,678	33,7395	100,182612	1,00182612
Rango SPL	11,68	10,966	93,8869863	0,93886986
Velocidad (palabras/min)	33,186	66,801	201,292714	2,01292714
Energía	1526,352	1366,97	89,5579788	0,89557979
Valor medio	0,653	0,6805	104,211332	1,04211332
Varianza	0,009	0,0095	105,555556	1,05555556

Tabla 273. Actor 7, comparaciones entre neutro y enfado

- **Miedo**

Emoción	Neutro 1	Miedo 1	Miedo 2
F0	718	250	187
Rango del pitch	0.296-1.0	0.416-1.0	0.43-1.0
SPL	33.678	33.769	33.818
Rango SPL	24.397-35.685	24.928-35.729	24.83-35.848
Velocidad (palabras/min)	33.186	15.306	15.957
Energía	1526.352	1838.844	1877.628
Valor medio	0.653	0.692	0.715
Varianza	0.009	0.007	0.006

Tabla 274. Actor 7, valores neutro y miedo

Emoción	Neutro Media	Miedo Media	Comparación	Variación
F0	718	218,5	30,4317549	0,30431755
Rango del pitch	0,704	0,577	81,9602273	0,81960227
SPL	33,678	33,7935	100,342954	1,00342954
Rango SPL	11,68	10,9095	93,4032534	0,93403253
Velocidad (palabras/min)	33,186	15,6315	47,1026939	0,47102694
Energía	1526,352	1858,236	121,743608	1,21743608
Valor medio	0,653	0,7035	107,733538	1,07733538
Varianza	0,009	0,0065	72,2222222	0,72222222

Tabla 275. Actor 7, comparaciones entre neutro y miedo

- Tristeza

Emoción	Neutro 1	Triste 1	Triste 2
F0	718	343	312
Rango del pitch	0.296-1.0	0.32-0.98	0.385-1.0
SPL	33.678	33.704	33.767
Rango SPL	24.397-35.685	24.495-35.62	24.683-35.844
Velocidad (palabras/min)	33.186	15.369	13.838
Energía	1526.352	1696.801	2034.171
Valor medio	0.653	0.663	0.692
Varianza	0.009	0.008	0.007

Tabla 276. Actor 7, valores neutro y tristeza

Emoción	Neutro Media	Tristeza Media	Comparación	Variación
F0	718	327,5	45,61281337	0,45612813
Rango del pitch	0,704	0,648	92,04545455	0,92045455
SPL	33,678	33,7355	100,1707346	1,00170735
Rango SPL	11,68	11,143	95,40239726	0,95402397
Velocidad (palabras/min)	33,186	14,6035	44,00500211	0,44005002
Energía	1526,352	1865,486	122,218597	1,22218597
Valor medio	0,653	0,6775	103,7519142	1,03751914
Varianza	0,009	0,0075	83,33333333	0,83333333

Tabla 277. Actor 7, comparaciones entre neutro y tristeza

8ª PERSONA

- Alegría

Emoción	Neutro 1	Neutro 2	Alegre 1	Alegre 2
F0	625	531	468	500
Rango del pitch	0.286-1.0	0.329-1.0	0.307-1.0	0.244-1.0
SPL	33.674	33.696	33.644	33.57
Rango SPL	24.415-35.552	24.199-35.568	24.058-35.567	23.807-35.709
Velocidad (palabras/min)	48.387	63.559	25.0	24.834
Energía	1043.649	834.328	1000.762	935.983
Valor medio	0.653	0.665	0.644	0.617
Varianza	0.011	0.014	0.016	0.022

Tabla 278. Actor 8, valores neutro y alegría

Emoción	Neutro Media	Alegría Media	Comparación	Variación
F0	578	484	83,7370242	0,83737024
Rango del pitch	0,693	0,725	104,617605	1,04617605
SPL	33,685	33,607	99,7684429	0,99768443
Rango SPL	11,253	11,7055	104,02115	1,0402115
Velocidad (palabras/min)	55,973	24,917	44,516106	0,44516106
Energía	938,9885	968,3725	103,129325	1,03129325
Valor medio	0,659	0,6305	95,6752656	0,95675266
Varianza	0,0125	0,019	152	1,52

Tabla 279. Actor 8, comparaciones entre neutro y alegría

- Enfado

Emoción	Neutro 1	Neutro 2	Enfado 1	Enfado 2
F0	625	531	468	437
Rango del pitch	0.286-1.0	0.329-1.0	0.331-1.0	0.314-1.0
SPL	33.674	33.696	33.66	33.738
Rango SPL	24.415-35.552	24.199-35.568	24.017-35.572	24.55-35.641
Velocidad (palabras/min)	48.387	63.559	26.786	86.207
Energía	1043.649	834.328	951.847	1291.825
Valor medio	0.653	0.665	0.65	0.683
Varianza	0.011	0.014	0.016	0.013

Tabla 280. Actor 8, valores neutro y enfado

Emoción	Neutro Media	Enfado Media	Comparación	Variación
F0	578	452,5	78,2871972	0,78287197
Rango del pitch	0,693	0,678	97,8354978	0,97835498
SPL	33,685	33,699	100,041562	1,00041562
Rango SPL	11,253	11,3225	100,617613	1,00617613
Velocidad (palabras/min)	55,973	56,4965	100,935272	1,00935272
Energía	938,9885	1121,836	119,472816	1,19472816
Valor medio	0,659	0,6665	101,138088	1,01138088
Varianza	0,0125	0,0145	116	1,16

Tabla 281. Actor 8, comparaciones entre neutro y enfado

- Miedo

Emoción	Neutro 1	Neutro 2	Miedo 1	Miedo 2
F0	625	531	625	593
Rango del pitch	0.286-1.0	0.329-1.0	0.435-1.0	0.314-1.0
SPL	33.674	33.696	33.809	33.66
Rango SPL	24.415-35.552	24.199-35.568	24.809-35.695	24.073-35.645
Velocidad (palabras/min)	48.387	63.559	94.34	53.571
Energía	1043.649	834.328	1276.739	964.067
Valor medio	0.653	0.665	0.714	0.652
Varianza	0.011	0.014	0.011	0.019

Tabla 282. Actor 8, valores neutro y miedo

Emoción	Neutro Media	Miedo Media	Comparación	Variación
F0	578	609	105,363322	1,05363322
Rango del pitch	0,693	0,626	90,3318903	0,9033189
SPL	33,685	33,7345	100,14695	1,0014695
Rango SPL	11,253	11,229	99,7867235	0,99786724
Velocidad (palabras/min)	55,973	73,9555	132,127097	1,32127097
Energía	938,9885	1120,403	119,320205	1,19320205
Valor medio	0,659	0,683	103,641882	1,03641882
Varianza	0,0125	0,015	120	1,2

Tabla 283. Actor 8, comparaciones entre neutro y tristeza

- **Tristeza**

Emoción	Neutro 1	Neutro 2	Tristeza 1	Tristeza 2
F0	625	531	718	656
Rango del pitch	0.286-1.0	0.329-1.0	0.383-1.0	0.323-1.0
SPL	33.674	33.696	33.773	33.67
Rango SPL	24.415-35.552	24.199-35.568	24.749-35.684	24.146-35.652
Velocidad (palabras/min)	48.387	63.559	28.736	13.889
Energía	1043.649	834.328	2007.807	1830.95
Valor medio	0.653	0.665	0.698	0.652
Varianza	0.011	0.014	0.012	0.013

Tabla 284. Actor 8, valores neutro y tristeza

Emoción	Neutro Media	Tristeza Media	Comparación	Variación
F0	578	687	118,858131	1,18858131
Rango del pitch	0,693	0,647	93,3621934	0,93362193
SPL	33,685	33,7215	100,108357	1,00108357
Rango SPL	11,253	11,2205	99,7111881	0,99711188
Velocidad (palabras/min)	55,973	21,3125	38,076394	0,38076394
Energía	938,9885	1919,3785	204,409159	2,04409159
Valor medio	0,659	0,675	102,427921	1,02427921
Varianza	0,0125	0,0125	100	1

Tabla 285. Actor 8, comparaciones entre neutro y tristeza

9ª PERSONA

- **Alegría**

Emoción	Neutro 1	Alegre 1
F0	312	906
Rango del pitch	0.252-0.98	0.39-1.0
SPL	33.718	33.716
Rango SPL	24.007-35.738	24.619-35.795
Velocidad (palabras/min)	32.051	11.364
Energía	842.785	2326.81
Valor medio	0.674	0.669
Varianza	0.014	0.008

Tabla 286. Actor 9, valores neutro y alegría

Emoción	Neutro Media	Alegría Media	Comparación	Variación
F0	312	906	290,3846154	2,90384615
Rango del pitch	0,728	0,61	83,79120879	0,83791209
SPL	33,718	33,716	99,99406845	0,99994068
Rango SPL	11,731	11,176	95,26894553	0,95268946
Velocidad (palabras/min)	32,051	11,364	35,45599201	0,35455992
Energía	842,785	2326,81	276,0858345	2,76085834
Valor medio	0,674	0,669	99,25816024	0,9925816
Varianza	0,014	0,008	57,14285714	0,57142857

Tabla 287. Actor 9, comparaciones entre neutro y alegría

- **Enfado**

Emoción	Neutro 1	Enfado 1
F0	312	781
Rango del pitch	0.252-0.98	0.339-0.97
SPL	33.718	33.629
Rango SPL	24.007-35.738	24.299-35.83
Velocidad (palabras/min)	32.051	49.02
Energía	842.785	984.166
Valor medio	0.674	0.635
Varianza	0.014	0.015

Tabla 288. Actor 9, valores neutro y enfado

Emoción	Neutro Media	Enfado Media	Comparación	Variación
F0	312	781	250,320513	2,50320513
Rango del pitch	0,728	0,631	86,6758242	0,86675824
SPL	33,718	33,629	99,736046	0,99736046
Rango SPL	11,731	11,531	98,2951155	0,98295116
Velocidad (palabras/min)	32,051	49,02	152,943746	1,52943746
Energía	842,785	984,166	116,775453	1,16775453
Valor medio	0,674	0,635	94,2136499	0,9421365
Varianza	0,014	0,015	107,142857	1,07142857

Tabla 289. Actor 9, comparaciones entre neutro y enfado

- **Miedo**

Emoción	Neutro 1	Miedo 1
F0	312	500
Rango del pitch	0.252-0.98	0.329-1.0
SPL	33.718	33.623
Rango SPL	24.007-35.738	24.455-35.71
Velocidad (palabras/min)	32.051	34.722
Energía	842.785	688.534
Valor medio	0.674	0.631
Varianza	0.014	0.013

Tabla 290. Actor 9, valores neutro y miedo

Emoción	Neutro Media	Miedo Media	Comparación	Variación
F0	312	500	160,2564103	1,6025641
Rango del pitch	0,728	0,671	92,17032967	0,9217033
SPL	33,718	33,623	99,71825138	0,99718251
Rango SPL	11,731	11,255	95,9423749	0,95942375
Velocidad (palabras/min)	32,051	34,722	108,3335933	1,08333593
Energía	842,785	688,534	81,69746733	0,81697467
Valor medio	0,674	0,631	93,62017804	0,93620178
Varianza	0,014	0,013	92,85714286	0,92857143

Tabla 291. Actor 9, comparaciones entre neutro y miedo

- **Tristeza**

Emoción	Neutro 1	Tristeza 1
F0	312	468
Rango del pitch	0.252-0.98	0.333-1.0
SPL	33.718	33.722
Rango SPL	24.007-35.738	24.702-35.724
Velocidad (palabras/min)	32.051	19.133
Energía	842.785	1413.647
Valor medio	0.674	0.674
Varianza	0.014	0.011

Tabla 292. Actor 9, valores neutro y tristeza

Emoción	Neutro Media	Tristeza Media	Comparación	Variación
F0	312	468	150	1,5
Rango del pitch	0,728	0,667	91,6208791	0,91620879
SPL	33,718	33,722	100,011863	1,00011863
Rango SPL	11,731	11,022	93,9561845	0,93956184
Velocidad (palabras/min)	32,051	19,133	59,6954853	0,59695485
Energía	842,785	1413,647	167,735188	1,67735188
Valor medio	0,674	0,674	100	1
Varianza	0,014	0,011	78,5714286	0,78571429

Tabla 293. Actor 9, comparaciones entre neutro y tristeza

10ª PERSONA

- Alegría

Emoción	Neutro 1	Alegre 1
F0	437	625
Rango del pitch	0.395-1.0	0.312-1.0
SPL	33.717	33.524
Rango SPL	24.775-35.717	24.581-35.74
Velocidad (palabras/min)	73.529	37.879
Energía	1793.168	559.507
Valor medio	0.669	0.593
Varianza	0.008	0.016

Tabla 294. Actor 10, valores neutro y alegría

Emoción	Neutro Media	Alegría Media	Comparación	Variación
F0	437	625	143,020595	1,43020595
Rango del pitch	0,605	0,688	113,7190083	1,13719008
SPL	33717	33524	99,42758846	0,99427588
Rango SPL	10,942	11,159	101,9831841	1,01983184
Velocidad (palabras/min)	73,529	37,879	51,51572849	0,51515728
Energía	1793,168	559,507	31,20215172	0,31202152
Valor medio	0,669	0,593	88,63976084	0,88639761
Varianza	0,008	0,016	200	2

Tabla 295. Actor 10, comparaciones entre neutro y alegría

- **Enfado**

Emoción	Neutro 1	Enfado 1
F0	437	437
Rango del pitch	0.395-1.0	0.323-1.0
SPL	33.717	33.643
Rango SPL	24.775-35.717	24.569-35.678
Velocidad (palabras/min)	73.529	51.37
Energía	1793.168	944.758
Valor medio	0.669	0.638
Varianza	0.008	0.01

Tabla 296. Actor 10, valores neutro y enfado

Emoción	Neutro Media	Enfado Media	Comparación	Variación
F0	437	437	100	1
Rango del pitch	0,605	0,677	111,900826	1,11900826
SPL	33717	33643	99,7805261	0,99780526
Rango SPL	10,942	11,109	101,526229	1,01526229
Velocidad (palabras/min)	73,529	51,37	69,8635912	0,69863591
Energía	1793,168	944,758	52,6865302	0,5268653
Valor medio	0,669	0,638	95,3662182	0,95366218
Varianza	0,008	0,01	125	1,25

Tabla 297. Actor 10, comparaciones entre neutro y enfado

- **Miedo**

Emoción	Neutro 1	Miedo 1	Miedo 2
F0	437	437	468
Rango del pitch	0.395-1.0	0.336-1.0	0.41-1.0
SPL	33.717	33.701	33.798
Rango SPL	24.775-35.717	24.601-35.701	24.975-35.769
Velocidad (palabras/min)	73.529	41.322	25.685
Energía	1793.168	2537.821	2294.897
Valor medio	0.669	0.664	0.707
Varianza	0.008	0.011	0.008

Tabla 298. Actor 10, valores neutro y miedo

Emoción	Neutro Media	Miedo Media	Comparación	Variación
F0	437	452,5	103,546911	1,03546911
Rango del pitch	0,605	0,627	103,636364	1,03636364
SPL	33,717	33,7495	100,096391	1,00096391
Rango SPL	10,942	10,947	100,045695	1,00045695
Velocidad (palabras/min)	73,529	33,5035	45,5650152	0,45565015
Energía	1793,168	2416,359	134,753632	1,34753632
Valor medio	0,669	0,6855	102,466368	1,02466368
Varianza	0,008	0,0095	118,75	1,1875

Tabla 299. Actor 10, comparaciones entre neutro y miedo

- Tristeza

Emoción	Neutro 1	Tristeza 1
F0	437	468
Rango del pitch	0.395-1.0	0.36-1.0
SPL	33.717	33.644
Rango SPL	24.775-35.717	23.848-35.669
Velocidad (palabras/min)	73.529	46.182
Energía	1793.168	2606.084
Valor medio	0.669	0.637
Varianza	0.008	0.008

Tabla 300. Actor 10, valores neutro y tristeza

Emoción	Neutro Media	Tristeza Media	Comparación	Variación
F0	437	468	107,0938215	1,07093822
Rango del pitch	0,605	0,64	105,785124	1,05785124
SPL	33717	33644	99,78349201	0,99783492
Rango SPL	10,942	11,821	108,0332663	1,08033266
Velocidad (palabras/min)	73,529	46,182	62,80787172	0,62807872
Energía	1793,168	2606,084	145,334068	1,45334068
Valor medio	0,669	0,637	95,21674141	0,95216741
Varianza	0,008	0,008	100	1

Tabla 301. Actor 10, comparaciones entre neutro y tristeza

11ª PERSONA

- **Alegría**

Emoción	Neutro 1	Alegre 1
F0	250	218
Rango del pitch	0.247-1.0	0.16-1.0
SPL	33.558	33.497
Rango SPL	24.46-35.569	24.193-35.634
Velocidad (palabras/min)	24.671	56.25
Energía	889.34	1124.754
Valor medio	0.604	0.586
Varianza	0.013	0.02

Tabla 302. Actor 11, valores neutro y alegría

Emoción	Neutro Media	Alegría Media	Comparación	Variación
F0	250	218	87,2	0,872
Rango del pitch	0,753	0,84	111,5537849	1,11553785
SPL	33,558	33,497	99,81822516	0,99818225
Rango SPL	11,109	11,441	102,9885678	1,02988568
Velocidad (palabras/min)	24,671	56,25	228,0004864	2,28000486
Energía	889,34	1124,754	126,4706411	1,26470641
Valor medio	0,604	0,586	97,01986755	0,97019868
Varianza	0,013	0,02	153,8461538	1,53846154

Tabla 303. Actor 11, comparaciones entre neutro y alegría

- **Enfado**

Emoción	Neutro 1	Enfado 1
F0	250	312
Rango del pitch	0.247-1.0	0.28-1.0
SPL	33.558	33.704
Rango SPL	24.46-35.569	24.595-35.785
Velocidad (palabras/min)	24.671	78.947
Energía	889.34	1306.81
Valor medio	0.604	0.649
Varianza	0.013	0.022

Tabla 304. Actor 11, valores neutro y enfado

Emoción	Neutro Media	Enfado Media	Comparación	Variación
F0	250	312	124,8	1,248
Rango del pitch	0,753	0,72	95,61752988	0,9561753
SPL	33,558	33,704	100,4350676	1,00435068
Rango SPL	11,109	11,19	100,7291385	1,00729139
Velocidad (palabras/min)	24,671	78,947	319,9991893	3,19999189
Energía	889,34	1306,81	146,9415522	1,46941552
Valor medio	0,604	0,649	107,4503311	1,07450331
Varianza	0,013	0,022	169,2307692	1,69230769

Tabla 305. Actor 11, comparaciones entre neutro y enfado

- Miedo

Emoción	Neutro 1	Miedo 1
F0	250	250
Rango del pitch	0.247-1.0	0.428-1.0
SPL	33.558	33.82
Rango SPL	24.46-35.569	25.044-35.716
Velocidad (palabras/min)	24.671	10.108
Energía	889.34	2967.382
Valor medio	0.604	0.716
Varianza	0.013	0.005

Tabla 306. Actor 11, valores neutro y miedo

Emoción	Neutro Media	Miedo Media	Comparación	Variación
F0	250	250	100	1
Rango del pitch	0,753	0,572	75,96281541	0,75962815
SPL	33,558	33,82	100,7807378	1,00780738
Rango SPL	11,109	10,672	96,06625259	0,96066253
Velocidad (palabras/min)	24,671	10,108	40,97118074	0,40971181
Energía	889,34	2967,382	333,661142	3,33661142
Valor medio	0,604	0,716	118,5430464	1,18543046
Varianza	0,013	0,005	38,46153846	0,38461538

Tabla 307. Actor 11, comparaciones entre neutro y miedo

- **Tristeza**

Emoción	Neutro 1	Tristeza 1
F0	250	375
Rango del pitch	0.247-1.0	0.341-0.97
SPL	33.558	33.704
Rango SPL	24.46-35.569	24.595-35.785
Velocidad (palabras/min)	24.671	11.329
Energía	889.34	2280.044
Valor medio	0.604	0.663
Varianza	0.013	0.007

Tabla 308. Actor 11, valores neutro y tristeza

Emoción	Neutro Media	Tristeza Media	Comparación	Variación
F0	250	375	150	1,5
Rango del pitch	0,753	0,629	83,53253652	0,83532537
SPL	33,558	33,704	100,4350676	1,00435068
Rango SPL	11,109	11,19	100,7291385	1,00729139
Velocidad (palabras/min)	24,671	11,329	45,9203113	0,45920311
Energía	889,34	2280,044	256,3748398	2,5637484
Valor medio	0,604	0,663	109,7682119	1,09768212
Varianza	0,013	0,007	53,84615385	0,53846154

Tabla 309. Actor 11, comparaciones entre neutro y tristeza

11.2 Anexo 2

En este apartado se expondrán las matrices de confusión de los resultados. Estos resultados corresponden a las personas del conjunto de test y se realizaron para cada una de las emociones respecto a cada uno de los audios con emoción neutra.

Pruebas

1ª Persona

- **Prueba 1**

Archivo neutro: 13_06_2012-17_41_58.wav

Matriz de confusión de alegría	
TP = 2	FP = 3
FN = 1	TN = 7

Tabla 310. Actor 1 – prueba 1, matriz de confusión para alegría

Matriz de confusión de enfado	
TP = 1	FP = 0
FN = 2	TN = 10

Tabla 311. Actor 1 – prueba 1, matriz de confusión para enfado

Matriz de confusión de miedo	
TP = 0	FP = 2
FN = 3	TN = 8

Tabla 312. Actor 1 – prueba 1, matriz de confusión para miedo

Matriz de confusión de tristeza	
TP = 4	FP = 1
FN = 0	TN = 8

Tabla 313. Actor 1 – prueba 1, matriz de confusión para tristeza

- **Prueba 2**

Archivo neutro: 13_06_2012-17_42_12.wav

Matriz de confusión de alegría	
TP = 1	FP = 2
FN = 2	TN = 8

Tabla 314. Actor 1 – prueba 2, matriz de confusión para alegría

Matriz de confusión de enfado	
TP = 2	FP = 0
FN = 1	TN = 10

Tabla 315. Actor 1 – prueba 2, matriz de confusión para enfado

Matriz de confusión de miedo	
TP = 2	FP = 2
FN = 1	TN = 8

Tabla 316. Actor 1 – prueba 2, matriz de confusión para miedo

Matriz de confusión de tristeza	
TP = 4	FP = 0
FN = 0	TN = 9

Tabla 317. Actor 1 – prueba 2, matriz de confusión para tristeza

- **Prueba 3**

Archivo neutro: 13_06_2012-17_42_49.wav

Matriz de confusión de alegría	
TP = 2	FP = 2
FN = 1	TN = 8

Tabla 318. Actor 1 – prueba 3, matriz de confusión para alegría

Matriz de confusión de enfado	
TP = 1	FP = 3
FN = 2	TN = 7

Tabla 319. Actor 1 – prueba 3, matriz de confusión para enfado

Matriz de confusión de miedo	
TP = 3	FP = 2
FN = 0	TN = 8

Tabla 320. Actor 1 – prueba 3, matriz de confusión para miedo

Matriz de confusión de tristeza	
TP = 0	FP = 0
FN = 4	TN = 9

Tabla 321. Actor 1 – prueba 3, matriz de confusión para tristeza

2ª Persona

- **Prueba 4**

Archivo neutro: 16_07_2012-18_50_46.wav

Matriz de confusión de alegría	
TP = 1	FP = 2
FN = 1	TN = 4

Tabla 322. Actor 2 – prueba 4, matriz de confusión para alegría

Matriz de confusión de enfado	
TP = 1	FP = 1
FN = 1	TN = 5

Tabla 323. Actor 2 – prueba 4, matriz de confusión para enfado

Matriz de confusión de miedo	
TP = 1	FP = 0
FN = 1	TN = 6

Tabla 324. Actor 2 – prueba 4, matriz de confusión para miedo

Matriz de confusión de tristeza	
TP = 2	FP = 0
FN = 0	TN = 6

Tabla 325. Actor 2 – prueba 4, matriz de confusión para tristeza

- **Prueba 5**

Archivo neutro: 16_07_2012-18_47_47.wav

Matriz de confusión de alegría	
TP = 0	FP = 0
FN = 2	TN = 6

Tabla 326. Actor 2 – prueba 5, matriz de confusión para alegría

Matriz de confusión de enfado	
TP = 2	FP = 2
FN = 0	TN = 4

Tabla 327. Actor 2 – prueba 5, matriz de confusión para enfado

Matriz de confusión de miedo	
TP = 0	FP = 0
FN = 2	TN = 6

Tabla 328. Actor 2 – prueba 5, matriz de confusión para miedo

Matriz de confusión de tristeza	
TP = 2	FP = 2
FN = 0	TN = 4

Tabla 329. Actor 2 – prueba 5, matriz de confusión para tristeza

3ª Persona

- Prueba 6

Archivo neutro: 18_07_2012-17_8_44.wav

Matriz de confusión de alegría	
TP = 0	FP = 0
FN = 2	TN = 6

Tabla 330. Actor 3 – prueba 6, matriz de confusión para alegría

Matriz de confusión de enfado	
TP = 2	FP = 2
FN = 0	TN = 4

Tabla 331. Actor 3 – prueba 6, matriz de confusión para enfado

Matriz de confusión de miedo	
TP = 0	FP = 0
FN = 2	TN = 6

Tabla 332. Actor 3 – prueba 6, matriz de confusión para miedo

Matriz de confusión de tristeza	
TP = 2	FP = 2
FN = 0	TN = 4

Tabla 333. Actor 3 – prueba 6, matriz de confusión para tristeza

- Prueba 7

Archivo neutro: 18_07_2012-17_12_50.wav

Matriz de confusión de alegría	
TP = 0	FP = 0
FN = 2	TN = 6

Tabla 334. Actor 3 – prueba 7, matriz de confusión para alegría

Matriz de confusión de enfado	
TP = 2	FP = 3
FN = 0	TN = 3

Tabla 335. Actor 3 – prueba 7, matriz de confusión para enfado



Matriz de confusión de miedo	
TP = 0	FP = 0
FN = 2	TN = 6

Tabla 336. Actor 3 – prueba 7, matriz de confusión para miedo

Matriz de confusión de tristeza	
TP = 2	FP = 1
FN = 0	TN = 5

Tabla 337. Actor 3 – prueba 7, matriz de confusión para tristeza