



Universidad de Valladolid

E.T.S DE INGENIERÍA INFORMÁTICA

TRABAJO DE FIN DE GRADO

Grado en Ingeniería Informática

**Evaluación de la Entonación en la
Enseñanza del Español como Segunda
Lengua**

Alumno: Diego Vallejo Alonso

Tutor: Valentín Cardeñoso Payo

RESUMEN

El Español es el segundo idioma más hablado en el planeta; es un valor de referencia no sólo en el ámbito cultural, sino también en otros de marcado carácter científico y económico. El número de hablantes y estudiantes de Español en todo el mundo aumenta anualmente frente a otros idiomas de importancia como el inglés o el chino. Por ello, son necesarios sistemas de autoayuda que complementen el aprendizaje en las aulas para mejorar la competencia en el idioma. Este proyecto surge como continuación al trabajo realizado en el proyecto "Aplicación Móvil de Ayuda a la Pronunciación del Español", en el que se desarrolló una aplicación móvil para dispositivos de la plataforma Android que hacía uso de las tecnologías del habla (síntesis del habla y reconocimiento automático de la voz) para tratar de ayudar a los usuarios a mejorar la pronunciación de nuestro idioma. En esta ocasión nos hemos centrado en la otra gran rama, la entonación. La entonación es importante porque transmite otra clase de información que no podemos expresar únicamente a través del mensaje: las emociones. Cada idioma tiene sus patrones entonativos característicos, nosotros trataremos de ofrecer una ayuda para los estudiantes del Español que deseen mejorar su destreza y adquirir una entonación más propia de las personas nativas mediante el desarrollo de un módulo de evaluación de la entonación que compare las locuciones de los estudiantes extranjeros con la de unos locutores de referencia españoles, emitiendo una puntuación a la entonación de la locuciones realizadas. Para concluir el trabajo, integraremos el módulo en la aplicación UVaPronunciación desarrollada el año pasado de modo que dé soporte tanto a la mejora de la pronunciación como a la de la entonación, proporcionando un servicio más completo al usuario.

AGRADECIMIENTOS

Gracias a mis compañeros del Curso de Adaptación al Grado, porque han hecho que el ambiente de este año sea magnífico, a pesar de que ha sido un curso complicado. En especial quiero darle las gracias a Begoña porque no sólo me ha demostrado ser una gran compañera, sino también una gran amiga. Gracias a mis nuevos compañeros de trabajo, porque son fantásticos y me lo han puesto muy fácil para adaptarme desde el primer día. A todos vosotros os tengo que dar un millón de veces las gracias, pero espero que sepáis disculparme el que ahora quiera hacer algo más especial.

Gracias a MaPi por estar siempre pendiente de mí, y por apoyarme en todo lo que haga sea cuando sea. Gracias a David Soler, mi primer compañero de trabajo, y quizás el mejor que nunca tendré, no sólo es un gran informático, también es un amigo espectacular. Gracias a Álvaro, Antonio, Bachi, Jaime, Julio, Alberto, Chuchi, Chema, Santi y José, porque aunque estemos cada uno en una punta seguimos siendo un gran grupo, sois geniales. Infinitos millones de gracias a Valentín, porque él sólo ha conseguido que yo sacara este proyecto adelante, porque siempre ha estado pendiente de mí y porque cada minuto que hablo con él aprendo algo nuevo. Gracias a Noelia por estar siempre ahí, por preocuparse de mí, aguantarme y ser la más especial. Gracias a mi tío Javier por compartir conmigo el secreto para hacernos los amos del mundo, ya podemos ponernos en marcha. Y por último, gracias a mis tías, InésPi y Beatriz, a mi Lita y a mis padres porque a pesar del año que les he dado han seguido sin darme por perdido y me han seguido apoyando más aún si cabe cada día. Os quiero. Os lo debo todo. Ahora Mamá, creo que ya voy a poder descansar un poco...

Índice

ÍNDICE	7
CAPÍTULO 1	13
INTRODUCCIÓN.....	13
1.1. <i>Motivación</i>	13
1.1.1. Importancia del Español.....	14
1.1.2. La Entonación en la Enseñanza del Español.....	14
1.1.3. Ampliación del Prototipo Inicial.....	15
1.2. <i>Esquema de la memoria</i>	16
CAPÍTULO 2	19
PLAN DE TRABAJO.....	19
2.1. <i>Visión General del Proyecto</i>	19
2.1.1. Descripción del Problema.....	19
2.1.2. Objetivos del Proyecto.....	19
2.1.3. Suposiciones y Restricciones.....	20
2.2. <i>Gestión del Proceso</i>	21
2.2.1. Planificación del Proyecto en Fases.....	21
2.2.2. Objetivos de las Fases.....	21
2.2.2.1. Fase de Estudio Inicial.....	22
2.2.2.2. Preparación del Corpus de Entrenamiento.....	22
2.2.2.3. Diseño y Desarrollo de un Módulo de Evaluación de la Entonación e Integración en una Aplicación Móvil Prototipo.....	22
2.2.2.4. Edición de la documentación elaborada y preparación de la memoria final.....	22
CAPÍTULO 3	23
LA EVALUACIÓN DE LA ENTONACIÓN EN LA ENSEÑANZA DEL ESPAÑOL.....	23
3.1. <i>Prosodia y Entonación</i>	23
3.2. <i>La Enseñanza de la Prosodia del Español</i>	24
3.3. <i>Evaluación de la pronunciación, Plan Curricular y Diplomas de Español</i>	25
3.4. <i>Aspectos de la Pronunciación en Español: Patrones de Entonación</i>	28
CAPÍTULO 4	35
MARCO CONCEPTUAL: ASPECTOS CLAVE PARA EL ANÁLISIS DE LA ENTONACIÓN.....	35
4.1. <i>Conceptos Generales de Interés</i>	35
4.2. <i>Descripción de la Entonación en la Enseñanza del idioma</i>	37
4.3. <i>Teorías Fonológicas de la Prosodia</i>	38
4.3.1. Teoría de Configuración.....	38
4.3.2. Teoría de la secuencia de tonos.....	38
4.4. <i>Análisis Acústico-Prosódico</i>	39
4.4.1. La onda sonora.....	40
4.4.2. Medidas de Duración.....	40
4.4.3. Frecuencia Fundamental del Habla (F0).....	41
4.4.4. Energía.....	42
4.4.5. Caracterización del habla con vectores de parámetros.....	42
CAPÍTULO 5	43
PREPARACIÓN DEL CORPUS.....	43
5.1. <i>¿Qué es un corpus?</i>	43

Evaluación de la Entonación en la enseñanza del Español como Segunda Lengua

5.2. Recopilación del Corpus.....	43
5.2.1. Protocolo de grabación	43
5.3. Contenidos del Corpus.....	44
5.3.1. Locutores.....	44
5.3.2. Frases	46
5.3.3. Estructura de Directorios.....	46
5.3.4. Ficheros de Datos	47
5.4. Procesado del Corpus	49
5.4.1. Etiquetado de las grabaciones con Praat	49
5.4.2. Procesamiento de las pistas	54
5.4.2.1. Extracción de archivos de sonido de las frases procedentes de las pistas de audio etiquetadas. 55	
5.4.2.2. Etiquetado de los nuevos archivos de las frases con la frase correspondiente.....	58
5.4.2.3. Extracción de los valores de frecuencia fundamental (F0) y energía (Intensity) de los nuevos archivos de audio.....	59
5.4.3. Generación de los ficheros con formato .f0p	63
5.4.4. Obtención de información de los archivos de audio con Snack	64
5.4.5. Generación de los ficheros con formato .f0a	69
5.4.6. Generación de información elaborada a partir de los datos de los ficheros .f0p y .f0a (ficheros .ppar y .spar).....	70
5.4.7. Cálculo de distancias con el algoritmo DTW.....	72
5.4.7.1. Algoritmo DTW.....	73
5.4.8. Generación de ficheros .csv.....	80
5.4.9. Estudio de las características extraídas con WEKA.....	81
5.4.9.1. Experimentos realizados para la selección de atributos	93
5.4.9.1.1. MODALIDAD 1: Pesos deducidos del análisis del fichero de datos globales (estáticos) contenidos con Praat considerando todas las muestras.....	94
5.4.9.1.2. MODALIDAD 2: Pesos deducidos del análisis del fichero de datos globales (estáticos) obtenidos con Praat distinguiendo sexos	95
A. LOCUTORES FEMENINOS.....	95
B. LOCUTORES MASCULINOS.....	96
5.4.9.1.3. MODALIDAD 3: Pesos deducidos del análisis del fichero de datos locales (dinámicos) obtenidos con Praat considerando todas las muestras	97
5.4.9.1.4. MODALIDAD 4: Pesos deducidos del análisis del fichero de datos locales (dinámicos) obtenidos con Praat distinguiendo sexos	98
A. LOCUTORES FEMENINOS.....	98
B. LOCUTORES MASCULINOS.....	99
5.4.9.1.5. MODALIDAD 5: Pesos deducidos del análisis del fichero de datos locales (dinámicos) obtenidos con Praat distinguiendo intervalos sonoros y sordos	100
A. INTERVALOS VOICED	100
B. INTERVALOS UNVOICED	102
5.4.9.1.6. MODALIDAD 6: Pesos deducidos del análisis del fichero de datos locales (dinámicos) obtenidos con Praat distinguiendo intervalos sonoros y sordos y distinguiendo sexos	103
A. LOCUTORES FEMENINOS E INTERVALOS VOICED.....	103
B. LOCUTORES FEMENINOS E INTERVALOS UNVOICED.....	104
C. LOCUTORES MASCULINOS E INTERVALOS VOICED.....	105
D. LOCUTORES MASCULINOS E INTERVALOS UNVOICED	106
5.4.9.1.7. MODALIDAD 7: Pesos deducidos del análisis del fichero de datos globales (estáticos) obtenidos con Snack considerando todas las muestras.....	107
5.4.9.1.8. MODALIDAD 8: Pesos deducidos del análisis del fichero de datos globales (estáticos) obtenidos con Snack distinguiendo sexos.....	108
A. LOCUTORES FEMENINOS.....	109
B. LOCUTORES MASCULINOS.....	110

5.4.9.1.9. MODALIDAD 9: Pesos deducidos del análisis del fichero de datos locales (dinámicos) obtenidos con Snack considerando todas las muestras	111
5.4.9.1.10. MODALIDAD 10: Pesos deducidos del análisis del fichero de datos locales (dinámicos) obtenidos con Snack distinguiendo sexos	112
A. LOCUTORES FEMENINOS	112
B. LOCUTORES MASCULINOS	113
5.4.9.1.11. MODALIDAD 11: Pesos deducidos del análisis del fichero de datos locales (dinámicos) obtenidos con Snack distinguiendo intervalos sonoros y sordos	114
A. INTERVALOS VOICED	114
B. INTERVALOS UNVOICED	115
5.4.9.1.12. MODALIDAD 12: Pesos deducidos del análisis del fichero de datos locales (dinámicos) obtenidos con Snack distinguiendo intervalos sonoros y sordos y distinguiendo sexos	116
A. LOCUTORES FEMENINOS E INTERVALOS VOICED	116
B. LOCUTORES FEMENINOS E INTERVALOS UNVOICED	117
C. LOCUTORES MASCULINOS E INTERVALOS VOICED	119
D. LOCUTORES MASCULINOS E INTERVALOS UNVOICED	120
5.4.10. Repetición del cálculo de distancias con el algoritmo DTW	121
5.4.11. Prueba del módulo de evaluación con los pesos deducidos en las pruebas y elección de la ponderación más adecuada	122
5.4.12. Sistema de Puntuación (Scoring)	127
CAPÍTULO 6	129
APLICACIÓN FINAL	129
6.1. <i>Análisis</i>	129
6.1.1. Requisitos del sistema	129
6.1.1.1. Requisitos funcionales	129
6.1.1.2. Requisitos No Funcionales	130
6.1.2. Modelo de Casos de Uso	131
6.1.2.1. Descripción General de los Actores	131
6.1.2.2. Diagrama del Modelo de Casos de Uso	132
6.1.2.3. Descripción de los Casos de Uso	132
6.1.3. Realización de los casos de Uso	134
6.1.3.1. Diagrama de Clases	134
6.1.3.2. Descripción de las Clases	134
6.1.3.3. Diagramas de Secuencia	136
6.2. <i>Diseño</i>	136
6.2.1. Arquitectura propuesta	136
6.2.1.1. Diseño de la Arquitectura	136
6.2.1.1.1. Topología del Sistema (Diagrama de Despliegue)	136
6.2.1.1.2. Visión Global	137
6.2.1.2.1. Diseño de las capas del Sistema	138
6.2.2. Modelo de Casos de Uso de Diseño	141
6.2.2.1. Descripción de los Casos de Uso	141
6.2.3. Realización de los Casos de Uso de Diseño	142
6.2.3.1. Diagrama de Clases	143
6.2.3.2. Diagramas de Secuencia	144
6.2.3.2.1. Diagrama de Secuencia de Diseño 1: Cargar Frase	145
6.2.3.2.2. Diagrama de Secuencia de Diseño 2: Escuchar Frase	146
6.2.3.2.3. Diagrama de Secuencia de Diseño 3: Grabar Ensayo	147
6.2.3.2.4. Diagrama de Secuencia de Diseño 4: Evaluar Ensayo	149
6.3. <i>Implementación</i>	151
6.3.1. Lenguajes de programación empleados	151
6.4. <i>Prueba de la Aplicación</i>	152

Evaluación de la Entonación en la enseñanza del Español como Segunda Lengua

6.4.1. Plan de Pruebas	152
6.4.1.1. Requisitos para las Pruebas	152
6.4.1.2. Estrategia para las Pruebas	159
6.4.1.2.1. Tipos de Pruebas	159
Pruebas de Funcionalidad	159
Pruebas de Interfaz de Usuario	159
Pruebas de Integridad de la Bases de Datos.....	160
6.4.1.3. Herramientas para las pruebas	160
6.4.2. Casos de Prueba	160
6.4.2.1. Pruebas de Funcionalidad	161
6.4.2.2. Pruebas de Interfaz de Usuario	168
6.4.2.3. Pruebas de Integridad de la Bases de Datos	173
6.4.3. Resultado de las Pruebas.....	175
6.4.3.1. Variaciones.....	175
6.4.3.2. Resumen de Resultados	175
6.4.3.3. Cobertura de las Pruebas basada en Requisitos.....	175
6.4.3.4. Tipos de Errores e Impacto en el Sistema	176
6.4.3.5. Acciones Sugeridas	176
CAPÍTULO 7.....	177
MANUAL DE USUARIO	177
7.1. Servidor	177
7.1.1. Requisitos para la Instalación	177
7.1.2. Instalación	178
7.1.2.1. Instalación de la base de datos	179
7.1.2.2. Instalación de ffmpeg.....	179
7.1.2.3. Instalación de Snack	179
7.1.3. Puesta en Marcha.....	179
7.1.4. Desconexión	180
7.2. Cliente Móvil Android	180
7.2.1. Requisitos para la Instalación	180
7.2.2. Instalación	180
7.2.3. Puesta en Marcha.....	182
7.2.4. Instrucciones	183
7.2.4.1. Configuración del Host	183
7.2.4.2. Registro	184
7.2.4.3. Identificación / Inicio de sesión en el sistema (Login)	186
7.2.4.4. Realizar un ejercicio	186
7.2.4.4.1. Escuchar la Grabación de Referencia.....	188
7.2.4.4.2. Grabación y Envío del Ensayo	189
7.2.4.4.3. Recepción de los Resultados.....	190
7.2.4.5. Cerrar Sesión (Logout).....	191
CAPÍTULO 8.....	193
CONCLUSIONES Y TRABAJO FUTURO	193
8.1. Resultados Finales	193
8.2. Conclusiones.....	195
8.3. Trabajo Futuro.....	196
CAPÍTULO 9.....	199
ANEXOS	199
A - Competencias Requeridas y Ejercicios Tipo de Exámenes para la obtención del DELE ..	199
A.1. Contenidos para la obtención del DELE por niveles en relación con la entonación	199

A.1.1. A1-A2.....	199
A.1.1.1. Segmentación del discurso en unidades melódicas.....	199
A.1.1.2. Identificación y producción de los patrones melódicos correspondientes a la entonación enunciativa	200
A.1.1.3. Identificación y producción de los patrones melódicos correspondientes a la entonación interrogativa	201
A.1.1.4. Identificación y producción de los patrones melódicos correspondientes a los distintos actos de habla	201
A.1.2. B1-B2.....	201
A.1.2.1. Segmentación del discurso en unidades melódicas.....	201
A.1.2.2. Identificación y producción de los patrones melódicos correspondientes a la entonación enunciativa	202
A.1.2.3. Identificación y producción de los patrones melódicos correspondientes a la entonación interrogativa	203
A.1.2.4. Identificación y producción de los patrones melódicos correspondientes a los distintos actos de habla	203
A.1.3. C1-C2.....	203
A.1.3.1. Segmentación del discurso en unidades melódicas.....	203
A.1.3.2. Identificación y producción de los patrones melódicos correspondientes a la entonación enunciativa	204
A.1.3.3. Identificación y producción de los patrones melódicos correspondientes a la entonación interrogativa	204
A.1.3.4. Identificación y producción de los patrones melódicos correspondientes a los distintos actos de habla	205
A.2. Ejercicios tipo de los exámenes para la obtención del DELE	206
A.2.1. A1	206
A.2.2. A2	207
A.2.3. B1	208
A.2.4. B2	209
A.2.5. C1	210
A.2.6. C2	211
<i>B - Reconocimiento de Voz con Google Voice</i>	<i>213</i>
<i>C - Soporte y Desarrollo.....</i>	<i>215</i>
C.1. Hardware.....	215
C.2. Software	216
<i>D - Glosario.....</i>	<i>217</i>
<i>E - Índice de Tablas</i>	<i>223</i>
<i>F - Índice de Ilustraciones.....</i>	<i>227</i>
<i>G - Contenido del CD.....</i>	<i>231</i>
<i>H - Bibliografía.....</i>	<i>233</i>

Introducción

En esta memoria se describe el trabajo realizado durante el Trabajo de Fin de Grado del alumno Diego Vallejo Alonso, tutelado por D. Valentín Cardeñoso Payo para la finalización de los estudios de la titulación de Grado en Ingeniería Informática; y que pretende dar continuidad al trabajo desempeñado durante el Proyecto de Fin de Carrera del mismo autor, "Aplicación Móvil de Ayuda a la Mejora de la Pronunciación del Español", correspondiente a la titulación de Ingeniería Técnica en Informática de Sistemas.

En aquel proyecto, se realizaba un estudio de las oportunidades que la plataforma Android nos proporcionaba en el campo de las Tecnologías del Habla, analizando sus puntos fuertes y debilidades a través de la programación de pequeños prototipos funcionales que ilustraran las diferentes funcionalidades (Análisis, Síntesis y Grabación de Voz) y finalmente desarrollando una aplicación que integrara dos de ellas (Análisis y Síntesis de Voz) aplicable en el campo de la enseñanza del español como segunda lengua como un sistema de autoayuda complementario a la formación en las aulas para mejorar la pronunciación.

Todo ello estaba fundamentado en una base teórica en la que se explicaron qué son las tecnologías del habla, los tipos de tecnologías que se pueden dar (Reconocimiento Automático del Habla, Conversión Texto a Voz, Sistemas de Gestión de Diálogo), en qué consistía cada tecnología y cómo se utilizaban estas herramientas como apoyo para el aprendizaje de una segunda lengua, en lo que se denomina CALL (Computer-Assisted Language Learning).

Los resultados de aquel trabajo fueron positivos en cuanto a que pudimos conocer cuál es el alcance de Android en relación con las Tecnologías del Habla y a que pudimos explotar sus posibilidades creando una aplicación que sirviera para el propósito deseado. Sin embargo, dadas las limitaciones que la plataforma nos imponía y a que el alcance del proyecto no lo contemplaba, no se realizaba una evaluación completa de una locución vocal, sino que nos centrábamos exclusivamente en la pronunciación y se omitía todo lo referente a la entonación.

Partiendo de esta base, se ha considerado la oportunidad de continuar el anterior proyecto añadiendo al estudio realizado la evaluación de la entonación, lo que supondría abarcar de una forma más amplia el análisis de la voz. Se pretende, además, elaborar un módulo de evaluación de la entonación que podamos integrar en la aplicación prototipo que se desarrolló como producto final del proyecto anterior, de modo que su servicio sea más completo.

1.1. Motivación

Como se apuntaba en el Proyecto de Fin de Carrera "Aplicación Móvil de Ayuda a la Mejora de la Pronunciación del Español", los datos sobre el número de usuarios de smartphones o tablets indicaban una tendencia ascendente que nos mostraban que estos dispositivos iban a estar cada vez más extendidos, con todas las oportunidades a nivel de conectividad y servicios que ello ofrecía. Parece lógico que durante el último año, la cifra de terminales de este tipo sigue en auge y que por lo tanto, la decisión de explorar este campo como objetivo de nuestro estudio ha sido un acierto.

Por otra parte, también se arrojaron algunos datos referentes al crecimiento de Android en los últimos años, su liderazgo en el mercado y las nuevas versiones para remarcar que esta plataforma está en constante evolución y que no se trata de una moda pasajera. Si antes de presentar aquel trabajo (Julio de 2012) Google acaba de anunciar el lanzamiento de la versión 4.1 Jelly Bean, desde entonces ya han sido anunciadas dos nuevas versiones, la 4.2 en Octubre de 2012 y más recientemente la 4.3 en Julio de 2013, ambas también Jelly Bean. Además, durante este periodo, Android está haciendo un esfuerzo importante en adaptar las nuevas funciones que ofrece su API a las versiones antiguas de la plataforma, proporcionando librerías de soporte para solucionar su famoso problema de fragmentación (demasiadas versiones y gran cantidad de modelos de dispositivos, muchos soportan versiones diferentes). Por consiguiente, parece que se confirma nuestra teoría esta plataforma y que podemos seguir apostando fuerte por ella para continuar nuestro estudio.

Reforzando nuestra confianza por el sector y la tecnología que hemos escogido como base de nuestro estudio, nos surgen otras motivaciones para continuar con el trabajo iniciado y seguir progresando en el campo de la aplicación de las tecnologías vocales en la enseñanza de una segunda lengua, desarrollando un producto para smartphones que soporten la plataforma Android.

1.1.1. Importancia del Español

Según los datos del Anuario de 2012 del Instituto Cervantes, más de 495 millones de personas en todo el mundo hablan español, siendo la segunda lengua del mundo en número de hablantes (superado solo por el chino mandarín, con más de 1000 millones) y el segundo idioma de comunicación internacional. Según las previsiones, esta cifra incrementará en los próximos años. Por otra parte, se afirma que el porcentaje de población mundial que habla el español de forma nativa está aumentando frente a otros idiomas de referencia como el chino o el inglés que están descendiendo. Además, se estima que el aproximadamente 18 millones de alumnos estudian el español como segunda lengua, situándose el crecimiento anual del número de matrículas de los estudiantes del idioma en un 8%. Se trata del tercer idioma más estudiado como lengua extranjera por detrás del francés y el inglés, y seguido por el alemán. En cuanto al nivel de penetración en Internet, el español es la tercera lengua más utilizada en la Red (por delante el inglés y el chino), con un crecimiento en su uso de más del 800% en la última década y constituyendo el cuarto idioma con mayor demanda de documentos. Si tenemos en cuenta que el chino es utilizado casi exclusivamente por hablantes nativos, su importancia aumenta más aún si cabe. Por último, podemos destacar también al español dentro de otros campos, como la economía o la ciencia, en los que forma un activo que incrementa las posibilidades de impulso económico y un instrumento significativo de difusión de los resultados de los estudios científicos, respectivamente. Todos estos datos nos hacen ver la importancia que el español, como idioma, tiene en el mundo y los beneficios (no sólo a nivel cultural, sino también en otros sectores) de su difusión; de modo que resulta bastante obvio que un producto que ayude a mejorar la destreza en la producción del idioma podría ser de especial interés para un gran sector de la población mundial.

1.1.2. La Entonación en la Enseñanza del Español

La evaluación de la entonación en la enseñanza de la segunda lengua supone un desafío por la dificultad de encontrar unos criterios que se puedan fijar para cuantificar la corrección de una locución. Existen autores que, además, señalan que los esfuerzos en este sentido han sido insuficientes o inapropiados, por lo que se hace necesario intensificar su estudio para obtener avances que permitan mejorar los criterios y sistemas que empleamos actualmente.

El estudio de esta materia se ha visto ampliado en los últimos años por compañías y grupos de investigación, y ya podemos ver en el mercado algunos productos comerciales que proporcionan sistemas de apoyo para el aprendizaje del inglés. Por tanto, sería importante desarrollar similar para nuestro idioma, el español, que permitiera a los estudiantes del idioma (e incluso a persona nativas con dificultades), mejorar sus competencias en el dominio del idioma.

1.1.3. Ampliación del Prototipo Inicial

Como ya se ha adelantado, el producto final de este trabajo toma como punto de partida la aplicación desarrollada como producto final del proyecto anterior. Esta aplicación estaba planteada como un “juego serio”, en el que se proponían una serie de frases al usuario, de modo que podía escucharlas (eran reproducidas por el sintetizador que provee el API Android), y a continuación podía realizar un ensayo (o intento, como se prefiera) que se calificaría según el texto reconocido por Google Voice (por defecto, el reconocedor proporcionado por el API de Android), calculando con el algoritmo de Levenshtein la distancia entre lo pronunciado por el usuario y lo que debería haber dicho. Además, el usuario podía ver su puntuación media y máxima de los intentos de cada frase. Se le ofrecía también la posibilidad de introducir sus propias frases en el ejercicio, de modo que el conjunto de frases no fuera cerrado y que pudiera ensayar sus propios desafíos.

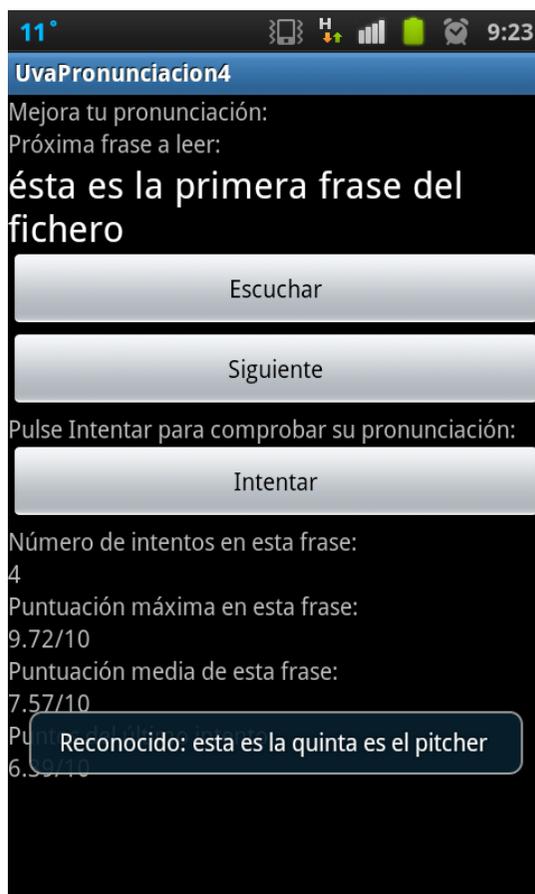


Ilustración 1: Prototipo de partida

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

En cuanto al propósito de la aplicación, cumplía con las expectativas, pero se tienen una serie de necesidades que no han sido completamente cubiertas. En primer lugar, parece que la voz sintética reproducida no resulta todo lo natural que debería. Este problema se ha visto agravado con la nueva voz que Google ha propuesto, pues da la impresión de ser “temblosa”, pudiendo resultar desagradable. Por tanto, quizás se debería dotar a la aplicación de algún mecanismo (otro sintetizador, descarga de grabaciones...) que minimice esta sensación en el usuario y que cumpla el mismo propósito.

Además, al utilizar el reconocedor de voz de Android (Google Voice), no se nos permite mantener la grabación de la locución del usuario, lo cual no nos beneficia de cara al objetivo de mejorar nuestro servicio, pues nos interesa acumular el mayor número de grabaciones de los intentos de los usuarios para poder entrenar repetidas veces el sistema e ir afinando su precisión en la clasificación como nativo o no nativo y dar una medida de la distancia lo más fiel posible para poder calificar de con mayor corrección el intento. Por otra parte, Google posee la mayor base de datos para dar soporte al reconocimiento de voz, con lo cual tenemos especial interés en, además de mantener la grabación, recuperar lo que ha reconocido Google Voice y sus parámetros estadísticos de estimación de confianza para poder tener una referencia con la que comparar de cara a futuras evoluciones del sistema (podríamos desear utilizar nuestro propio reconocedor de voz en futuras versiones del producto), por lo que se desea investigar si existe alguna posibilidad, un “work around” que nos ayude en este sentido.

Por último, nuestra aplicación se centra solamente en la pronunciación (plano segmental). Quizás para un usuario que desea mejorar concretamente ese aspecto podría resultar provechoso el uso de nuestro producto, pero se desea aportar más valor dotándole de la capacidad de evaluar también la entonación. De este modo, el servicio daría soporte para la mejora de otro de los tres grandes bloques que podemos distinguir en la producción del habla (base articulatoria, plano suprasegmental y plano segmental, la entonación pertenecería a este último).

1.2. Esquema de la memoria

Podemos organizar esta memoria en cuatro bloques bien diferenciados. El primero de ellos versaría sobre la presentación del trabajo que realizaremos, y estaría compuesto por los Capítulos 1 y 2. En ellos trataremos asuntos como qué nos ha llevado a la elección del tema de la evaluación de la entonación en la enseñanza del español como segunda lengua como materia principal del trabajo, qué esperamos conseguir en el transcurso de este proyecto y cuál será la forma de proceder durante su realización.

El siguiente bloque incluiría los conceptos teóricos que debemos tener en cuenta durante el desarrollo del proyecto. A su vez, podríamos subdividir esta sección en dos bloques: uno que podríamos denominar "de alto nivel" y otro más específico. En el primero se exponen aspectos como la situación actual de la enseñanza de la entonación en nuestro idioma, los criterios considerados para la evaluación de la entonación y los niveles establecidos por una institución de referencia como el Instituto Cervantes para la obtención de los Diplomas de Español como Lengua Extranjera (DELE), algunos ejercicios típicos de los exámenes que dan acceso a estos certificados y un breve apunte sobre los patrones entonativos del español. Por otra parte, en el segundo bloque nos centramos más en los aspectos concretos referentes a la voz humana que vamos a manejar para tratar de abordar el problema. Este apartado se correspondería con los Capítulos 3 y 4.

Un tercer grupo estaría constituido por los Capítulos 5, 6 y 7, referentes a la parte más técnica de este documento. En ellos veremos con detalle los pasos seguidos para el diseño y desarrollo del módulo de evaluación de la entonación, desde la recopilación del corpus de entrenamiento, pasando por su procesamiento, hasta su inclusión en una aplicación prototipo que ilustre cómo se podría emplear un sistema de este tipo. En el Capítulo 7, además, se añade un manual de usuario de ésta.

Capítulo 1: Introducción

El último bloque del trabajo lo podríamos catalogar como un bloque de evaluación, en él expondremos los resultados obtenidos y los enfrentaremos con los objetivos que nos marcamos inicialmente para comprobar que el trabajo cumplió las expectativas y poder elaborar una conclusión en un sentido u otro en función de ello. Además, basándonos en la experiencia de haber podido manejar ya un prototipo que trate de dar solución al problema propuesto, seguramente tengamos varias oportunidades para ampliar las fronteras de este estudio, las cuales comentaremos. Este apartado lo constituye el Capítulo 8.

Como documentación extra, y con el fin de complementar el material expuesto en el resto de la memoria, se propone un conjunto de Apéndices en el que se amplían otros datos que podrían resultar de interés y que en algún momento han sido considerados para este trabajo. Podemos encontrarlos en el capítulo 9.

Plan de Trabajo

En este capítulo se presenta el plan que se utilizará para la realización del Proyecto. Se definen de forma global las fases que constituirán el desarrollo del trabajo, añadiendo una estimación aproximada de lo que durará cada una, con el fin de realizar el seguimiento y control del Proyecto. Se determinarán los objetivos para las fases, de modo que en cada hito de final de fase se pueda comprobar si se han logrado las metas predeterminadas.

También se definen los objetivos del Proyecto, y se comentan las suposiciones y restricciones iniciales que podrían motivar ciertas decisiones en la planificación. A continuación, se expone el plan de trabajo ideado para abordar el desarrollo del Proyecto, indicando las duraciones esperadas para cada fase. Aunque podríamos hacer una reseña sobre los recursos software y hardware utilizados durante la consecución del trabajo aquí, se ha optado por incluir un apéndice "Soporte y Desarrollo" al final de la memoria con tal contenido.

2.1. Visión General del Proyecto

En este apartado se describe el problema y se definen los objetivos del trabajo, así como la descripción de aquellas circunstancias que pudieran influir directamente en la planificación del mismo.

2.1.1 Descripción del Problema

Como punto de partida, se dispone de un prototipo de aplicación de ayuda para el aprendizaje del español como segunda lengua para plataformas Android, el cual se desea ampliar con mecanismos de evaluación de la calidad de la entonación de las locuciones generadas por los usuarios.

La aplicación existente plantea un escenario en el cual el usuario debe repetir las frases en español que le muestra el sistema. Estas frases están elegidas por un educador y pueden ir acompañadas de material audiovisual. La medida de la calidad de la pronunciación se basa en el uso de la tecnología de vocal de reconocimiento de voz de Google, Google Voice, por medio de la cual es posible obtener un conjunto ponderado de transcripciones posibles que se hacen corresponder con la locución del usuario. Cada una de ellas está acompañada de una probabilidad (likelihood) que indica la posibilidad de que lo que ha transcrito el reconocedor sea correcto. De todas las transcripciones, se selecciona la primera, de modo que la medida de distancia de dicha transcripción con respecto a la frase problema (actualmente calculada con el algoritmo de Levenshtein), permite asociar a cada pronunciación una puntuación.

2.1.2. Objetivos del Proyecto

La comparación entre la frase propuesta y la locución de la misma en el prototipo existente se realiza exclusivamente en términos fonéticos puesto que el sistema de reconocimiento (Google Voice) no tiene en cuenta los aspectos prosódicos de dicha locución. Por ello, en este trabajo se plantea como objetivo principal el **desarrollo de un módulo de comparación de entonación entre locuciones de**

una misma frase que permita obtener una medida cuantitativa de similitud entre la entonación esperada de esa frase y la producida por el locutor.

Para lograr este objetivo central, se plantea el siguiente conjunto de objetivos y tareas elementales:

1. Recopilación de un corpus vocal en el que se disponga de varias pronunciaciones de un conjunto de frases controlado (siguiendo las directrices marcadas para la obtención de los DELE) por varios locutores y que pueda emplearse para elaborar el modelo de comparación. El corpus debe incluir grabaciones de los locutores repitiendo los ejercicios en diferentes situaciones (repetición del mismo ejercicio 5 veces seguidas, repetición del ejercicio a las dos horas, a la semana...), para poder comprobar las variaciones intra-locutor .
2. Preparación del corpus anterior para extraer de cada locución el conjunto de parámetros prosódicos que se vayan a emplear para caracterizar dicha locución. Para ello se analizarán diversas herramientas (SPTK, Praat, Snack...) existentes que facilitan la extracción de la secuencia de valores de la frecuencia fundamental (F0) de una señal hablada y a partir de dichas secuencias se elaborarán los parámetros adecuados según el tipo de unidad de habla elegida (sílabas, palabras, grupo acentual, grupo de entonación). Básicamente, se usarán en principio valores máximos, mínimos, medios y evolución temporal de las duraciones, pausas y niveles de F0. Se desea que las herramientas que extraigan las secuencias de valores de la frecuencia fundamental puedan ser ejecutadas en un servidor, y que el tiempo que tardan en procesar una locución sea razonable.
3. Evaluación experimental del método de comparación de entonación. Para ello se analizarán diversos métodos de clasificación que permitan determinar cuándo dos locuciones de una misma frase son iguales o distintas, dentro de unos márgenes de error suficientemente pequeños. En este sentido, las pronunciaciones perceptualmente equivalentes de una misma frase, realizadas por el mismo locutor o por locutores diferentes, deberán ser clasificadas como iguales de forma que las que no resulten equivalentes a éstas estén lo más separadas que sea posible de ellas.

Un aspecto fundamental de este estudio es que la calidad del módulo de evaluación de la distancia entre dos locuciones cualesquiera ha de ser comparable a la que proporcionaría un hablante nativo de español. Para ello, será indispensable disponer de un corpus de locuciones que sea sometido a evaluación previa por escuchantes humanos, quienes decidirían si dos locuciones son o no equivalentes perceptualmente en términos de entonación. Metodológicamente, se trataría de presentar al experto una locución de referencia y un conjunto pequeño de alternativas que debe valorar si son o no son similares a la primera. Otro enfoque que debería considerarse sería aquél en el que se le muestren varias alternativas de locución y deba marcar cuáles percibe como equivalentes en términos de entonación (entendiendo por esto que transmiten la misma información pragmática sobre la intención informativa, el estado del hablante o el contexto de discurso). A partir de estas etiquetas proporcionadas por nativos expertos, el sistema que se quiere desarrollar deberá intentar emular lo más fielmente que sea posible esas capacidades, sin merma de la capacidad de la generalización que caracteriza a un ser humano.

2.1.3. Suposiciones y Restricciones

El trabajo se realizará durante el segundo cuatrimestre del curso 2012/2013, debiéndose compaginar su elaboración con la asignatura de Profesión y Sociedad y con las Prácticas en empresa programadas en el Plan de Estudios para el Curso de Adaptación al Grado en Ingeniería Informática.

Los plazos estipulados para entregar la solicitud de defensa y presentación de proyectos son hasta el 5 de julio de 2013 para la convocatoria ordinaria, y hasta el 31 de julio de 2013 para la convocatoria extraordinaria. En este segundo caso el plazo de presentación de trabajos expira el 6 de septiembre de 2013.

Teniendo en cuenta estas circunstancias, con las tareas paralelas y el examen programado para el 17 de junio de 2013, se pretende finalizar el trabajo como muy tarde el día 2 de septiembre de 2013, con el fin de poder encargar la impresión de esta memoria y poder entregarla dentro de los plazos establecidos. Por otra parte, el Trabajo de fin de Grado se corresponde con 12 ECTS equivalentes a 300 horas, por lo que se espera que la actividad se prolongue durante aproximadamente ese tiempo.

2.2. Gestión del Proceso

A continuación se presenta toda la información referente a las fases en que se divide el proyecto, incluyendo los objetivos a cumplir y la duración estimada de cada una de ellas.

Las estimaciones de duración que se muestran están basadas en la experiencia de anteriores proyectos y en las suposiciones y restricciones que se citaron en el apartado anterior. El esfuerzo que se empleará para llevar a cabo el trabajo será realizado por un único recurso humano.

2.2.1. Planificación del Proyecto en Fases

De acuerdo con los objetivos y suposiciones especificados en la sección anterior, el plan de trabajo propuesto se compone de las siguientes fases:

Fase	Tarea	Estimación de duración
Fase 1	Fase de estudio inicial: caracterización del habla, extracción de F0, herramientas, métodos de evaluación. Recopilación de corpus.	4-6 semanas
Fase 2	Preparación del Corpus de Entrenamiento	6-8 semanas
Fase 3	Diseño y Desarrollo de un Módulo de Evaluación de la Entonación e Integración en una Aplicación Móvil Prototipo	1-2 semanas
Fase 4	Edición de la Documentación Elaborada y Preparación de la Memoria Final.	2-3 semanas

Tabla 1: Fases del Proyecto

2.2.2. Objetivos de las Fases

Después de haber dividido el plan de trabajo en estas cuatro etapas, podemos, a su vez, desglosar de forma natural las fases definidas en los objetivos y tareas que se presentan en este apartado.

2.2.2.1. Fase de Estudio Inicial

- Estudio de algoritmos de extracción de la frecuencia fundamental (F0) del habla y métodos de evaluación de la entonación.
- Estudio de herramientas que faciliten la secuencia de valores de la frecuencia fundamental (F0) del habla (cómo funcionan, que algoritmo de extracción utilizan, posibilidad de ejecutarlas en un servidor).
- Selección de una herramienta que se adapte a las necesidades descritas (ejecución en servidor, respuesta en tiempo razonable).
- Selección de un conjunto de frases que incluyan diferentes modalidades de entonación (enunciativa, interrogativa, exclamativa...), siguiendo las pautas que se indican para la obtención del DELE.
- Recopilación de un corpus vocal compuesto por grabaciones de distintos locutores pronunciando las frases de seleccionadas en el punto anterior. Se deberán realizar varias repeticiones por cada locutor de cada ejercicio.
- Selección de un vector de parámetros que se puedan extraer de la secuencia de valores de la frecuencia fundamental (F0) del habla extraída de una locución y que permitan evaluarla.
- Estudio de la tecnología que se utilizará para el desarrollo del módulo.

2.2.2.2. Preparación del Corpus de Entrenamiento

- Etiquetado de las grabaciones que componen el corpus vocal.
- Extracción de los valores de los parámetros prosódicos seleccionados de las grabaciones que componen el corpus vocal.
- Generación de ficheros de datos a partir de la información extraída del corpus vocal para su posterior análisis.
- Análisis de los datos recopilados.
- Elección de la herramienta más adecuada y que asegure un mayor índice de acierto en la evaluación.

2.2.2.3. Diseño y Desarrollo de un Módulo de Evaluación de la Entonación e Integración en una Aplicación Móvil Prototipo

- Diseño del módulo de comparación de la entonación teniendo en cuenta la herramienta seleccionada.
- Desarrollo del módulo de comparación de la entonación que integre la herramienta seleccionada.
- Prueba y evaluación del prototipo incluyendo el módulo desarrollado.

2.2.2.4. Edición de la documentación elaborada y preparación de la memoria final

- Recopilación de los documentos elaborados durante el transcurso del proyecto e integración de su contenido en la memoria final.
- Añadido de la información necesaria para completar la memoria final.
- Formateado final de la memoria. Revisión y corrección de errores.

La Evaluación de la Entonación en la Enseñanza del Español

En este capítulo nos centraremos en presentar los aspectos de más alto nivel del trabajo. En primer lugar, presentaremos el concepto central que nos ocupa en este proyecto, la entonación, y la situaremos dentro del contexto lingüístico conociendo sus orígenes y rasgos más generales. A continuación, analizaremos la situación actual de la evaluación de la entonación en la enseñanza del español como segundo idioma (L2), describiendo algunas opiniones sobre posibles criterios de evaluación y exponiendo cuáles son los Diplomas de Español como Lengua Extranjera (DELE) y sus niveles y qué requisitos debe cumplir un candidato para acceder al certificado de un nivel concreto. Finalmente, se realiza un breve apunte sobre las propiedades de la entonación en el español y los patrones que presenta.

3.1. Prosodia y Entonación

Como el trabajo que nos importa está directamente relacionado con la entonación, en primer lugar debemos comprender a qué nos referimos cuando hablamos de este concepto y en qué contexto lo situamos dentro de las áreas que engloba la lengua española. Tal como indica el Diccionario de la Real Academia Española de la Lengua, la entonación es la "Modulación de la voz en la secuencia de sonidos del habla que puede reflejar diferencias de sentido, de intención, de emoción y de origen del hablante, y que, en algunas lenguas, puede ser significativa". Viendo esta definición, podemos ubicar a la entonación como parte de un aspecto más general, la prosodia.

Etimológicamente, la palabra *prosodia* proviene del griego clásico προσῳδία: πρὸς ('cerca', 'al lado') ἢ ᾠδῆ ('canto'). El término latino *accentus* (*ad - cantus*), 'acento', es muy parecido al griego.

En un principio se refería a una canción acompañada de música instrumental. Posteriormente se empleó en literatura griega y latina para referirse al conjunto de reglas relativas a la acentuación y a la cantidad de las vocales y de las sílabas, imponiéndose a otros términos como *tenor*, *nota uocis*, *moderamentum*, *accentiunculum* o *ouoculatio*. Finalmente, el término ha sido adoptado por disciplinas modernas como la fonética y la fonología, con el fin de designar una rama de estudio dedicada a los fenómenos fonéticos que implican más de un fonema o segmento, los fenómenos suprasegmentales. De este modo, la prosodia estudia la entonación, el acento y el ritmo, que incluye a su vez timbre, pausas y velocidad de elocución.

Según el Diccionario de la Real Academia Española de la Lengua, la prosodia es la "parte de la gramática que enseña la recta pronunciación y acentuación". En otra acepción, se califica como el "estudio de los rasgos fónicos (perteneciente o relativo a la voz) que afectan a la métrica, especialmente

de los acentos y de la cantidad". Además, se añade una tercera que la define como la "parte de la fonología (rama lingüística que estudia los elementos fónicos, atendiendo a su valor distintivo y funcional) dedicada al estudio de los rasgos fónicos que afectan a unidades inferiores al fonema (cada una de las unidades fonológicas mínimas que en el sistema de una lengua pueden oponerse a otras en contraste significativo; dentro de cada fonema caben distintos alófonos), como las moras (unidad de medida de la cantidad silábica, equivalente a una sílaba breve), o superiores a él, como las sílabas u otras secuencias de la palabra u oración".

Otra institución de referencia es el Centro Virtual Cervantes, en donde podemos buscar qué se expone de la prosodia en relación con el objetivo final Clave de ELE, encontramos que se refiere a la prosodia como "el conjunto de fenómenos fónicos que abarcan más de un fonema o segmento -entonación, acentuación, ritmo, velocidad de habla, etc. (denominados fenómenos suprasegmentales)". Por otra parte, se alude a otras funciones de la prosodia en la organización e interpretación de un discurso y a la información (emotiva, sociolingüística y dialectal) que se puede transmitir en éste al mismo tiempo gracias a su influencia.

En el español, los fenómenos prosódicos más relevantes son la acentuación y la entonación. Otro fenómeno es el ritmo, cuya función es agrupar los sonidos del discurso en bloques (*palabras fónicas o grupos rítmicos*), con el fin de facilitar la descodificación y comprensión del mensaje. Las pausas contribuyen de forma decisiva a caracterizar los patrones rítmicos y entonativos de la lengua.

Podemos tratar de comprender el sentido del término "suprasegmental" a través de un ejemplo. Al oír una sílaba aislada [/sa/], no es posible saber si la vocal es átona o tónica. Por el contrario, al oír [/lamésa/] o [/losábe/], sí podemos establecer una distinción: /sa/ es átona en [/lamésa/] y tónica en [/losábe/]. Se necesita un punto de apoyo átono para poder realzar la vocal tónica, por lo que el fenómeno de la acentuación abarca varios fonemas o segmentos.

Para describir los fenómenos suprasegmentales de una lengua, se estudian las características o *rasgos prosódicos* de sus sonidos: el tono, el timbre, la intensidad y la duración. Asimismo, los análisis prosódicos también se ocupan de las características relativas al tipo o cualidad de la voz -susurrada, ronca, inestable, nasalizada...- o a la articulación -precisa o imprecisa-.

Normalmente, los estudiantes de L2 conservan rasgos fónicos (segmentales y suprasegmentales) característicos de su lengua materna, aunque hayan alcanzado un dominio excelente de la L2 en otros niveles (léxico, gramatical, estilístico, pragmático...). Se entiende que la pronunciación y la entonación son los componentes más personales de la identidad lingüística del alumno por la estrecha relación entre el componente afectivo y el entonativo. De hecho, la entonación informa sobre otros aspectos del hablante más allá del contenido de su discurso: actitud, estado de ánimo, sentimientos, emociones..., lo que puede causar malentendidos en caso de que un extranjero emplee una entonación inapropiada al dirigirse a otro interlocutor, por ejemplo cuando se desea pedir una ayuda (el nativo puede percibir una orden y no una solicitud).

3.2. La Enseñanza de la Prosodia del Español

Profesores y lingüistas reconocen que la prosodia española es compleja; por ello, es conveniente prestarle atención. El peso de la prosodia en el aprendizaje de la lengua ha variado a lo largo de la historia, en función de las corrientes metodológicas. En el Método gramática-traducción o en el Enfoque del código cognitivo la prosodia no se tiene en cuenta. Por el contrario, en el Método directo y en los modelos didácticos basados en el conductismo (Método audiolingüe, Enfoque oral, Enfoque

situacional), el alumno escucha e imita la pronunciación del profesor (o de grabaciones). Actualmente, se utiliza un Enfoque por tareas, en el cual se incluye la prosodia. En cualquier caso, es esencial que los alumnos practiquen la prosodia a la par que otros componentes del lenguaje y en discursos de habla espontánea.

Según LListerri en su artículo "La evaluación de la pronunciación en la enseñanza del español como segunda lengua" (2003), la enseñanza de la pronunciación es uno de los temas a los que menor atención se ha dedicado en la enseñanza del español como segunda lengua (E/L2), de tal modo que la evaluación de la pronunciación no ha sido lo suficientemente abordada como para establecer unos criterios concretos. Los aspectos fonéticos de la lengua poseen determinadas características especiales que hacen necesaria una reflexión sobre la forma en que se evalúa (y por lo tanto, cómo se enseña) la pronunciación. Es por esto que la enseñanza de la pronunciación resulta especialmente compleja, quizá más difícil de abordar que cualquier otra de las facetas que integran la labor de un profesor de ELE.

A su vez LListerri postula que se puede evaluar la pronunciación con al menos cuatro objetivos diferenciados: determinar el nivel de la enseñanza en el que el alumno debe integrarse (pruebas de nivel que se realizan en los centros cuando se incorporan los nuevos estudiantes); valorar la competencia fonética adquirida por el estudiante en un tiempo conforme a determinados criterios (cuando se requiere certificar el conocimiento de la lengua); establecer los objetivos para la enseñanza de la pronunciación (ligado al análisis de las necesidades cuando se inicia el trabajo con un grupo); y conocer las dificultades del alumno con el fin de aplicar las estrategias de corrección más adecuadas (vinculado con la labor diaria en el aula y el análisis preciso del error que conlleva toda corrección). Se apunta que los tres primeros objetivos se corresponden con lo que se llama "evaluación para el diagnóstico" – diagnostic evaluation (Celce-Murcia et al., 1996:341) –, mientras que la intención del cuarto es proporcionar al alumno información que le ayude a mejorar su producción oral.

La competencia fonética implica dos destrezas estrechamente relacionadas, producción y percepción, así que en la evaluación tendrán que considerarse conjuntamente también. La relación entre ambas es objeto de debate (Llisterri, 1995) y el peso que se dé a cada una en la enseñanza de una L2 puede llevar a adoptar estrategias que prioricen la articulación (colocar correctamente y poner en contacto los órganos activos (labios, lengua y velo del paladar) sobre un órgano pasivo (dientes, alvéolos y paladar duro)), o la discriminación (distinguir entre dos sonidos). Se suele asumir que la producción se evalúa en las pruebas de expresión oral y la percepción en las de comprensión oral. Aunque los resultados de estas pruebas proporcionan unos indicadores generales, no reflejan con detalle las capacidades fonéticas de los estudiantes precisamente por su genericidad.

3.3. Evaluación de la pronunciación, Plan Curricular y Diplomas de Español

Tal como sucede con otros ámbitos de la enseñanza de la lengua, la evaluación de la pronunciación debe seguir unos objetivos previamente establecidos en un diseño curricular. En el caso del español como segunda lengua, uno de los documentos de referencia es el Plan Curricular del Instituto Cervantes. El Plan Curricular del Instituto Cervantes desarrolla y fija los niveles de referencia para el español según las recomendaciones que, en 2001, publicó el Consejo de Europa en su Marco Común Europeo de Referencia para las lenguas: aprendizaje, enseñanza, evaluación; pasando de los cuatro niveles —inicial, intermedio, avanzado y superior— que habían venido constituyendo hasta la fecha el ciclo curricular completo de la oferta de cursos del Instituto, a seis niveles —A1-A2, B1-B2, C1-C2— relacionados con la propuesta de niveles comunes del MCER. Puesto que los criterios de evaluación se definen en el plan curricular en función de las cuatro destrezas clásicas (expresión oral, expresión

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

escrita, comprensión auditiva y comprensión lectora), es esperable que la evaluación de la pronunciación forme parte de la valoración de la expresión oral en los distintos niveles propuestos.

Dado que el tema en el que se centra este trabajo es la entonación, la destreza que nos interesa concretamente es la expresión oral. En el apartado de "Pronunciación" del Plan Curricular deberíamos encontrar las directrices que nos guíen en la enseñanza y evaluación de este aspecto de la lengua española, en el que se incluye también lo concerniente a la entonación. Si atendemos al texto, parece claro que el objetivo final es conseguir que el alumno alcance un buen nivel comunicativo por encima de la corrección de la pronunciación (sin renunciar a ésta), aunque también remarca la dificultad del mismo. Se afirma que *"El profesor tiene que transmitir al alumno la idea de que es factible adquirir una pronunciación clara, aceptable y comprensible para el hablante nativo, pero que no lo es tanto conseguir una pronunciación cuasi perfecta, equivalente a la de los hablantes nativos, precisamente porque en el proceso de adquisición entran en juego muchas variables, que, con frecuencia, escapan a su propio control"*, así como que *"En circunstancias poco favorables, lograr una expresión oral aceptable, fluida y apropiada puede quizá resultar un propósito inalcanzable, mientras que conseguir una buena pronunciación, en cambio, puede ser una tarea perfectamente realizable."*

Desde el punto de vista del modo en el que el maestro deberá abordar la tarea de enseñar cómo pronunciar, no se determina una metodología específica: *"Dada la complejidad que implican tanto la enseñanza como el aprendizaje de la pronunciación, parece lógico pensar que el profesor no tiene por qué ceñirse a un único modelo metodológico"*. Por otra parte, se descarta que los conceptos gramaticales ayuden de una forma clara para mejorar la pronunciación de los alumnos: *"Todo el debate existente en el campo de la enseñanza de lenguas acerca del grado mayor o menor de atención que se le debe conceder a la forma (frente al contenido o la función) tiene una cierta razón de ser si nos referimos al componente estrictamente gramatical de la lengua (morfología, sintaxis), pero deja de tenerla si estamos hablando de pronunciación, porque en este campo la atención a la forma es simplemente imprescindible"*.

Se razona esta teoría basándose en la posibilidad de que los estudiantes se centren demasiado en hablar formulando oraciones gramaticalmente correctas, más aún que en transmitir otros aspectos emocionales del mensaje: *"Existe el peligro de que este énfasis en el aspecto formal provoque en los aprendices cierta dificultad para emitir enunciados «reales» o responder a ellos de modo natural, lo que genera un tipo de habla mecánica caracterizada, por ejemplo, por la ausencia de modulaciones entonativas o incorrecciones acentuales. [...] El papel que desempeña la prosodia es, en este sentido, fundamental"*. Siguiendo con el razonamiento de la imposibilidad de señalar una metodología determinada (demostrando una vez más lo especial de la materia), se destaca la figura del profesor sobre otros instrumentos de docencia, aunque no se renuncia al empleo de los mismos como ayuda: *"El profesor que aspire a obtener un buen resultado en esta tarea no solo debe estar preparado académicamente para «adaptar sobre la marcha» el método seguido a las necesidades manifestadas por los alumnos, sino que ha de saber integrar las prácticas de fonética correctiva en el conjunto de actividades interactivas del modo más natural posible. Por consiguiente, su papel no lo puede desempeñar ni un programa informático ni una grabación de vídeo, ni ninguna otra herramienta de las que a menudo se ofertan como recursos sustitutivos del profesor, quien, por lo demás, puede aprovechar las ventajas que todas ellas le ofrecen como meros instrumentos complementarios"*.

Una vez se han introducido los propósitos y las estrategias para la enseñanza de la pronunciación, el Plan Curricular del Instituto Cervantes establece una *"Ordenación jerárquica de los grandes bloques en que pueden distribuirse los distintos aspectos fonéticos en función del grado de relevancia que tengan para la calidad global de la producción en la nueva lengua"*:

A. Disposición (base) articulatoria del español

B. Plano suprasegmental: sílaba, acento, ritmo y entonación.

1. La sílaba
2. El acento en español
 - i. El acento léxico
 - ii. El acento oracional
 - iii. El acento enfático
3. Distribución y tipos de pausas
 - i. El grupo fónico
 - ii. El tempo
4. El ritmo en español
5. La entonación en español
 - i. Las formas entonativas básicas
 - ii. Las modalidades expresivas y afectivas

C. Plano segmental - Fonemas vocálicos y consonánticos: alófonos en distribución complementaria y variantes libres de los distintos fonemas.

1. Las vocales del español
 - i. Modificaciones contextuales
 - ii. Secuencias vocálicas
 - iii. Fenómenos dialectales
2. Las consonantes del español
 - i. Variantes contextuales
 - ii. Consonantes agrupadas
 - iii. Variantes dialectales

Se trata de justificar la jerarquía anterior analizando en base al objetivo final, es decir, a una comunicación eficaz por parte del estudiante: *"En un intercambio comunicativo es más importante la correcta entonación de los enunciados que lo componen que la pronunciación más o menos correcta de un sonido dado, y para conseguir una buena calidad general de las emisiones son más decisivos los ajustes articulatorios a medio y largo plazo que los limitados a la producción de un único elemento"*.

A partir de aquí se razona el porqué de cada uno de los tres grandes bloques, comenzando por el plano segmental y el suprasegmental, remarcando este último, el cual es de gran importancia desde el punto de vista del objetivo principal (que se reafirma), sobre todo por los aspectos que se pueden transmitir mediante su uso: *"De acuerdo con las últimas tendencias didácticas y con la orientación que preside, en general, el Plan curricular del Instituto y los objetivos de los Niveles de referencia para el español, la enseñanza de la pronunciación debe, por principio, basarse en la interacción comunicativa más que en la práctica repetitiva de las formas aisladas, y, consecuentemente, el plano suprasegmental, que tanta importancia reviste para la comunicación en cuanto que es el que más directamente refleja los factores situacionales y contextuales de los que depende la cadena hablada, se ordena en primer lugar"*. Si nos fijamos, además, insta a mejorar las aptitudes comunicativas de los alumnos a través de la interacción sobre ejercicios de repetición.

Por otra parte, se apunta que la adopción de la articulación del español por parte del estudiante es un proceso que se prolonga durante toda su enseñanza: *"La asimilación de la base de articulación no admite niveles, el alumno debe captarla y «reprogramarse» desde el comienzo hasta el final de su proceso de aprendizaje"*; quizás restándole responsabilidad o influencia al profesor en este bloque (que domina a los dos anteriores) y dejando entrever una cierta dependencia de que en las fases iniciales del aprendizaje el alumno se familiarice con la configuración articulatoria de nuestra lengua.

Por último, se establecen unos niveles que establezcan unos límites para la evaluación y certificación, basándose en la calidad de las producciones de los alumnos aplicadas a diferentes situaciones tipo que se le propongan: *"La división en niveles de los contenidos de carácter fonético puede mantenerse única y exclusivamente como un mero recurso expositivo y organizativo. La progresión en este campo de la fonética no se mide tanto por la cantidad y la cualidad de la información que se proporciona al estudiante en cada etapa, sino por la calidad que sus emisiones van alcanzando conforme se suceden los cursos. Y, desde una perspectiva comunicativa, la calidad de la pronunciación viene determinada no solo por el dominio global que el alumno demuestra de los rasgos suprasegmentales y segmentales, sino también por la medida en que es capaz de servirse con éxito de ellos adaptándolos a las diversas situaciones comunicativas en que se halle. No se trataría de evitar la exposición del estudiante a unos determinados contenidos fonéticos ni de dosificar la corrección por parte del profesor en función de los niveles de suficiencia establecidos, sino de marcar unos puntos de referencia generales para la evaluación final"*.

Estos niveles, también se pueden agrupar por fases que reflejen de un modo más general en qué punto del aprendizaje se encuentra el estudiante: *"La división en niveles solo establecerá el grado mínimo de dominio de la fonética del español se le debe exigir a un alumno en cada uno de dichos estadios. Convendría hablar de fases, de etapas globales que progresan desde un primer contacto hasta el estadio en el que el estudiante ya se siente seguro al expresarse en el nuevo idioma porque su pronunciación ha ido perfeccionándose gradualmente"*:

- **Fase de aproximación**, etapa A (que incluye los niveles A1-A2).
- **Fase de profundización**, etapa B (que incluye los niveles B1-B2).
- **Fase de perfeccionamiento**, etapa C (que incluye los niveles C1-C2).

A continuación se describen cada una de las fases, comenzando por la de aproximación: *"En la fase de aproximación se pone énfasis en la toma de conciencia por parte del estudiante de las características fundamentales de la pronunciación del español. Se persigue como objetivo fundamental el reconocimiento de los patrones fónicos del español y la producción de sus esquemas básicos.*

En la fase de profundización se busca que el alumno ajuste cada vez más su pronunciación a la del español y sea capaz de expresar determinados estados emocionales a través de ella. El estudiante deberá aproximarse más a la base articulatoria del español, pronunciar correctamente las secuencias vocálicas y consonánticas en el interior de una palabra y en el contexto de los enunciados, y emitir estos con las inflexiones tonales adecuadas.

Finalmente, en la fase de perfeccionamiento, la disposición articulatoria general de la que parte el estudiante en sus enunciados es muy similar a la del español. Se intentarán pulir los matices de pronunciación de manera que el emisor se asemeje propiamente a un nativo. La entonación debe ser adecuada no solo a los estados de ánimo, sino también a las intenciones pragmáticas (ironía, cortesía...) y el hablante debe alcanzar la capacidad de modificar el tempo y la articulación de acuerdo con la situación comunicativa y el registro."

3.4. Aspectos de la Pronunciación en Español: Patrones de Entonación

En este apartado se resumen algunos rasgos generales del español relacionados con la prosodia, en contraste con otros idiomas, para poder comprobar los efectos que pueden provocar estas características

Capítulo 3: La Evaluación de la Entonación en la Enseñanza del Español

en el habla de nuestro idioma por parte de estudiantes extranjeros. Además, se incluyen los patrones de entonación básicos junto con el esquema básico de un ejemplo de sus realizaciones.

1) Entonación

- El español es una lengua tonalmente grave en comparación a otras como el italiano, francés, inglés o alemán.
- Se trata de una lengua con una amplitud tonal (fluctuaciones tonales) media, menor que el italiano o el inglés, y mayor que el francés o el alemán. Es normal que hablantes italianos o estadounidenses tiendan a producir grandes modulaciones tonales (creando un efecto tonal de exageración para nativos), y que hablantes procedentes de Francia o Alemania tiendan a lo contrario, aparentando cierta "monotonía" tonal cuando hablan español. Dentro de nuestro propio país, las variedades andaluza o canaria presentan una mayor amplitud tonal media que la castellana.

2) Pausas, ritmo y tempo

- El español tiene un ritmo de isocronía silábica (todas las sílabas se pronuncian aproximadamente con la misma duración y cualidad sonora), frente al inglés o el alemán, que tienen una isocronía acentual (mismo tiempo de realización silábica de las sílabas acentuadas independientemente del número de sílabas átonas que haya entre ellas).
- Suelen producirse grupos de entonación que oscilan entre cuatro y doce sílabas entre pausa y pausa natural, dependiendo de la situación. En las mismas circunstancias, estos grupos suelen ser más cortos que en italiano y más largos que en francés.
- Como en todas las lenguas románicas, se realizan enlaces entre las vocales iniciales de las palabras y las vocales o consonantes finales de las palabras que las preceden. En otras lenguas (inglés, danés, alemán, sueco) estas realizaciones no son frecuentes, puesto que tienen más consonantes contiguas que vocales y otras pronunciaciones distintas a partir de golpes de glotis.

3) Acentuación

- La sílaba tónica ocupa una posición libre, frente al francés (posición final fija) o el checo (posición inicial fija). En su realización interviene principalmente la intensidad, aunque pueden participar otros factores como la duración y el tono.

Los patrones de entonación son las curvas melódicas que describimos al pronunciar una frase, y suelen indicar su propósito comunicativo. En general, la entonación de una oración en español consta de una primera rama ascendente que comprende desde el primer sonido hasta el primer acento tónico. A partir de aquí presenta variaciones en ciertas sílabas tónicas y va subiendo y bajando, hasta la parte del último acento hasta el final, donde dependiendo de la modalidad que se adopte, terminará de forma ascendente en oraciones interrogativas; y descendente en oraciones enunciativas.

A continuación se exponen algunos ejemplos procedentes del "Manual de entonación española"(NAVARRO TOMÁS, T., 1974) recogidos de la página personal de Joaquim LListerri de la Universitat Autònoma de Barcelona:

Formas enunciativas

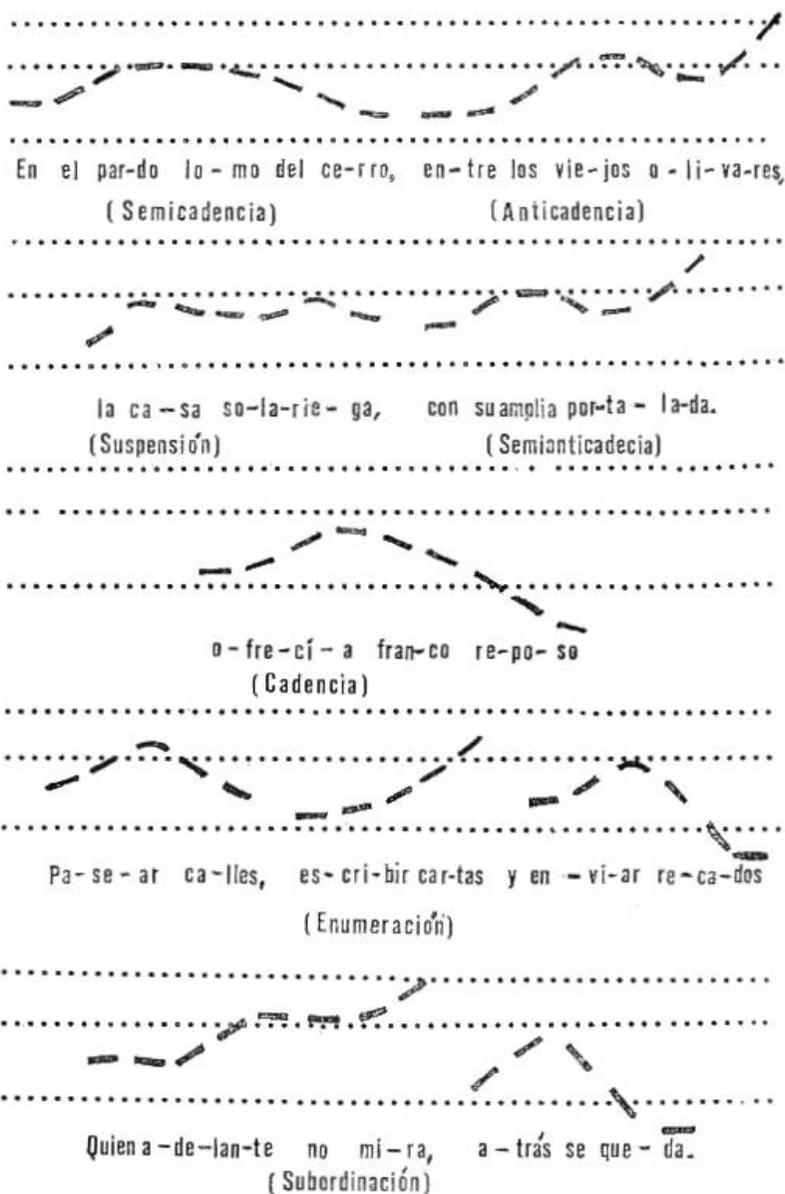


Ilustración 2: Formas Enunciativas (Fuente: LLISTERRI, Joaquim. "La descripción fonética y fonológica del español: los elementos suprasegmentales")

Formas interrogativas

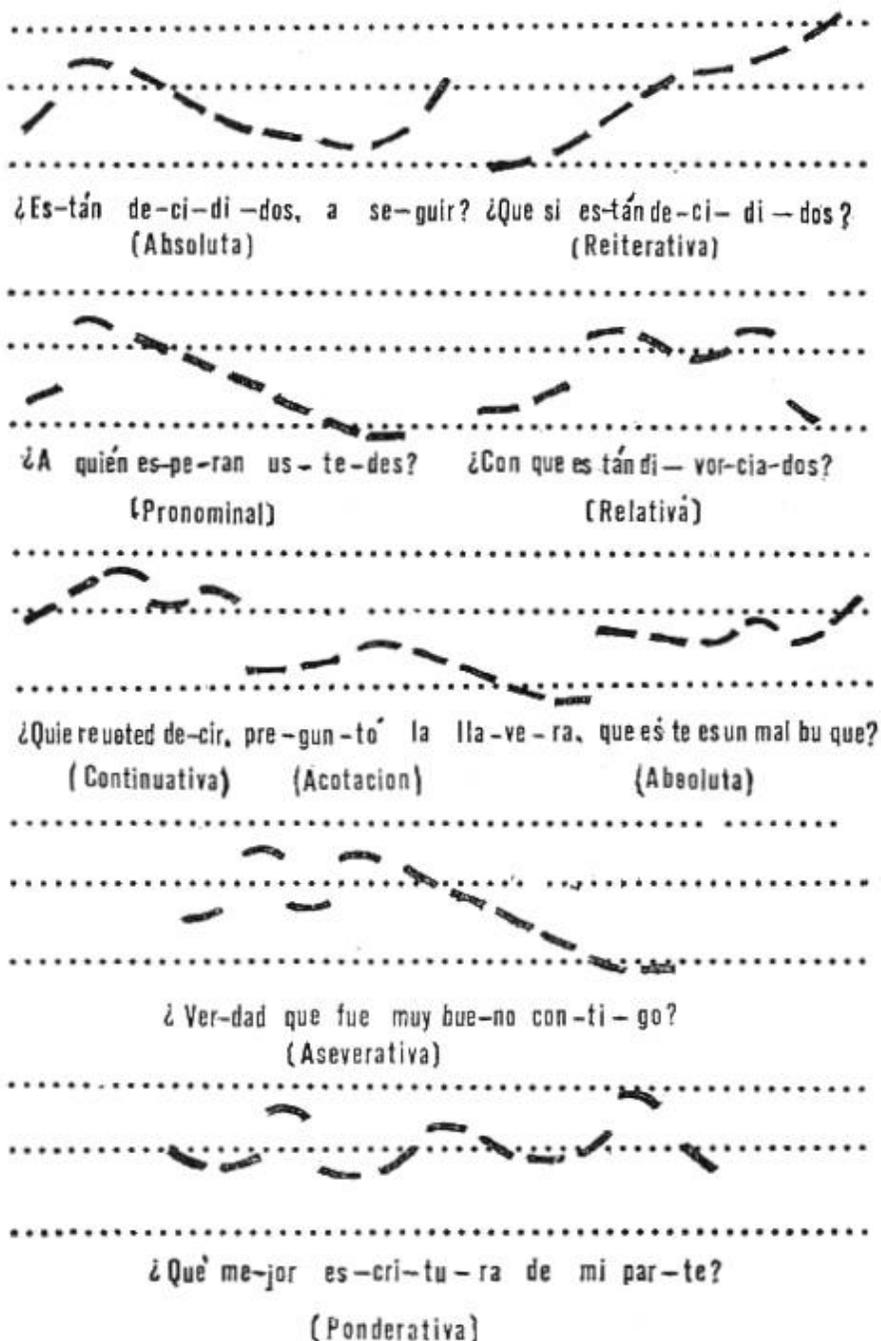
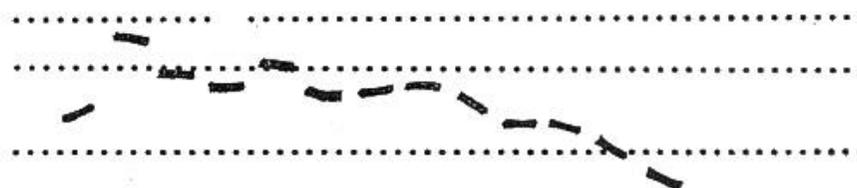
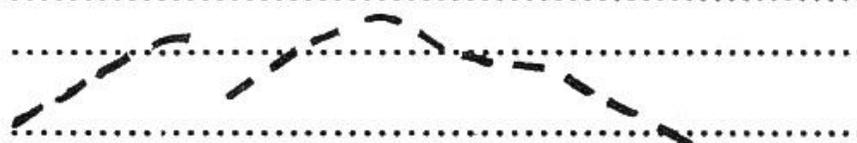


Ilustración 3: Ilustración 2: Formas Interrogativas (Fuente: LLISTERRI, Joaquim. "La descripción fonética y fonológica del español: los elementos suprasegmentales")

Formas volitivas



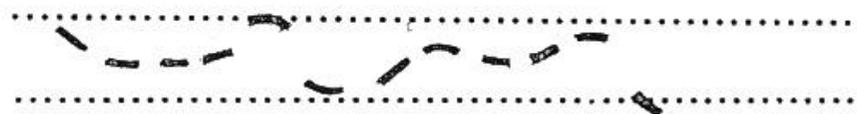
Da-re-mos u - na vuel-ta hasta la ho-ra de co-mer
(Invitación)



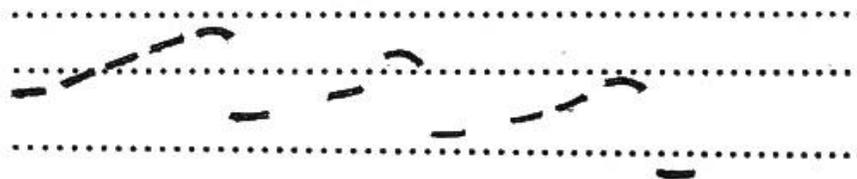
Hay que ac-ti-var, hay que ac-ti-var e-sa hue-na o-bra
(Exhortación)



An-tes que te ca-ses, mí-ra lo que ha-ces.
(Recomendación)



¡O-ra por nos-o-tros, se-ñor de los tris-tes!
(Ruego)



¡Ven-gaa-cá con-mi-goi !No te-mai !no se a-pu-reí
(Apelación tranquilizadora)

Ilustración 4: Ilustración 3: Formas Volitivas (Fuente: LLISTERRI, Joaquim. "La descripción fonética y fonológica del español: los elementos suprasegmentales")

Formas afectivas

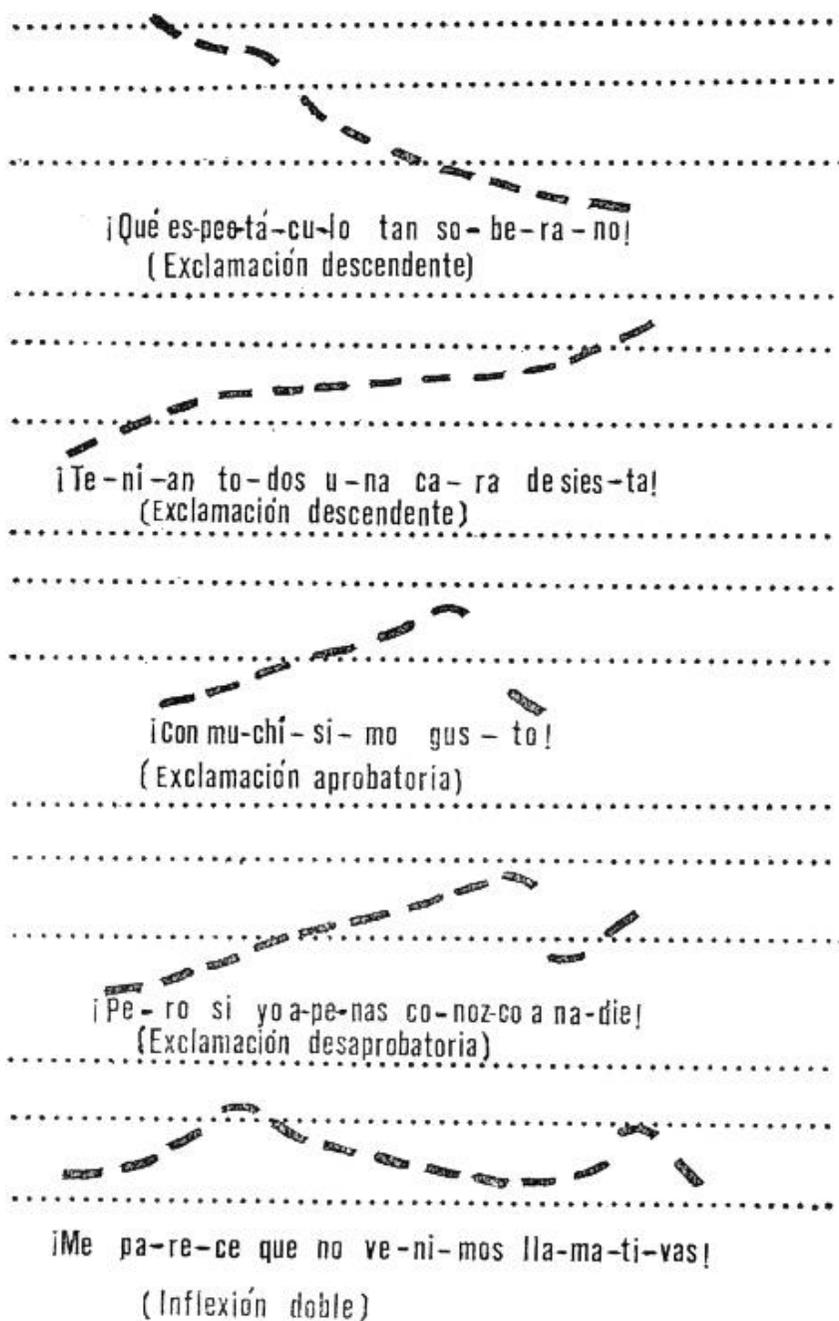


Ilustración 5: Ilustración 4: Formas Afectivas (Fuente: LLISTERRI, Joaquim. "La descripción fonética y fonológica del español: los elementos suprasegmentales")

Marco Conceptual: Aspectos Clave para el Análisis de la Entonación

Para analizar la entonación, necesitamos tener presentes ciertos conceptos que suelen aparecer de forma recurrente en los estudios referentes al habla humana. En este capítulo se introducen los aspectos característicos del habla en los que nos centraremos para desarrollar nuestro módulo de evaluación de la entonación. Se realizará un breve apunte sobre el acento en las sílabas y las teorías fonológicas, y posteriormente se explicarán los parámetros específicos que consideraremos en nuestro trabajo para caracterizar las locuciones de los usuarios y estudiar las posibles diferencias que existan entre las mismas y las de los locutores de referencia.

4.1. Conceptos Generales de Interés

Una representación acústico-fonética (procesamiento lingüístico de la señal; se trata de identificar los segmentos o fonemas del habla; las claves acústicas se acoplan a los rasgos fonéticos que son representaciones abstractas mediadoras entre los planos físicos (acústico) y lingüístico (fonético)) de la prosodia tiene aplicación práctica como herramienta para la enseñanza de la pronunciación asistida por ordenador.

El término acento (stress) se refiere a la prominencia (elevación de algo sobre lo que está alrededor o cerca de ello) perceptual relativa de las unidades del habla (discurso) mayores que segmentos fonémicos. El término entonación se refiere a la manipulación del tono (pitch) con propósitos lingüísticos y paralingüísticos sobre el nivel de los segmentos fonémicos.

Según el Diccionario de la Real Academia Española de la Lengua, cuando hablamos de fonética nos referimos a todo aquello que es perteneciente o relativo a la voz humana. Además, dicho de un alfabeto, de una ortografía o de un sistema de transcripción se refiere a la representación de los sonidos con mayor exactitud que la ortografía convencional. Una última acepción nos indica que la fonética es el conjunto de los sonidos de un idioma.

"Segmento", es un término usado en fonética y lingüística principalmente para hacer referencia a cualquier unidad discreta que se puede identificar en la cadena hablada, ya sea por medios físicos o auditivos: en el primer caso se pueden encontrar puntos de cambio acústico o articulatorio en los límites de los segmentos; en el segundo caso, la base de la división son los cambios perceptibles en la cualidad o la cantidad, que a menudo muestran la influencia de los fonemas de la lengua. El término se usa especialmente en fonética, en donde la unidad discreta más pequeña que se percibe recibe el nombre de "Fono".

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

También como se apunta en el Diccionario de la RAE, un fonema es cada una de las unidades fonológicas mínimas que en el sistema de una lengua pueden oponerse a otras en contraste significativo. Por otra parte, se indica que dentro de cada fonema caben distintos alófonos.

Así, los alófonos serían cada una de las variantes que se dan en la pronunciación de un mismo fonema, según la posición de este en la palabra o sílaba, según el carácter de los fonemas vecinos, etc.; p. ej., la *b* oclusiva de *tumbo* y la fricativa de *tubo* son alófonos del fonema /b/.

Para ofrecer un diagnóstico de la pronunciación de un estudiante con respecto a características de la prosodia, es necesario describir la composición prosódica del habla del estudiante y determinar cualquier desviación significativa de una pronunciación cercana a la nativa. De este modo, Bagshaw postula en su tesis doctoral "Automatic Prosodic Analysis For Computer Aided Pronunciation Teaching" (1994), que el único medio seguro disponible para describir la prosodia en la enseñanza de un idioma extranjero consiste en una representación acústico-fonética. Por otra parte, se afirma que para la evaluar la prosodia los parámetros que más influyen son la medidas de duración, la frecuencia fundamental del habla, la energía y la calidad de las vocales, aunque este último caso es descartado al no poderse demostrar, a través de su estudio, tal influencia.

El análisis automático del habla en la enseñanza de idiomas asistida por ordenador utiliza la onda acústica digital muestreada como único parámetro de entrada. De esta señal se deriva una representación acústico-fonética de los aspectos prosódicos del habla. Los datos procedentes de esta onda son los que trataremos de comparar entre diferentes locuciones para calcular la distancia de una locución de una frase en contraste con otra que haya servido de modelo, de modo que podamos evaluar este ensayo asignándole una puntuación.

La definición de sílaba (sonido o sonidos articulados que constituyen un solo núcleo fónico entre dos depresiones sucesivas de la emisión de voz) está asociada con los grados relativos de sonoridad de los segmentos fonémicos. La energía tiene una correlación acústica con la sonoridad de los segmentos fonémicos. Sin embargo, las medidas relacionadas con la prominencia de las sílabas dependen de la definición de la sílaba. Además de desempeñar un rol a bajo nivel en la extracción de medidas acústicas como la duración, las sílabas se utilizan como el dominio fundamental de las descripciones prosódicas.

El proceso de localizar sílabas prominentes es complicado por la interacción de las características acústicas (duración, energía, y frecuencia fundamental) en la manifestación del acento (stress) y por el hecho de que cada una de las características está influenciada por otros factores que no son el acento. Por tanto, la relación de estas características no es simple.

La silabificación del habla conectada constituye una parte central del análisis prosódico automático, desempeñando tres roles:

- 1) Los aspectos prosódicos del habla se describen en dominio silábico, el acento (stress) se refiere a la prominencia perceptual relativa de las sílabas y la descripción de la entonación involucra la asociación de acentos tonales (pitch accents) (relieve que se le da a una sílaba en una palabra mediante una elevación en el tono de la voz) con sílabas prominentes.
- 2) La extracción de parámetros acústicos depende de la definición e identificación de sílabas.
- 3) (Relacionado con la integración de los parámetros acústicos) El dominio silábico abarca de forma inherente información segmental y de energía, que puede ser transmitida a los análisis de las medidas de duración, energía y F0.

4.2. Descripción de la Entonación en la Enseñanza del idioma

El término "acento" (stress) se refiere a la prominencia perceptual relativa de las unidades del habla (discurso) mayores que segmentos fonémicos (sílabas o palabras) en un contexto particular, mientras que "entonación" se refiere a la manipulación del tono (pitch) con propósitos lingüísticos y paralingüísticos sobre el nivel de los segmentos fonémicos.

Podemos distinguir varias categorías de acentos:

- Acento léxico: se refiere a la posición de las sílabas relativamente prominentes en palabras habladas de forma aislada.
- Acento gramatical: se refiere a los patrones de acento léxico que aparecen en frases o en habla conectada, que se rigen por la gramática del idioma. El uso impropio de las fuerzas relativas de este tipo de acento puede resultar en ambigüedades sintácticas.
- Acento enfático: se refiere a las sílabas que se hacen prominentes conscientemente para poner énfasis en una palabra partícula.
- Acento contrastivo: se refiere a la posición de las sílabas prominentes para aclaraciones sintácticas.

Los tres acentos que se producen en habla conectada, gramatical, enfático y contrastivo, suelen ser denominados como "acento oracional".

Las sílabas prominentes se pueden subdividir en sílabas "accented" y en sílabas "stressed" ("non-accented"). Según Beckman, las sílabas prominentes están señaladas por la duración, intensidad y calidad de las vocales. Los acentos tónicos ("pitch accents") se toman como características de entonación (sin embargo antes se indicó que estaban asociados con las sílabas prominentes y por tanto sirven como señal de las mismas).

Las sílabas prominentes que no están asociadas con un acento tónico ("pitch accent") son las sílabas "stressed", mientras que aquellas que los están son las sílabas "accented". "Non-stress accent" se refiere a los tonos que son características entonacionales no relacionadas con una sílaba prominente. El resto de las sílabas son "unstressed".

Los incrementos en la energía promedio de una sílaba y/o los incrementos de duración de una sílaba completa, y los incrementos en el pico de la frecuencia fundamental de una sílaba pueden ser utilizados para manifestar varios grados de énfasis (Godfrey & Brodsky, 1986). No se debería confundir con los grados de acento oracional, aunque están estrechamente relacionados.

Para enseñar la entonación de un idioma a un estudiante, es esencial comprender y describir los fenómenos suprasegmentales propios de tal idioma. También es importante describir la composición prosódica del habla del estudiante, para determinar desviaciones significativas de una pronunciación cercana a la nativa, y para poder ofrecer algún tipo de corrección.

El habla de un hablante no nativo puede estar altamente influenciada por la prosodia del idioma fuente y por los estereotipos del idioma objetivo. Esto hace que una representación fonológica (mientras que la fonética estudia la naturaleza acústica y fisiológica de los sonidos o alófonos, la fonología describe el modo en que los sonidos funcionan) de la prosodia específica del idioma sea inapropiada para describir los aspectos prosódicos del habla de un estudiante extranjero.

4.3. Teorías Fonológicas de la Prosodia

La fonología se define como el estudio de las unidades sonoras en un idioma junto con el estudio de la estructura de la relación entre esas unidades de sonido. El uso de las teorías fonológicas puede ser muy útil para enseñar un idioma. Si los patrones de acento y entonación de un idioma específico van a ser enseñados a un estudiante, se requiere saber cómo describirlos previamente. Será también necesario describir los patrones de acento y entonación del estudiante para determinar posibles desviaciones de una prosodia similar a la nativa, y poder ofrecer un diagnóstico correctivo.

4.3.1. Teoría de Configuración

La primera de estas teorías es la de Crystal (1969). Se postula que la entonación de una locución puede ser descrita en términos de unidades tónicas, que son consideradas como la unidad primaria de estructura entonacional.

Así, cada unidad tónica está compuesta de, por los menos, cuatro componentes (precabeza, cabeza, núcleo y cola). El núcleo es la única parte obligatoria de una unidad tónica.

La sílaba prominente final (acentuada, "accented") de la unidad tónica forma el núcleo de la unidad tónica, y está asociada con el acento primario ("primary stress", pico de prominencia) en la unidad tónica. La cola consiste en cualquier sílaba adicional que aparezca después del núcleo. La cabeza la componen las sílabas de la unidad tónica que van desde la primera sílaba prominente acentuada hasta (sin incluirla) la sílaba tónica. La precabeza representa cualquier sílaba que soporta acento terciario y sílabas no acentuadas que existen antes de la primera sílaba prominente (o de la cabeza, o del núcleo si no hay cabeza) acentuada en la unidad tónica.

Cristal propone cuatro tipos primarios de movimientos de núcleo del acento (nuclear pitch movements), "rise", "fall", "rise-fall", "fall-rise". Cada movimiento está representado fonéticamente por unos niveles de tono (high onset, mid onset, low onset...) y unos rangos de tono (ancho, estrecho). Hay cuatro posibles combinaciones de estos tonos primarios que pueden suceder dentro de una única unidad tónica:

- "rise"+"fall"
- "fall-rise"+"fall"
- "fall"+"rise"
- "rise-fall"+"rise"

Esta descripción está basada en la entonación de un Inglés nativo, por tanto puede ser inadecuada para describir la entonación de un hablante extranjero.

4.3.2. Teoría de la secuencia de tonos

La siguiente teoría fonológica (Pierrehumber, 1980) describe una curva de tono como una serie de interpolaciones entre tonos-objetivo ("pitch targets") sucesivos. Los tonos-objetivo son realizaciones fonéticas de dos segmentos tónicos fonológicamente distintos, un tono alto (H) y un tono bajo (L).

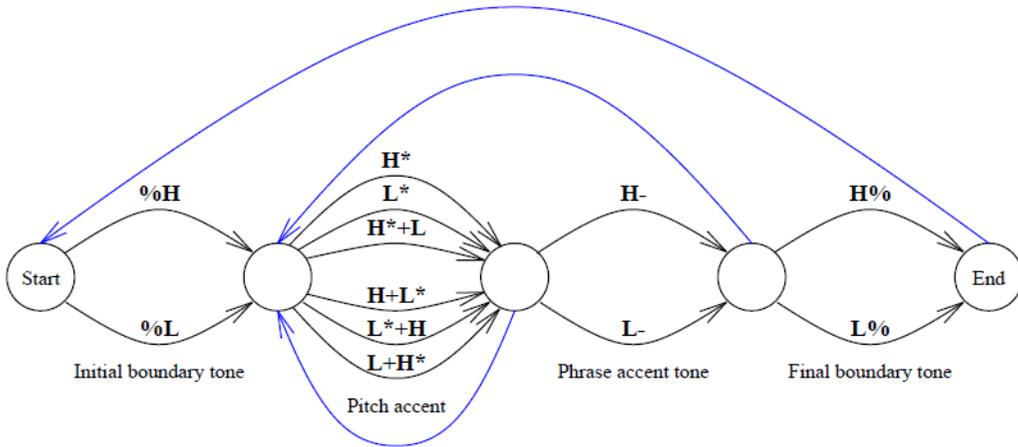


Ilustración 6: Diagrama de estados finito de posibles secuencias de tonos H/L (Pierrehumber, 1980)

Se distinguen:

- Tonos relacionados con el acento tónico ("pitch accent related tones"): se designan a la localización de sílabas que soportan acento oracional. Para representar estos tonos asociados con sílabas prominentes acentuadas ("accented"), se utiliza "*". Los seis acentos tónicos ("pitch accents") son H*, L*, H*+L, H+L*, L*+H y L+H*. Cada acento tónico se realiza a través de cualquier número de sílabas adyacentes, denominadas como palabra prosódica.
- Tonos de acento de frase ("Phrase accent tones"): son especificados por la gramática para existir únicamente entre el tono final relacionado con un acento tónico ("pitch accent") y un tono límite ("boundary tone"). Se marcan con "H-" y denominan frases intermedias.
- Tonos límite ("boundary tones"): se asocian con un mayor nivel de fraseo ("phrasing") que los tonos de acento de frase ("phrase accent tones"). Se marcan con "H%" y denominan frases entonacionales.

Una observación común es una tendencia descendente en el tono, como norma general, en el transcurso de una frase entonacional. Pierrehumber lo atribuyó al efecto fonológico de "downstep" (sin traducción), una reducción gradual de altos tonos-objetivo ("pitch targets") en momentos controlados por el hablante; reducción del rango de tonos que disminuye la realización de F0 de cualquier tono alto que siga a un determinado "trigger". Se marca como "!H". Esta tendencia de forma prolongada puede ser atribuida a una propiedad fisiológica conocida como "declinación" (Vassière, 1983).

4.4. Análisis Acústico-Prosódico

El análisis automático del habla en la enseñanza de la pronunciación asistida por ordenador utiliza una señal con forma de onda digitalmente muestreada como único parámetro de entrada. Puede ser capturada fácilmente con un micrófono sin ser molesto para el hablante o, como en nuestro caso, con el sistema de entrada vocal de un smartphone que recoge una grabación.

La señal acústica del habla transmite información lingüística y paralingüística. La señal con acústica con forma de onda se percibe como una secuencia de segmentos fonémicos que constituyen las

palabras de una locución. También muestra fenómenos suprasegmentales que están relacionados con las estructuras sintácticas y prosódicas de una locución. Se trata de una señal extremadamente compleja que transmite muchos otros aspectos del habla además de la prosodia. Para realizar un análisis prosódico automático del habla es necesario identificar parámetros acústicos correlacionados con la prosodia. Los que se exponen a continuación son los que consideraremos en nuestro estudio.

4.4.1. La onda sonora

La fonación es el proceso mediante el cual se produce en la voz, resultado de la acción de la laringe. Las cuerdas vocales se abren y se cierran con un movimiento ondulatorio, mientras que los músculos que las controlan regulan el grosor y la tensión para producir diferentes frecuencias de vibración.

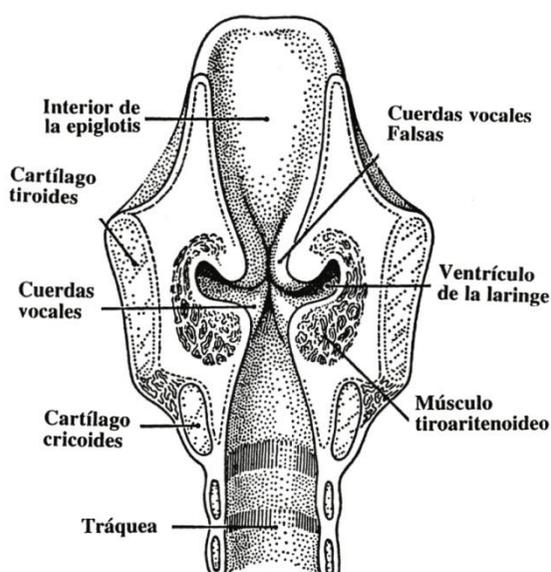


Ilustración 7: Interior de la Cavidad Laríngea (Fuente: La Fonación - Página Web personal de Joaquim Listerri)

La onda sonora producida en la fonación es el resultado del paso de aire por la glotis en la emisión de una serie de sucesivas bocanadas de aire al ritmo de abertura y cierre de las cuerdas vocales. Las moléculas de aire vibran al pasar por los pliegues vocálicos y se produce el sonido. Puede definirse en función de la duración del movimiento de vibración, su frecuencia y su energía.

4.4.2. Medidas de Duración

Cuando hablamos de duración, nos referimos a la longitud de un componente particular del discurso. En el contexto del análisis prosódico, este componente suele ser un tipo de unidad fonética o fonológica. La duración no es estrictamente un parámetro acústico, sin embargo, se considera que tiene una correlación acústica con los fenómenos prosódicos.

Para este trabajo, utilizaremos medidas de duración locales (relativas a las secuencias temporales de una locución) y globales (relativas al total de la locución) de determinados puntos de interés de la señal (máximos y mínimos de frecuencia fundamental y energía). Como estas medidas en principio no nos aportarán mucha información para una comparación (por regla general dos locuciones se realizarán a velocidades diferentes, y por consiguiente no serán comparables temporalmente), emplearemos además medidas de tiempo normalizadas, a priori más razonables puesto que consideran los puntos de interés en el contexto un entorno concreto.

4.4.3. Frecuencia Fundamental del Habla (F0)

La frecuencia fundamental (F0) del habla se define como la proporción de pulsos glotales generados por la vibración de las cuerdas vocales durante la reproducción en voz de los segmentos. El tono del discurso es la correlación perceptual de F0. Las escalas psicoacústicas son lineales sólo en relativas bajas frecuencias. Sin embargo, se asume que hay una correlación lineal entre el tono y F0 en bajos rangos de frecuencias que son relevantes para la reproducción del habla en hombres y mujeres (aproximadamente 50-250Hz y 120-400Hz, respectivamente).

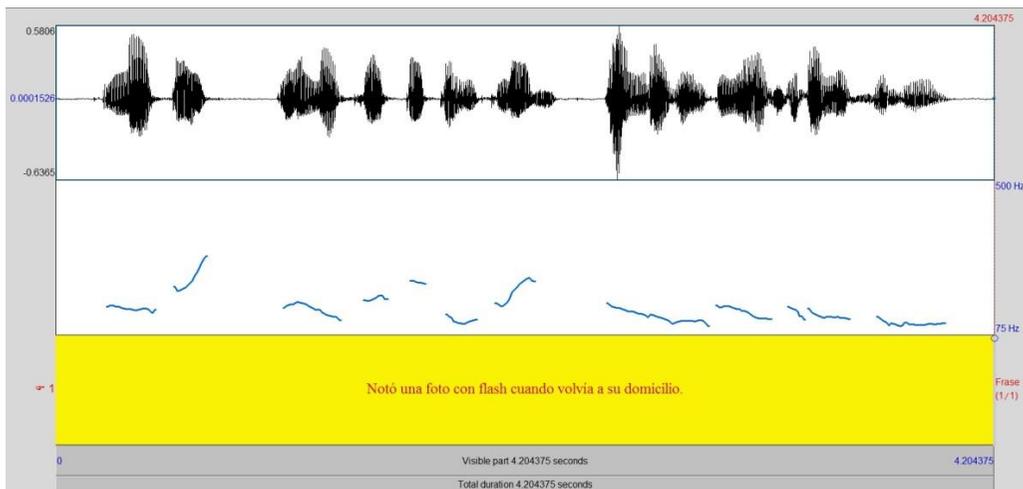


Ilustración 8: Señal de la Onda Sonora (parte superior) y Frecuencia Fundamental (parte central)

La frecuencia fundamental del habla desempeña un papel importante en las características prosódicas de acento y entonación, tiene una correlación con la prominencia percibida de las sílabas y la melodía característica de una locución. Su extracción de la señal acústica con forma de onda es necesaria como proceso inicial para el análisis prosódico de fenómenos suprasegmentales. La determinación de F0 no es una tarea sencilla.

Los métodos para determinar la frecuencia fundamental del habla operan o en el dominio de la frecuencia, o en el dominio del tiempo, o con una función de correlación. Los algoritmos basados en tiempo y correlación utilizan filtrado para intentar mejorar su rendimiento, mientras que los basados en frecuencia utilizan la señal en bruto.

Para extraer los valores de frecuencia fundamental del habla, el algoritmo de extracción pertinente recorre los datos de la señal acústica, recogiendo los valores de la onda sonora según su frecuencia de

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

muestreo. Estos valores son tratados por lo que se denomina frames o ventanas de análisis, que analizan los datos en un pequeño intervalo (ancho del frame). Este análisis local está basado en la redundancia en las señales del habla (periodicidad y variación relativamente lenta), de modo que a cada intervalo analizado por un frame, se le asigna el valor promedio de los datos extraídos de la onda sonora en ese intervalo (generalmente en su punto medio). Los frames no se suceden uno detrás de otro secuencialmente, sino que se solapan para que no haya pérdidas de información en los límites de las transiciones. El conjunto de los valores asignados a cada frame describirán la evolución de la frecuencia fundamental del habla (curva del tono, "pitch contour"), lo cual representa las variaciones de la frecuencia fundamental (eje vertical) a lo largo del tiempo (eje horizontal).

4.4.4. Energía

Según se expone en la página personal de Joaquim Llisterri (en la que se hace referencia al trabajo de Juana Gil, "Los sonidos del lenguaje"), en la curva de la señal acústica, podemos ver la energía como la amplitud, esto es, la distancia entre la posición de reposo y el punto de máximo desplazamiento. Depende de la fuerza aplicada inicialmente y de la resistencia que ofrezca el medio en el que se produce la vibración. Se cuantifica en unidades de presión sonora desde el punto de vista físico o en unidades de intensidad (decibelios, dB) desde el punto de vista perceptivo.

La curva de la señal acústica se utiliza como una representación acústica de las variaciones de energía percibidas en una locución. La curva de energía se reduce desde el nivel del marco ("frame") en el que se recoge la muestra a un nivel de sonido, de manera que sonido de una locución se asocia con una única medida de energía.

4.4.5. Caracterización del habla con vectores de parámetros

Una vez ya conocemos los parámetros con los que vamos a trabajar, necesitamos elaborar un vector que combine las diferentes medidas que obtengamos de los mismos procedentes de las grabaciones de las locuciones realizadas por los usuarios. Este vector de parámetros lo utilizaremos para caracterizar dichas locuciones y compararlas con las que utilizemos como referencia. Como se ha podido comprobar tras la consulta de varios estudios y documentos de referencia, esta clase de vector suele incluir valores estadísticos de los parámetros seleccionados.

En nuestro caso, como nos vamos a centrar en medidas relacionadas con la duración, la frecuencia fundamental del habla y la energía, podemos probar con medidas de máximos y mínimos, valores promedios, desviaciones e incluso rangos. Estas medidas no sólo las podremos utilizar en la comparación de las locuciones en global, sino que también seremos capaces de realizar comparaciones de intervalos dentro de cada locución, logrando así efectuar un análisis más detallado, considerando los datos a nivel de diferentes tipos de unidad entonativa: frase, palabra y sílaba.

Por la naturaleza de la entonación, como trabajaremos directamente sobre los datos de la onda que representa el contorno de la señal acústica, no podemos (a no ser que realicemos un etiquetado previo), conocer qué partes de la onda se corresponden con una palabra o una sílaba. Sin embargo, sí que obtendremos del procesamiento de la señal qué segmentos son sonoros o sordos. Los datos de estos intervalos también pueden contener información valiosa para el análisis de la entonación, por ello, son los que emplearemos para realizar la comparación a un nivel más "local".

Preparación del Corpus

El objetivo principal del proyecto es el desarrollo de un módulo de evaluación de la entonación capaz de determinar la distancia de una locución en comparación con una de referencia y de dictaminar si una locución ha sido lo suficientemente buena como para clasificarla como una entonación propia de un nativo. Como se expuso anteriormente, este sistema debemos entrenarlo con un corpus del que podamos extraer una serie de datos que controlemos y que utilicemos como referencia al computar los ensayos de los usuarios. En este capítulo nos centraremos en cómo se ha preparado el corpus recopilado con tal motivo y las decisiones que se han tomado fruto del análisis de sus datos.

5.1. ¿Qué es un corpus?

Un corpus es un conjunto (habitualmente) amplio de ejemplos reales de uso de una lengua. En nuestro caso, se trata de un conjunto de muestras orales almacenadas en formato electrónico, las cuales procesaremos para realizar un análisis lingüístico que nos servirá para elaborar un sistema de evaluación de la entonación del español, aplicable en sistemas de aprendizaje del español como segunda lengua (L2).

5.2. Recopilación del Corpus

5.2.1. Protocolo de grabación

- El protocolo de grabación consta, básicamente, de las siguientes fases:
 - Bienvenida y explicación: se indica al usuario qué vamos a hacer y por qué.
 - Firma de autorización: para uso de sus datos con fines de investigación.
 - Control de niveles de grabación: el operador charla informalmente con el donante mientras comprueba los niveles de audio
 - Cada hoja supone un cambio de pista.
 - En el experimento de repetición, se dispone de un tablet con las frases de referencia grabadas por un locutor profesional y se activa el audio del mismo justo en el momento que se necesita.
 - En la lectura de noticias, se le da un tiempo prudencial para leer las noticias en voz baja para las 10 primeras noticias. Para las últimas, se lee a primera vista.
- **Tareas:**
 - **Frases:**
 - Leer frases a primera vista (10): tres repeticiones.
 - Repetir frases leídas por un locutor profesional (10). Las cinco primeras son las últimas cinco de la secuencia de frases anterior: tres repeticiones
 - **Fábula:** Se le da el tiempo que precise para leer en voz baja y entender la fábula de Esopo "El Sol y el Viento del Norte". Se le pide que la lea como si se tratase de un cuento o narración corta dirigida a un niño.

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

- **Noticias:** Lee hasta 15 noticias, agrupadas en páginas que contienen una o dos, dependiendo de la longitud de la noticias.
- **Cuestiones generales:**
 - Cada página representa una grabación en pista diferente.
 - Se anota en una hoja de control el número de pista y el tipo de grabación de la misma (ver modelo **20130503-PlantillaRegistroGrabacion.docx**)
 - Cada pista se almacena en el SD de la grabadora con el nombre 10NN.wav, siendo NN el número de pista.
 - En cada grabación siempre estaban presentes dos controladores del grupo. Uno actúa como operador y el otro de apoyo.
 - Se emplea un micrófono de ambiente profesional Neumann y una grabadora digital Marantz.
 - Se utiliza grabación D-mono: un micro y dos canales, uno de ellos atenuado automáticamente por la grabadora (dubbing mono-stereo).
 - Se graba con calidad CD en formato WAV.

5.3. Contenidos del Corpus

5.3.1. Locutores

- Locutores de referencia: Locutores femeninos y masculinos del corpus glissando. Ver tabla siguiente:

NumLoc	Sexo	SpeakerID	recordingFolder
1	F	fr1	sp_f16a
2	F	fr2	sp_f11r
3	F	fr3	sp_f13r
4	F	fr4	sp_f15a
1	M	mr1	sp_m09a
2	M	mr2	sp_m10a
3	M	mr3	sp_m12r
4	M	mr4	sp_m14r

Tabla 2: Locutores de Referencia del Corpus y sus correspondientes sexos, identificadores y directorios

- El recordingFolder es el nombre de la carpeta dentro de call2013/frases en la que se encuentran los wavs en crudo de ese locutor.
- El speakerID será el nombre de la carpeta en call2013/corpus/data en que se almacenará todos los ficheros de datos asociados a las locuciones de ese locutor.

- Locutores extranjeros: Ver la tabla siguiente

NumLoc	Sexo	Nacionalidad	SpeakerID	recordingOrder	nickName	recordingFolder
01	F	USA	f01	01	robin	loc01-robin
02	F	USA	f02	03	hannah	loc03-hannah
03	M	USA	m03	06	hiromasa	loc06-hiromasa
04	F	JPN	f04	02	amanda	loc02-amanda
05	F	USA	f05	13	ashley	loc13-ashley
06	F	USA	f06	04	kristie	loc04-kristie
07	F	USA	f07	07	crystal	loc07-crystal
08	M	USA	m08	14	michael	loc14-michael
09	F	USA	f09	08	danielle	loc08-danielle
10	F	USA	f10	05	kendall	loc05-kendall
11	F	JPN	f11	09	tomoko	loc09-tomoko
12	F	JPN	f12	11	miyabi	loc11-miyabi
13	F	JPN	f13	10	ikumi	loc10-ikumi
14	F	JPN	f14	12	ayana	loc12-ayana

Tabla 3: Locutores Extranjeros del Corpus y sus correspondientes sexos, nacionalidades, identificadores, nick names, y directorios

- El recordingFolder es el nombre de la carpeta dentro de call2013/recordings en la que se encuentran los wavs en crudo de ese locutor.
- El speakerID será el nombre de la carpeta en call2013/corpus/data en que se almacenará todos los ficheros de datos asociados a las locuciones de ese locutor.

5.3.2. Frases

s01	La coalición interpuso esta querrela por prevaricación el viernes pasado.
s02	52 denuncias por faltas graves en dos años, 18 de ellas graves por carecer de licencia de funcionamiento. Y el bar sigue abierto.
s03	Para una gala que se celebrará el 8 de febrero del próximo año.
s04	ATT prevé eliminar 12.000 empleos y reducir inversiones de capital.
s05	Notó una foto con flash cuando volvía a su domicilio.
s06	<i>En la cartelera de cine no hay este fin de semana mucha poesía que digamos.</i>
s07	<i>¿Qué sería de una Navidad sin su cesta?</i>
s08	<i>Más de un millón de mujeres trabajan actualmente por cuenta propia.</i>
s09	<i>Y en los mercados los números rojos se extienden hoy por todas las bolsas europeas.</i>
s10	<i>No les han ofrecido hotel, ni tan siquiera a un vaso de agua.</i>
s11	Sin embargo, también hay una buena noticia. Existen soluciones.
s12	Sigue con sus trabajos de investigación, en los que ya constan sus conversaciones con la presidenta regional.
s13	Todos ellos, según las últimas informaciones del diario El País, fueron también víctimas de seguimientos.
s14	Esta investigación interna no ha dado aún ningún dato concluyente, y no tiene fecha límite.
s15	Hoy, hay huelga en las escuelas infantiles.

Tabla 4: Conjunto de Frases que componen el Corpus

En negrita cursiva las frases que aparecen tanto en la lectura a primera vista como en la repetición de la frase escuchada.

5.3.3. Estructura de Directorios

- .doc: Contiene toda la documentación sobre el corpus
- .data: Contiene, estructurados por locutor, todos los datos y wavs originales de cada locutor y tipo de experimento (frases, fábula, noticias, ...)
- .data/frases: Contiene, organizado por locutores, todas las frases leídas por los diferentes locutores en la tarea tanto de lectura a primera vista como repetición.
- .data/fabula: Contiene, organizado por locutores, las frases que componen la fábula leída por cada locutor.

- `.data/noticias`: Contiene, organizado por locutores, las frases de las noticias leídas por cada locutor.
- `.../[fm][0-9][0-9]/...`: Directorio del locutor correspondiente.
 - Cada locutor lee 15 frases diferentes repetidas tres veces:
 - 10 leídas sin escuchar: 5 se repetirán luego y 5 no.
 - 10 leídas tras escuchar (repetición): 5 nuevas y 5 de las anteriores.
- `[fm][0-9][0-9]s[0-9][0-9]r[0-9][0-9]`: Nombre base del fichero de frases para un locutor dado, indicando la frase leída (s01-s15), el ordinal de la repetición (r01-r06).
- `[fm][0-9][0-9]f[0-9][0-9]`: Nombre base del fichero de la frase fNN de la fábula leída por un locutor dado.
- `[fm][0-9][0-9]n[0-9][0-9]s[0-9][0-9]`: Nombre base del fichero de la frase sNN de la noticia nMM leída por un locutor dado.

5.3.4. Ficheros de Datos

Con el nombre base definido en el apartado anterior, tenemos, dentro de las carpetas correspondientes en la estructura de directorios establecida, los ficheros de datos extraídos de los archivos de audio, con las siguientes extensiones:

- `.csv`: Contiene, ordenados por locutor, número de frase y repetición, los datos que aparecen en los ficheros con formato `.ppar` y `.spar` de todas las cabeceras y de todos los intervalos de todas las frases, separados por `','`. Existen dos archivos `headersP.csv` y `headersS.csv`, con los valores de cabecera obtenidos de Praat y Snack, respectivamente. Con el mismo criterio existen dos archivos, `intervalsP.csv` e `intervalsS.csv`, con los valores de los intervalos de cada fichero de audio. A ambos tipos de fichero (cabeceras e intervalos) se les ha añadido por línea los campos de locutor (`loc`), frase (`sent`), repetición (`rep`), nativo (`nat`) y sexo (`sex`).
- `.ene`: Contiene, ordenados por tiempo, los valores temporales de las muestras tomadas sobre un fichero de audio con sus correspondientes valores de energía, obtenidos con SNACK.
- `.f0`: Contiene, ordenados por tiempo, los valores temporales de las muestras tomadas sobre un fichero de audio con sus correspondientes valores de f_0 (extraídas mediante ESPS), sus probabilidades de sonoridad, medidas cuadráticas de raíces locales y valores de correlación cruzada normalizados de pico; obtenidos con SNACK.
- `.f0a`: Contiene, ordenados por tiempo, los valores temporales de las muestras tomadas de un fichero de audio con sus correspondientes valores de f_0 , categorizaciones como muestras sonoras o sordas, valores de energía y coeficientes de correlación; obtenidos con SNACK.
- `.f0p`: Contiene, ordenados por tiempo, los valores temporales de las muestras tomadas de un fichero de audio con sus correspondientes valores de f_0 , categorizaciones como muestras sonoras o sordas, valores de energía y coeficientes de correlación; obtenidos con PRAAT.
- `.Intensity`: Contiene, ordenados por número de muestra, los valores de la energía de las muestras tomadas sobre un fichero de audio. En la cabecera aparecen los valores de tiempo de inicio y final del audio, número de muestras, paso y tiempo de la primera muestra.
- `.Pitch`: Contiene, ordenados por número de muestra, los valores de f_0 de cada candidato por cada muestra de las muestras tomadas sobre un fichero de audio. En la cabecera aparecen los valores de tiempo de inicio y final del audio, número de muestras, paso y tiempo de la primera

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

muestra, así como la mayor máxima frecuencia permitida y el número máximo de candidatos por muestra.

- ".PitchTier": Contiene, ordenados por muestra, los valores temporales de las muestras consideradas sonoras y sus valores correspondientes de f0, obtenidos con PRAAT. En la cabecera aparecen los valores de tiempo de inicio y de final del audio y el número de muestras.
- ".ppar": Contiene, ordenados por número de intervalo, valores temporales, de categorización de las muestras como sonoras o sordas, de f0 y de energía calculados a partir del fichero ".f0p" de cada intervalo y del total del fichero de audio.
- ".spar": Contiene, ordenados por número de intervalo, valores temporales, de categorización de las muestras como sonoras o sordas, de f0 y de energía calculados a partir del fichero ".f0a" de cada intervalo y del total del fichero de audio.
 - Cabeceras de los ficheros ".ppar" y ".spar": Compuestas por 9 líneas que contienen los parámetros de resumen calculados para el intervalo equivalente a la longitud total del fichero de audio. Por líneas, el contenido sería el siguiente:
 1. Duración total del fichero de audio.
 2. Número de intervalos, número de intervalos sonoros y número de intervalos sordos.
 3. Número de muestras, número de muestras sonoras, número de muestras sordas y número de valores espurios.
 4. F0 máxima, punto temporal en el que se alcanza F0 máxima y tiempo normalizado (punto temporal dividido entre el tiempo total, con resultado en [0,1]) en el que se alcanza F0 máxima.
 5. F0 mínima, punto temporal en el que se alcanza F0 mínima y tiempo normalizado en el que se alcanza F0 mínima.
 6. Rango de F0 (valor máximo menos valor mínimo, excluyendo valores sordos (con F0 igual a 0.0)), valor medio de F0 y desviación típica de F0.
 7. Energía máxima, punto temporal en el que se alcanza la energía máxima y tiempo normalizado en el que se alcanza la energía máxima.
 8. Energía mínima, punto temporal en el que se alcanza la energía mínima y tiempo normalizado en el que se alcanza la energía mínima.
 9. Rango de energía, valor medio de la energía y desviación típica de la energía.
 - Cuerpo de los ficheros ".ppar" y ".spar": Compuesto por líneas en formato registro que contienen los valores, separados por un espacio, correspondientes a cada intervalo en el que se divide el fichero de audio (una línea por cada intervalo), atendiendo a su categorización como intervalo sonoro o sordo, con el siguiente orden:
 1. Número de intervalo
 2. Categorización sonoro(1)/sordo(0)
 3. Tiempo de inicio del intervalo
 4. Tiempo de final del intervalo
 5. Duración del intervalo
 6. Número de muestras del intervalo
 7. F0 máxima en el intervalo
 8. Tiempo en el que se alcanza F0 máxima en el intervalo (con respecto al total)
 9. Tiempo normalizado en el que se alcanza F0 máxima en el intervalo

10. F0 mínima en el intervalo
 11. Tiempo en el que se alcanza F0 mínima en el intervalo (con respecto al total)
 12. Tiempo normalizado en el que se alcanza F0 mínima en el intervalo
 13. Rango de F0 en el intervalo
 14. F0 media en el intervalo
 15. Desviación típica de F0 en el intervalo
 16. Energía máxima en el intervalo
 17. Tiempo en el que se alcanza la energía máxima en el intervalo
 18. Tiempo normalizado en el que se alcanza la energía máxima en el intervalo
 19. Energía mínima en el intervalo
 20. Tiempo en el que se alcanza la energía mínima en el intervalo
 21. Tiempo normalizado en el que se alcanza la energía mínima en el intervalo
 22. Rango de la energía en el intervalo
 23. Energía media en el intervalo
 24. Desviación típica de la energía en el intervalo
- ".TextGrid": Contiene, ordenados por número de intervalo, los valores temporales de inicio y fin de cada intervalo y su etiqueta de frase (o silencio) correspondiente. En la cabecera aparecen los valores temporales de inicio y fin del audio y el número de intervalos.
 - ".wav": Fichero de audio que contiene una grabación de una frase realizada por un locutor.

5.4. Procesado del Corpus

Para procesar los datos procedentes de este corpus, nos hemos servido de dos productos software, Praat y Snack, utilizados para el estudio científico del habla. Se conocen otros sistemas similares (SPTK y HTK) que no se han empleado para el estudio del corpus.

Previo a la elaboración del sistema de evaluación, se deben preparar los datos extraídos del corpus para que puedan ayudar a tomar las decisiones oportunas que nos lleven a realizar una clasificación correcta de las locuciones en un futuro. En primera vuelta, se propondrá un "juego", en la que el sistema, después de un ensayo del usuario, deberá valorar el intento, estimando si el usuario pronuncia tan bien como un nativo o no.

Con este fin, se ha dividido la preparación del corpus en las tareas que se exponen a continuación.

5.4.1. Etiquetado de las grabaciones con Praat

Como se ha explicado en el apartado de "Protocolo de Grabación", tenemos las grabaciones correspondientes a las locuciones de 14 locutores extranjeros (12 de género femenino y 2 de género masculino) y de 8 locutores profesionales españoles (4 de género masculino y otras 4 de género masculino), agrupadas en "pistas" de 10 frases con el criterio ya comentado.

Para poder procesar cada locución de cada frase por separado, necesitamos, en primer lugar, etiquetar las grabaciones distinguiendo dentro de cada pista los segmentos que constituyen las

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

diferentes frases (10 en ejercicios de lectura, 20 en ejercicios de audición/ensayo), así como los silencios intermedios. Esta tarea ha sido realizada con la ayuda de Praat, para lo cual se ha procedido como se describe a continuación.

1. *Se selecciona la pista a etiquetar: en la pestaña "Open" de la barra de herramientas de Praat, pulsamos en la opción "Read from file..." del menú desplegable. Se mostrará la típica ventana de explorador de archivos en la cual podemos buscar y elegir qué archivo es el que vamos a cargar, en este caso, uno de sonido con extensión .wav.*

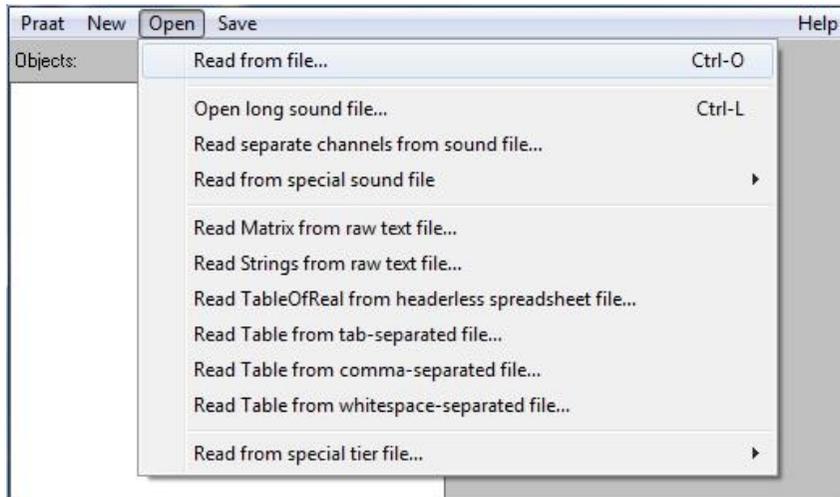


Ilustración 9: Cargar un archivo en Praat

El resultado es que aparece en la ventana de objetos de Praat, un objeto definido como [tipo_objeto][id], donde [tipo objeto] es la clase de artefacto (de los que reconoce Praat) que hemos cargado en el programa, e [id] es el nombre con el que aparece en nuestro sistema de archivos.

2. *Puesto que es posible que las pistas estén grabadas en formato estéreo, con el fin de facilitar el etiquetado, extraemos uno de los dos canales: con el objeto (Sound) seleccionado, pulsamos en el botón "Convert -" dentro del menú de opciones de la derecha del panel de objetos.*

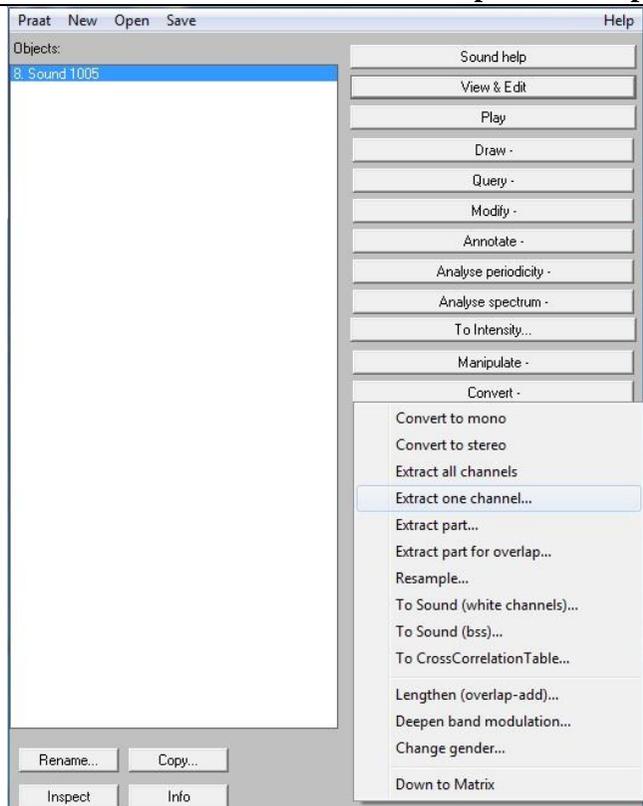


Ilustración 10: Extraer un canal de un objeto de sonido (Sound)

Se ha decidido extraer el canal de sonido en el cual la señal de audio tiene una amplitud mayor, es decir, el canal en el que a simple vista se aprecia mejor la onda del sonido (en el otro la señal tiene un tamaño menor y se distingue peor la curva). Este canal es el izquierdo. Seleccionando la opción "Extract one channel..." del menú que se despliega al pulsar en "Convert -", se muestra un cuadro de diálogo en el que podremos definir qué canal es el que utilizaremos.

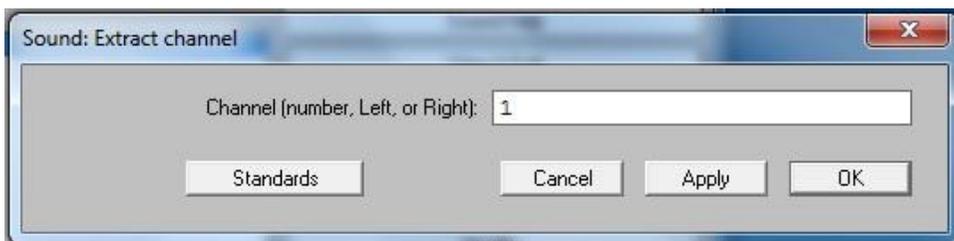


Ilustración 11: Extraer el canal izquierdo del objeto de sonido

Se crea un nuevo objeto llamado [tipo_objeto][id_ch1], el cual aparece en el panel de objetos añadido al final. Es sobre este objeto sobre el que realizaremos el etiquetado.

3. *Teniendo seleccionado el objeto de sonido de un único canal, pulsamos en el botón "Annotate -" del menú de la derecha, y la opción "To TextGrid..." del menú que aparece automáticamente.*

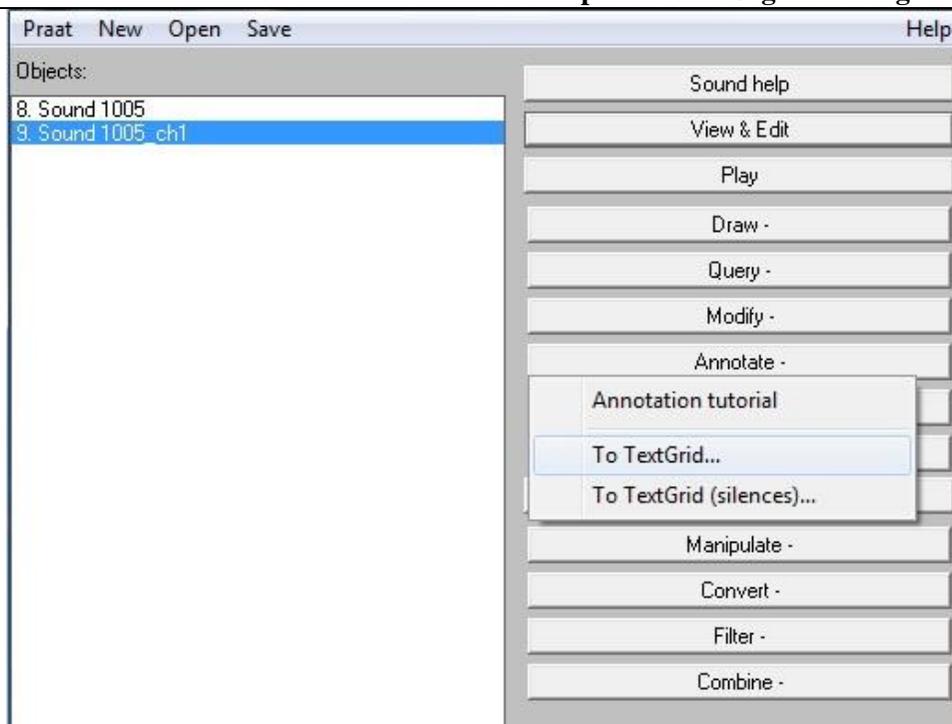


Ilustración 12: Etiquetar un archivo de sonido

Se muestra un cuadro de diálogo en el que podremos seleccionar el número de capas (Tiers), así como la denominación de referencia de las capas. Para esta tarea se ha necesitaba exclusivamente una capa, a la cual se ha llamado "Fragmentos".

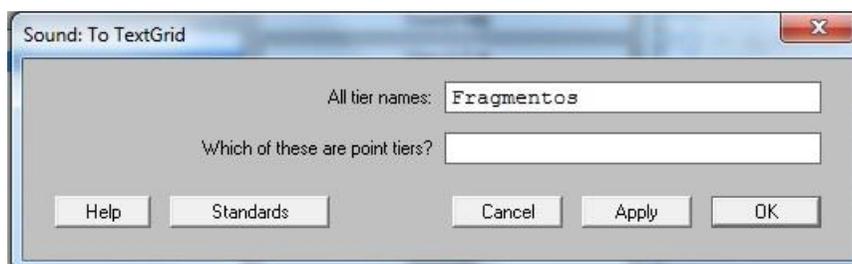


Ilustración 13: Nombre de las capas que se utilizarán en el etiquetado del objeto de sonido

4. En el panel de objetos se crea un nuevo objeto TextGrid [id_ch1], el cual seleccionamos conjuntamente con el objeto Sound [id_ch1]. En el menú de la derecha aparece la opción "View & Edit" que pulsaremos para visualizar a la vez la señal del archivo de audio y la capa sobre la que etiquetaremos.

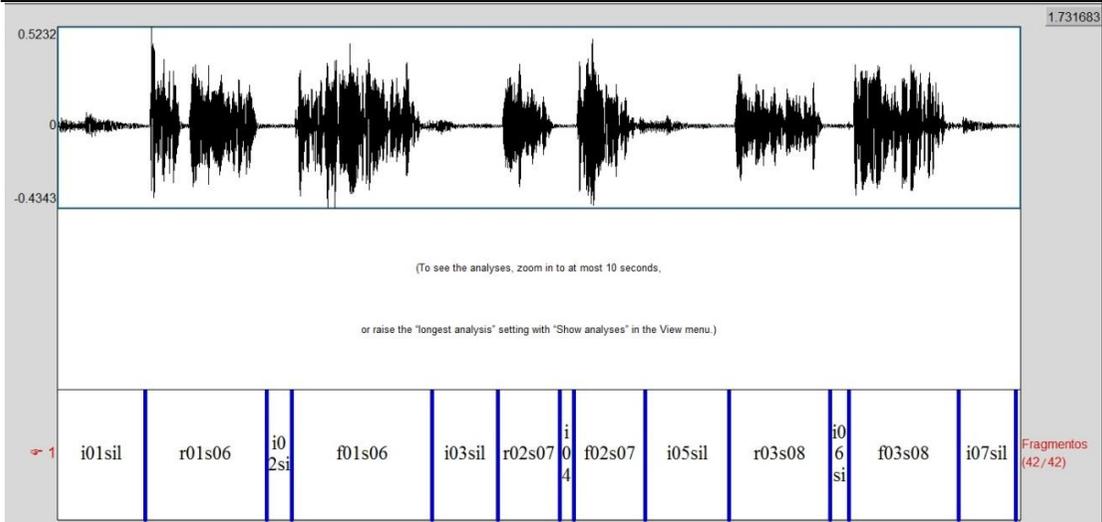


Ilustración 14: Objeto de sonido etiquetado

Sobre la capa "Fragmentos" es donde se realizarán las divisiones correspondientes para delimitar el principio y el final de cada frase de las pistas. Esto se realiza señalando con una "Boundary" el principio y final de las frases. Cada par de "Boundaries" especifican los límites de un "Interval" que contendrá o el sonido de la locución de una frase, o un silencio, o segmentos sonoros sin interés (voz, ruidos...).

Para etiquetar cada intervalo se ha seguido la siguiente codificación:

- i[0-9][0-9][sil][voz]: Indica que el intervalo contiene información no útil, o bien silencio, o bien voces que se han grabado y que no forman parte del ejercicio. También quedan dentro de esta categoría algunas difluencias que por su duración o pausa anterior y posterior podemos eliminar con facilidad (pensando en un futuro proceso de automatización en el cual consideremos como información no útil a intervalos sonoros demasiado cortos precedidos y seguidos de silencios muy prolongados, para lo cual deberemos tomar algunas decisiones como a partir de cuánto tiempo consideramos un intervalo como silencio).
- r[0-9][0-9]s[0-9][0-9]: Indica que el intervalo contiene la locución de un locutor de referencia (r), de una frase cuyo orden local es el del número (dos cifras, de 01 a 10) que le sigue, y cuyo id (en referencia a la tabla del apartado frases) es el correspondiente a los 3 últimos caracteres de la etiqueta (estos intervalos aparecen en las pistas de los ejercicios de audición/ensayo, por tanto, como se ha visto en el apartado protocolo de grabación, serán códigos de s06 a s15).
- f[0-9][0-9]s[0-9][0-9]: Indica que el intervalo contiene la locución de un locutor extranjero (f), de una frase cuyo orden local es el del número (dos cifras, de 01 a 10) que le sigue, y cuyo id (en referencia a la tabla del apartado frases) es el correspondiente a los 3 últimos caracteres de la etiqueta (de s01 a s10 en pistas de ejercicios de lectura de frases y de s06 a s15 en ejercicios de audición/ensayo).

El propósito de esta codificación es que en el proceso posterior de segmentación y extracción de las frases de forma aislada, podamos procesar rápidamente las etiquetas y conocer qué intervalo nos

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

interesa y cuál no, además de etiquetar (ahora sí, con la frase correspondiente identificada a través del id) los archivos de audio resultantes de la extracción de las frases procedentes de las pistas.

5. Una vez se ha etiquetado la pista completa, se selecciona en el panel el objeto TextGrid [id_ch1] y lo guardamos en la carpeta del locutor pertinente en la que tenemos ubicada la grabación de la pista. Para ello, pulsamos en la pestaña "Save" de la barra de herramientas de Praat y elegimos la opción "Save as text file...".

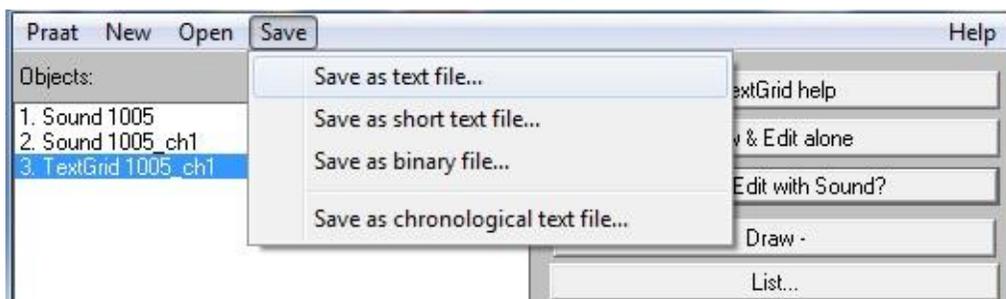


Ilustración 15: Guardar el objeto TextGrid

Finalmente, se determina en la típica ventana de guardado la ubicación en la que deseamos almacenar el fichero de texto con los datos del TextGrid y pasamos a la siguiente tarea.

5.4.2. Procesamiento de las pistas

Después de haber etiquetado las pistas de audio, debemos procesar las mismas para extraer los sonidos y los datos que nos interesan referentes a las locuciones de las frases que están contenidas en ellas.

Si calculamos, tenemos 960 locuciones de las frases ensayadas por 14 locutores extranjeros (60 locuciones de 15 frases por cada locutor, en total 840) y por 8 locutores profesionales nativos españoles (15 locuciones de 15 frases, en total 120). Como es lógico, trataremos de automatizar la realización de tarea, puesto que haciéndolo a mano nos podría costar bastante tiempo y la tarea sería muy tediosa.

Este procesamiento, a su vez, lo dividiremos en varias subtarefas cuyos resultados iremos combinando para poder obtener los archivos de sonido de las frases por separado y sus datos con la simple ejecución de un script.

Las subtarefas de las que nos ocuparemos en esta etapa son:

- Extracción de archivos de sonido de las frases procedentes de las pistas de audio etiquetadas.
- Etiquetado de los nuevos archivos de las frases con la frase correspondiente.
- Extracción de los valores de frecuencia fundamental (F0) y energía (Intensity) de los nuevos archivos de audio.

Al igual que en la fase anterior, nos hemos servidos del software Praat, que nos proporciona las herramientas necesarias para conseguir los ficheros que queremos y un lenguaje de script sencillo que nos facilitará la tarea.

Para crear un script en Praat, debemos pulsar en la opción "New Praat script" de la pestaña Praat de la barra de herramientas.

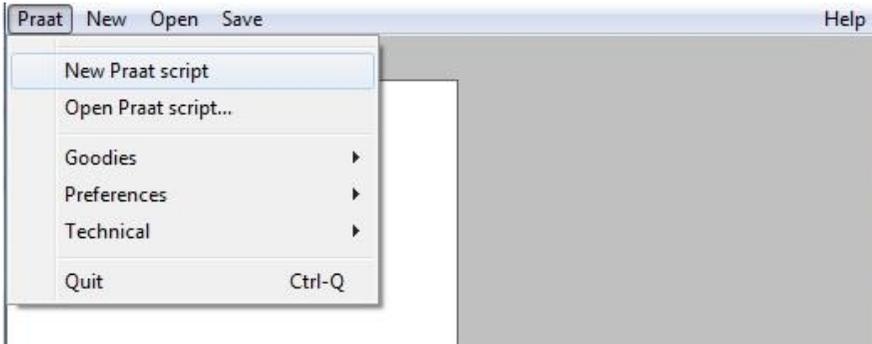


Ilustración 16: Nuevo script

Automáticamente aparecerá en pantalla una nueva ventana en blanco en la que podremos escribir nuestras órdenes. Podremos ejecutar cualquier orden que realice cualquier función de las que podemos lanzar a través de la interfaz gráfica del programa. De hecho, si seleccionamos varias acciones y después nos dirigimos a la ventana de edición de script, podremos ver la sintaxis de las acciones que hemos lanzado pulsando sobre "Paste history" en la pestaña "Edit".

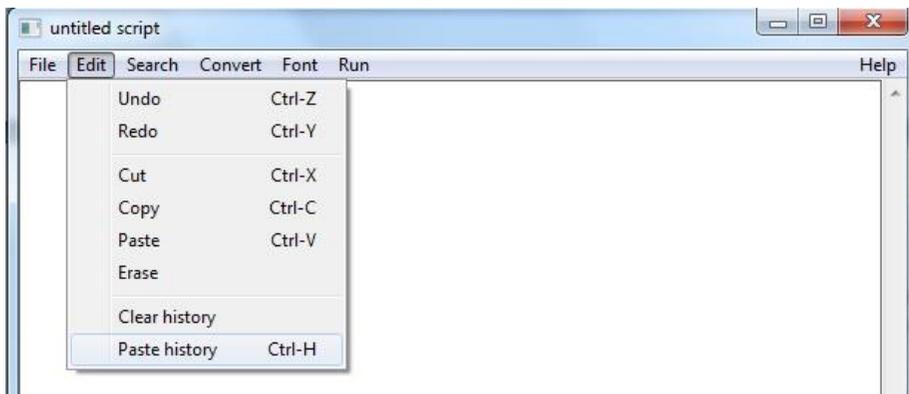


Ilustración 17: Pegar historial de acciones realizadas

Por otra parte, Praat además nos permite combinar estas funciones y automatizar su lanzamiento a través de sentencias de control típicas de los lenguajes de programación (condiciones, bucles...) y métodos ya definidos en su API (conversión de tipos, tratamiento de cadenas...). Podemos encontrar un amplio manual con ejemplos en la página oficial de sus autores, Paul Boersma y David Weenink.

Éstas han sido las herramientas de las que nos hemos servido para desempeñar esta tarea. Pasamos a detallar el procedimiento.

5.4.2.1. Extracción de archivos de sonido de las frases procedentes de las pistas de audio etiquetadas.

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

Como punto de partida de esta nueva fase, disponemos de las pistas de audio con las grabaciones de las locuciones de 10 frases, y de los archivos TextGrid con los datos referentes al etiquetado de esas pistas, es decir, tiempo total de la pista (xmax), nombre de las capas sobre las que se ha etiquetado (name), los tiempos de inicio (xmin) y fin (xmax) de los intervalos (intervals[i]) en que se dividen las pistas, y las etiquetas propias de cada intervalo (text).

Lo que realmente nos interesa es disponer de los datos de las locuciones de cada frase por separado, por tanto, debemos extraer un archivo de audio con el sonido de cada frase de forma aislada.

En primer lugar debemos abrir el archivo de audio de la pista y el archivo .TextGrid que contiene los datos de etiquetado. Como se indicó previamente, podemos hacerlo con la opción "Read from file..." de la pestaña "Open" de la barra de herramientas.

Tal como se hizo en la tarea anterior, antes de iniciar el tratamiento del archivo de sonido se ha extraído el canal izquierdo para poder trabajar mejor.

Con los dos objetos de audio y el TextGrid creados en el panel de objetos, se selecciona el objeto TextGrid para poder acceder a las funciones que nos ofrece Praat en relación con este objeto. En especial nos interesan el número de intervalos de la pista ("Get number of intervals..."), el tiempo de inicio y final de cada intervalo ("Get start/end point..."), y la etiqueta del intervalo ("Get label of interval..."). Podemos acceder a ellas de forma interactiva a través del botón "Query -" del menú de opciones de la derecha del panel de objetos.

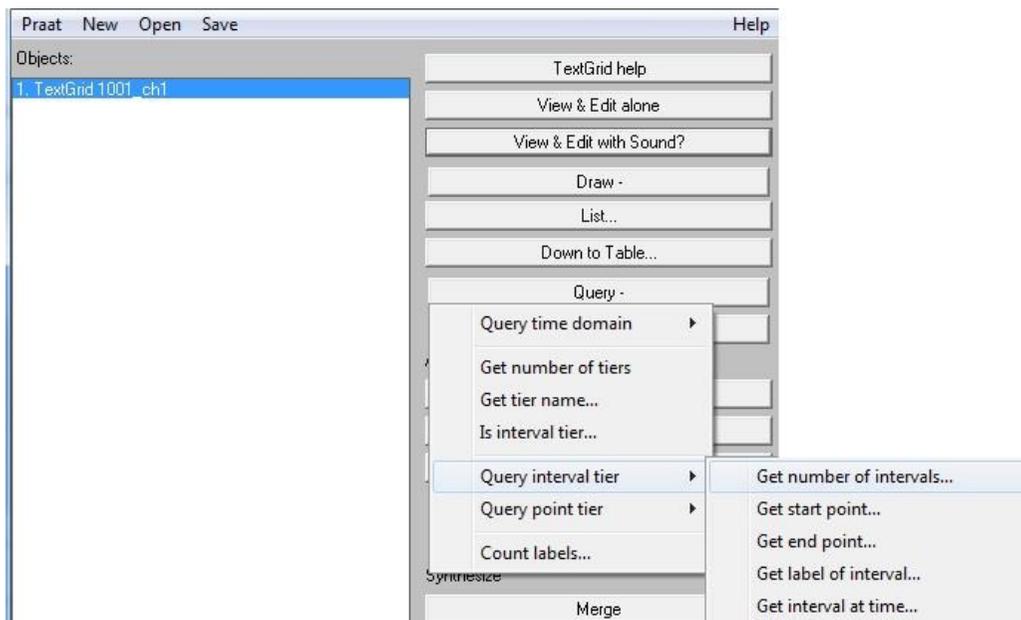


Ilustración 18: Funciones relacionadas con objetos .TextGrid

Lo que se pretende es, por cada intervalo de la pista, obtener el tiempo de principio y final del intervalo, y después de seleccionar el objeto Sound pertinente, utilizar la función "Extract part..." pasando como parámetro esos tiempos para obtener un nuevo objeto Sound en el panel de objetos con el sonido de las frases por separado.

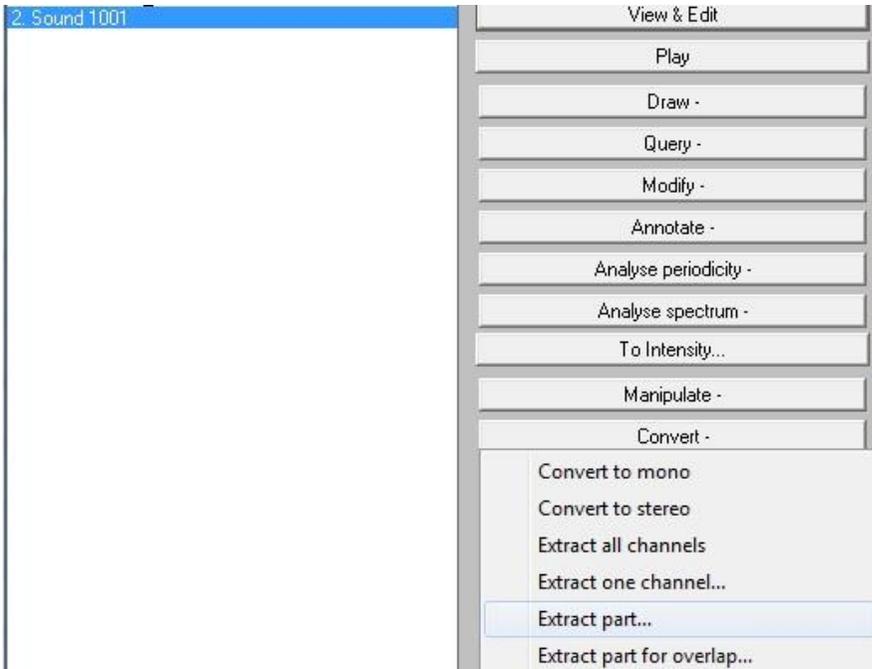


Ilustración 19: Extraer una parte de un objeto de sonido

Este nuevo objeto Sound [id_part] es el que almacenaremos para su posterior estudio, para ello, habiéndolo seleccionado, utilizamos la función "Save as WAV file..." que podemos encontrar en la pestaña Save de la barra de herramientas.

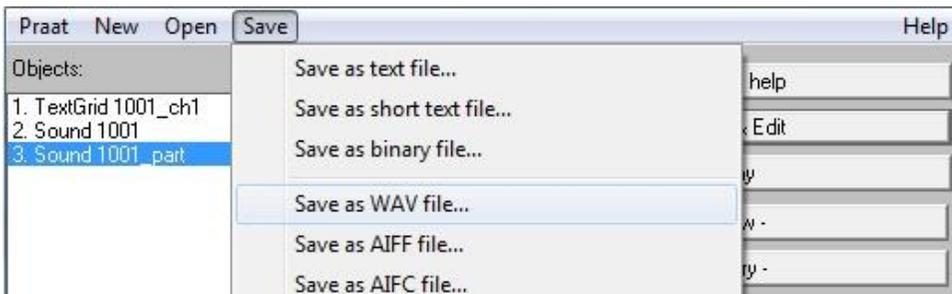


Ilustración 20: Guardar objeto de sonido con formato WAV

Si utilizamos la interfaz gráfica de Praat, se abrirá una ventana típica de guardado de archivos en donde podremos elegir la ubicación del sonido extraído; en caso de invocar la función a través de un script, deberemos pasarla como argumento.

Cuando ya no se va a utilizar más un objeto, conviene "borrarlo" del panel de objetos con el fin de ahorrar recursos de memoria en el sistema. Podemos hacerlo utilizando "do ("Remove")", o bien seleccionando el objeto en el panel y pulsando el botón "Remove" situado en la parte inferior del mismo. Esto no quiere decir que lo borre de nuestro sistema de archivos, tan sólo deja de ocupar espacio en la memoria utilizada por el programa.

5.4.2.2. Etiquetado de los nuevos archivos de las frases con la frase correspondiente.

Después de haber obtenido los archivos de audio de cada frase aislada, se desea poder etiquetarla de modo que se pueda identificar el contenido pronunciado (ensayado) por el locutor correspondiente.

En esta subtarea se hará uso de otra función relativa a los objetos TextGrid, insertar una etiqueta a un intervalo ("Set interval text..."), además de la que ya comentamos en el apartado anterior y que utilizamos para recuperar la etiqueta de un intervalo ("Get interval text...").

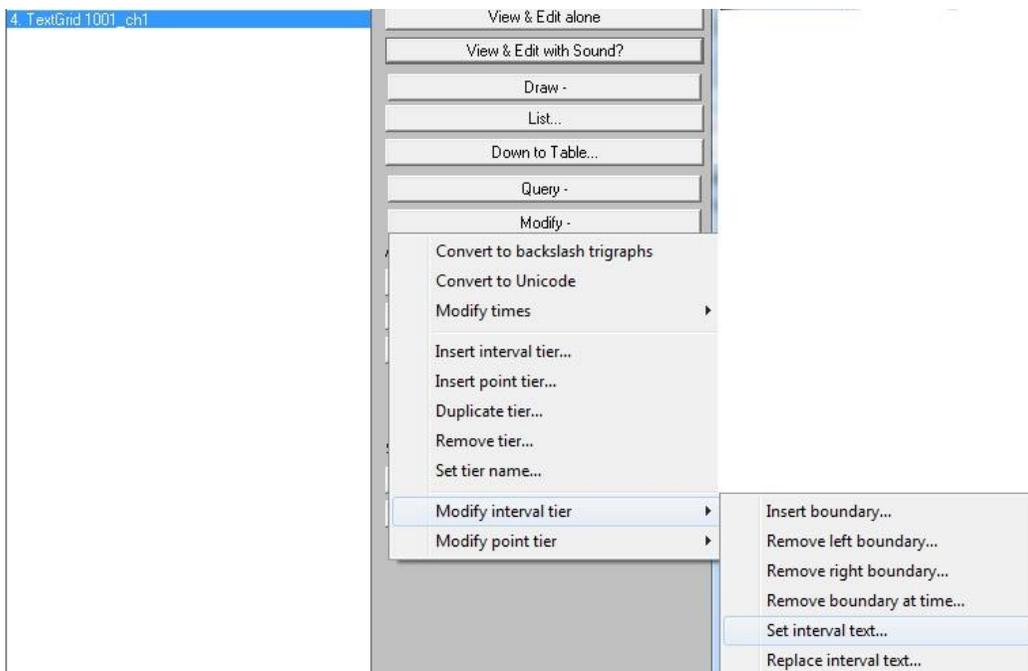


Ilustración 21: Establecer la etiqueta de un intervalo

Lo que se pretende es leer las etiquetas de los intervalos de las pistas y, según el código que tengan anotado (atendiendo al convenio descrito previamente), establecer como etiqueta el texto de la frase correspondiente. Para dar cierta independencia al script del conjunto de frases creado, se ha optado por escribir las frases en un archivo de texto plano ".txt", del cual se leen las frases para su posterior comparación y etiquetado. Esto ofrece la ventaja de que si en algún momento se pretende aumentar el número de frases, tan sólo se deberá copiar en el fichero los nuevos textos y el programa seguirá funcionando correctamente. Para poder leer este archivo se utiliza la función "Read Strings from raw text file..." al que se le pasa como parámetro el path de la ubicación del archivo.

Para identificar las frases en función del código del intervalo se han empleado las funciones de tratamiento de cadenas y conversión de tipos (de cadena a entero y viceversa) que el API de Praat proporciona. Esto ha permitido extraer los prefijos y sufijos que nos interesan para distinguir los locutores, frases y repeticiones, que no sólo utilizaremos para etiquetar la frase oportuna, sino también para almacenar los archivos en la carpeta pertinente dentro del árbol de directorios según lo definido en el apartado "Estructura de Directorios" de la sección

"Contenidos del Corpus" de este documento. En este caso, el script se ciñe completamente a las especificaciones de las grabaciones realizadas, es decir, se contemplan las 15 frases con las correspondientes repeticiones según el caso. Se propone como mejora la parametrización del mismo, de modo que se pueda ampliar en el futuro el número de frases del sistema. Como se expuso señaló anteriormente, la lectura de las frases del fichero de texto ya está implementada y la combinación con este módulo sería muy interesante de cara a la escalabilidad del sistema.

Al igual que en la etapa inicial, al terminar el etiquetado guardamos el objeto TextGrid como archivo de texto ("Save as text file..."), y finalmente eliminamos el objeto del panel.

5.4.2.3. Extracción de los valores de frecuencia fundamental (F0) y energía (Intensity) de los nuevos archivos de audio.

De los archivos de audio de cada frase por separado nos interesa extraer la información referente a la frecuencia fundamental del habla (F0) y a la energía de las locuciones. Praat nos permite acceder a esta información a través de la creación de varios tipos de objetos que utilizaremos para esta subtarea, y que contienen los datos con los que trabajaremos en futuras fases de estudio de este corpus.

Para extraer la frecuencia fundamental (F0), nos serviremos de los objetos Pitch y PitchTier. El primero contiene la información completa de la frecuencia, lo cual incluye la duración completa del archivo de audio (xmax), el número de muestras que se han tomado (nx), el tiempo de la primera muestra (x1), el paso entre las muestras (dx), la frecuencia más alta (ceiling) y los valores de los candidatos (candidate[i]) de la frecuencia fundamental (frequency) en los puntos temporales correspondientes a cada muestra (centro del frame), con una medida de fuerza (strength) y de energía (intensity) que como podremos comprobar en nada se parece a la que obtendremos mediante el objeto Intensity.

Este objeto Pitch se crea invocando a la orden "To Pitch..." que podemos encontrar en el menú desplegable que se muestra al pulsar en la opción "Analyse periodicity -" del menú de comandos de la derecha del panel de objetos.

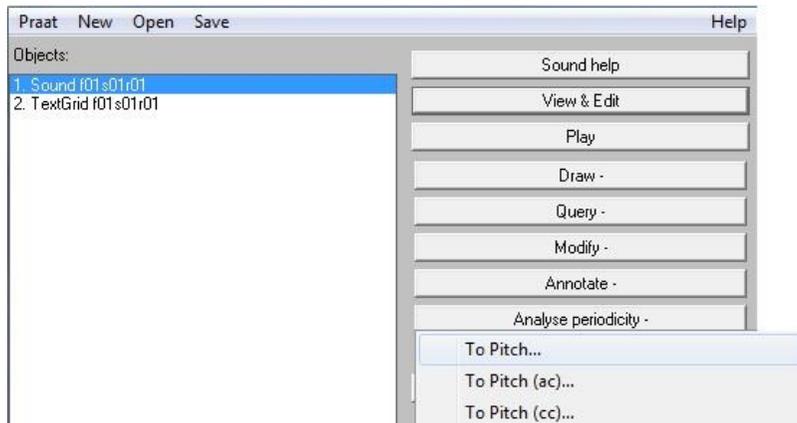


Ilustración 22: Obtención de un objeto Pitch a partir de un objeto Sound

Aparecerá en pantalla una nueva ventana en la que podremos especificar algunos parámetros para la extracción del tono y que necesitaremos conocer, como veremos a

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

continuación, especialmente para sincronizar la información procedente de este objeto con los datos nacidos del objeto Intensity.

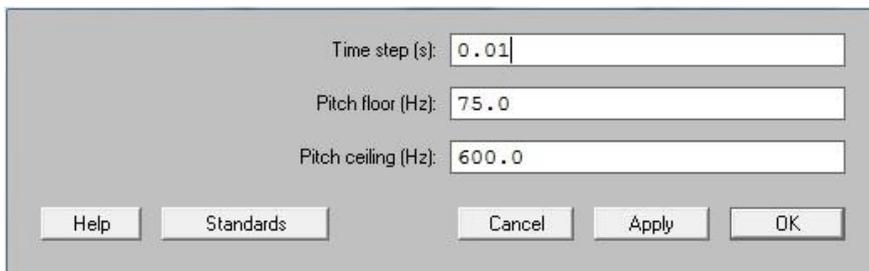


Ilustración 23: Parámetros para la extracción del Pitch

Se ha decidido mantener los valores umbral por defecto (75Hz y 600Hz), por debajo y por encima de los cuales las muestras son catalogadas como "unvoiced". Además se ha fijado el paso entre muestras a 0.01 segundos. Según la documentación, éste es el valor que asigna automáticamente Praat a este parámetro, pero por seguridad, se ha "forzado" su valor. Para guardar los datos de este objeto, nos dirigimos a "Save as text file..." como hicimos anteriormente.

El otro objeto relacionado con el tono es el objeto PitchTier. Éste, contiene la duración del archivo de sonido (xmax), el número de puntos considerados sonoros en él (points:size) y la sucesión de pares (tiempo, frecuencia) ((number, value)) de estos puntos sonoros. Para crearlo es necesario seleccionar en el panel el objeto Pitch [id] y pulsar en la opción "Down to PitchTier".

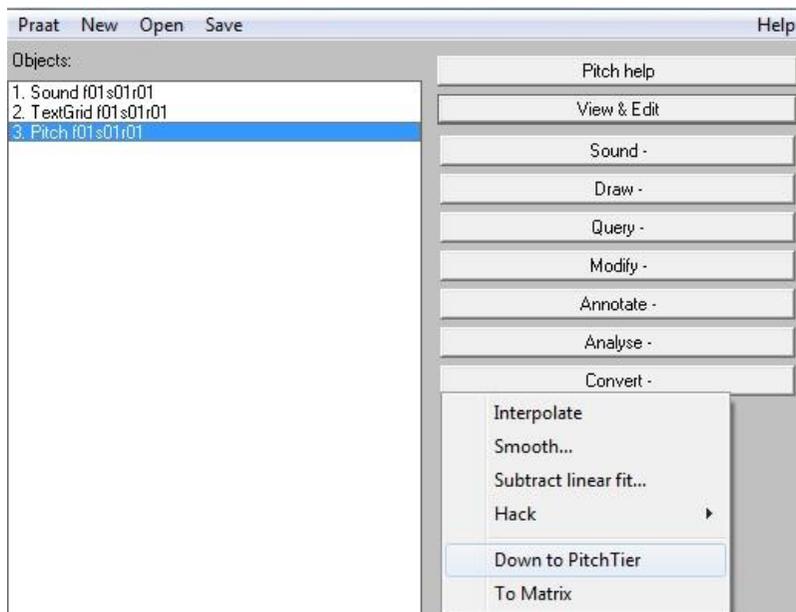


Ilustración 24: Creación del objeto PitchTier a partir de un objeto Pitch

Capítulo 5: Preparación del Corpus

La forma de guardar esta información como un archivo de texto es la misma que la que se ha seguido hasta ahora.

El último tipo de objeto que tiene interés para nuestro estudio es el que recoge los valores referentes a la energía, el objeto Intensity. Este objeto se crea a partir del mismo objeto Sound que hemos cargado y a partir del cual hemos realizado el etiquetado en el TextGrid, y creado los objetos que nos permitirán medir las características relativas a la frecuencia fundamental (Pitch y PitchTier). Creamos este objeto pulsando en la opción "To Intensity -" del menú de la derecha del panel de objetos.

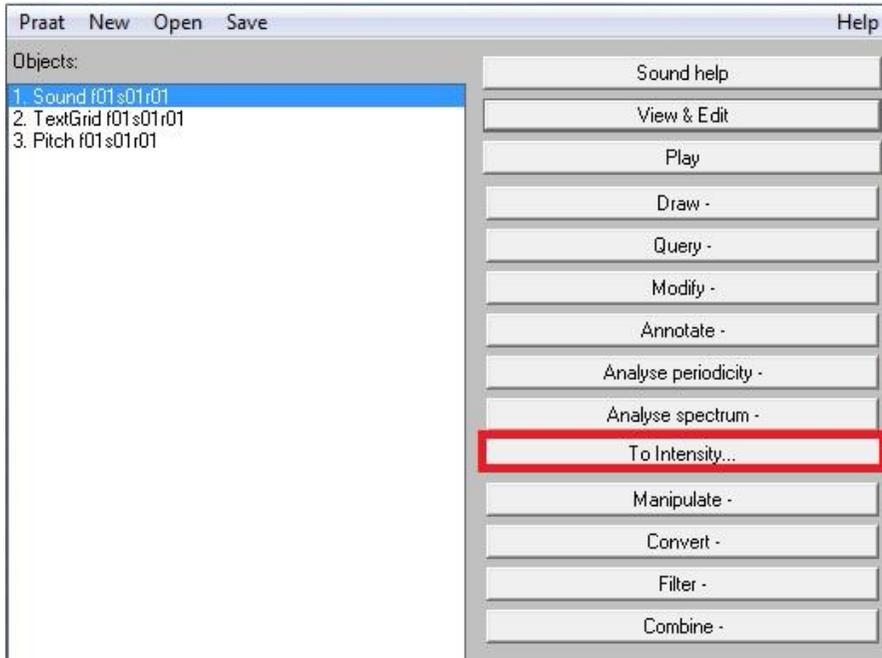


Ilustración 25: Creación de un objeto Intensity a partir de un objeto Sound

La parametrización del objeto Intensity es vital para la sincronización temporal de las muestras de energía recogidas en este objeto y las que obtengamos a través de los objetos creados con el uso de las órdenes "ToPitch..." y "Down to PitchTier".

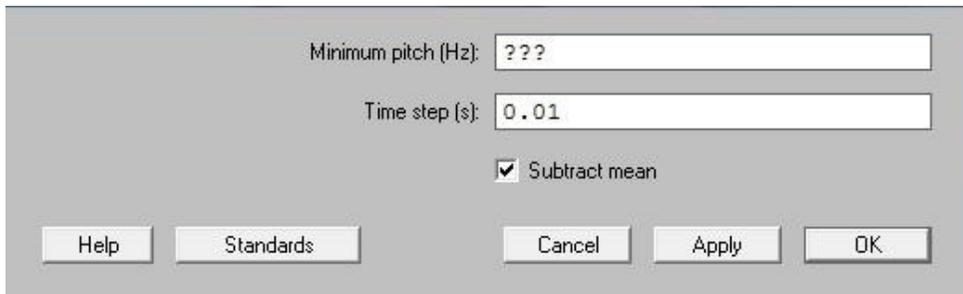


Ilustración 26: Parámetros del objeto Intensity

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

Como es lógico, si queremos sincronizar temporalmente el objeto Intensity, debemos ajustar el tiempo de paso (dx) entre dos muestras al mismo valor al que lo fijamos para el objeto Pitch, en nuestro caso, 0.01 segundos; pero el factor determinante es el parámetro "Minimum pitch (Hz)" y que no podemos asignar aleatoriamente. Según la documentación de Praat, el algoritmo para la obtención de la energía primero eleva al cuadrado los valores del sonido para luego realizar una convolución con una ventana Gaussiana de análisis (Kaiser -20; lóbulos laterales -190dB). La duración efectiva de esta ventana de análisis es de $\pi/\text{minimum_pitch}$, asignando a π un valor de 3.2, y no de 3.14 como solemos acostumbrar. Por tanto, y teniendo en cuenta que queremos sincronizar a partir del primer elemento (x_1) que aparece en el objeto Pitch, sustituiremos los "???" que aparecen en la ilustración 18 (y de forma análoga en la llamada a su función en el script) por el valor equivalente a $3.2/x_1$. Esto nos permitirá calcular los valores de la energía exactamente en los mismos puntos en los que medimos la frecuencia.

Aún así, la aplicación de esta fórmula ha presentado algunos problemas a la hora de automatizar la extracción de los datos de la energía en el mismo cantidad de puntos que los datos de la frecuencia (quizás por errores de redondeo). En ocasiones ha sido necesario reajustar este parámetro para que la información adquirida sea la deseada. En caso de que el número de muestras tomadas para la intensidad fuera mayor que el número para el tono, se ha disminuido ligeramente (en 1) el valor de `minimum_pitch`, hasta conseguir el resultado esperado. En el caso contrario (número de muestras de frecuencia mayor que el de energía), se aumenta de la misma manera (en 1) `minimum_pitch`. En todas las ejecuciones nunca se ha llegado a incrementar o reducir el parámetro `minimum_pitch` en más de una unidad.

Una vez se ha explicado cómo se han planteado cada una de estas subtareas por separado, conviene comentar cómo se ha realizado la automatización final del procesamiento para todas las pistas referentes a los ejercicios de la lectura de frases realizadas por todos los locutores, incluidos los de referencia.

Dentro de la carpeta `C:\call2013\recordings` se encuentran los archivos wav en bruto correspondientes a los locutores extranjeros, cada uno con su propia carpeta, según lo determinado en el apartado "Locutores" de la sección "Contenidos" de este documento. Asimismo, en la carpeta `C:\call2013\frases` están ubicados los archivos wav de las grabaciones de los locutores de referencia, también ordenados en sus carpetas oportunas como se estableció previamente.

En la carpeta `C:\call2013\corpus\scripts` se encuentran el fichero de texto `frases.txt` y tres scripts de Praat: `auto_lanzador.praat`, `auto_TextGrid_todas.praat` y `auto_Pitch_Intensity.praat`. El fichero `frases.txt` contiene en cada línea cada una de las 15 frases del corpus, ordenadas tal como se ha indicado en el apartado "Frases" de la sección "Contenidos del Corpus" del presente informe. El script `auto_lanzador.praat` está compuesto por un conjunto de llamadas al script `auto_TextGrid_todas.praat` mediante la directiva `execute auto_TextGrid_todas.praat [origen] [loc_destino] [fich_audio] [n_rep]`, donde `[origen]` es la carpeta de origen de la pista de audio que se va a procesar; `[loc_destino]` es el código del locutor (conforme a lo especificado en el apartado "Locutores" de la sección "Contenidos del Corpus" de este documento), y por consiguiente el nombre del directorio destino de ese locutor; `[fich_audio]` es el wav que se va a procesar; y `[n_rep]` es un número con rango [01-06] que identifica el número de repetición. Mediante la ejecución íntegra de este script (Run), se desencadena el procesado al completo de todos las pistas de audio recogidas en las líneas de su código. Si lo deseáramos podríamos ejecutar líneas por separado de este script (Run selection), limitando su campo de acción.

El script `auto_textGrid_todas.praat` automatiza la extracción de los nuevos archivos de audio que contienen los sonidos relacionados exclusivamente con cada frase por separado, partiendo de las pistas iniciales y los ficheros `.TextGrid` con el etiquetado de las frases codificado como se expuso en la tarea

anterior. Una vez se extrae el archivo de audio, se guarda en su carpeta correspondiente atendiendo al locutor y repetición pasados como parámetro al script, y al código de frase obtenido de las etiquetas de los intervalos de las pistas de audio en bruto; y posteriormente se ejecuta el script `auto_Pitch_Intensity.praat` con la orden `"execute auto_Pitch_Intensity.praat [ubi] [nom_audio] [id_frase]"`, donde `[ubi]` es la ubicación del archivo de audio con nombre `[nom_audio]`, referente a la frase `[id_frase]`.

El script `auto_Pitch_Intensity.praat`, etiqueta el nuevo archivo de audio creado en `auto_TextGrid_todas.praat` comparando el id de la frase con las frases del fichero `frases.txt` y, teniendo en cuenta las restricciones que se han comentado para la sincronización de los objetos Pitch e Intensity, calcula los parámetros necesarios para la extracción de estas características y del objeto `PitchTier`, obteniendo los datos de estos objetos y guardándolos como ficheros de texto en la carpeta del locutor, frase y repetición adecuados.

De este modo, y tras finalizar la ejecución de estos scripts de forma combinada, tendremos en cada carpeta del árbol de directorios correspondiente a un locutor, frase y repetición concreta un archivo de audio en formato `.wav` con el sonido de la frase, un fichero `.TextGrid` con la etiqueta de la frase, y los ficheros `.Pitch`, `.PitchTier` y `.Intensity` con las características de la frecuencia y la energía.

5.4.3. Generación de los ficheros con formato `.f0p`

Después de haber recogido los datos de cada archivo de audio de cada frase por separado, necesitamos reunir esa información en un único fichero de texto por cada frase, de manera que tenga un formato más "legible" y que posteriormente podamos trabajar con ella de una única lectura (sin tener que abrir y leer varios ficheros a la vez). Con tal propósito, se han diseñado los ficheros `".f0p"`, descritos en el apartado "Ficheros de Datos" de la sección "Contenidos del Corpus" previamente comentada.

Antes de introducir cómo se planteó la creación de estos ficheros, se debe señalar que tras la finalización de las tareas descritas anteriormente, se habilitó un directorio en el servidor remoto `tauro.eca-simm.uva.es` en el que almacenar el trabajo realizado, así como los resultados de las tareas que se desempeñarían posteriormente. Este directorio es `/home/p/pfc/dvall/pcompare`. Dentro de él se copió el árbol de directorios (con raíz en `/corpus`) procedente de la carpeta `c:\call2013\` que antes estaba ubicado de forma local en una de las máquinas del laboratorio 2L018; transfiriendo los ficheros con la ayuda del software `FileZilla`. A partir de este momento, todos los programas y scripts desarrollados para continuar con la preparación de los datos del corpus fueron realizadas a través de un cliente SSH, con el software `Xshell 4`.

Los ficheros `".f0p"` fueron creados mediante la utilización de un programa Python, `leePitchAuto.py`, con ubicación en `/pcompare` y que se ejecuta con la orden `"python leePitchAuto.py"`, encontrándose este programa fuera del directorio `/corpus` (por algún detalle de programación al intentarlo ejecutar desde el directorio `/pcompare/corpus/scripts` no se comporta de la misma forma y produce errores en tiempo de ejecución).

En cuanto a la forma de proceder del programa para generar los ficheros `".f0p"`, la primera intención era basarse en los ficheros `.Pitch` y `.Intensity` con las características de frecuencia y energía extraídas con la ayuda de Praat, siempre comparando con los valores de los candidatos sonoros contenidos en los ficheros `.PitchTier`. Esta aproximación al final se desestimó, pues no se conocía el algoritmo exacto con el que Praat consideraba o descartaba candidatos para la frecuencia fundamental en los puntos temporales en los que recogía una muestra. Se conocían detalles como que si el primer candidato tenía una frecuencia (y fuerza) nula, la muestra se catalogaba como "unvoiced". A pesar de

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

los intentos, no se llegó a conseguir los mismos datos de los ficheros .PitchTier basándonos en los candidatos de los ficheros .Pitch. La forma en la que Praat calcula los valores de los objetos PitchTier viene descrita en el documento "Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-to-Noise Ratio of a Sampled Sound" (Paul BOERSMA, 1993), pero para ahorrar tiempo se optó por un método de trabajo alternativo. Aún así, se recomienda el estudio del documento para una mejor comprensión de la extracción de los valores de F0 en Praat y para el perfeccionamiento de los mecanismos de procesamiento de los archivos de audio en el futuro. Esto nos permitirá conocer con qué criterios se acepta o rechaza a los candidatos como posibles valores pertenecientes a la frecuencia fundamental y mantener la información restante de modo que se puedan considerar otros candidatos que nos permitan estudiar el comportamiento de la señal de una forma más completa y proponer, quizás, otras aproximaciones obtenidas mediante otros métodos.

Puesto que tenemos sincronizadas temporalmente las muestras de los ficheros .Pitch y .Intensity, y puesto que los datos de los ficheros .PitchTier (pares (tiempo, F0) correspondientes a las muestras catalogadas como "voiced") se generan a partir de los comprendidos en los ficheros .Pitch correspondientes, se desestimaron los ficheros .Pitch para la elaboración de los ficheros ".f0p", manteniéndolos en nuestro sistema para un futuro estudio. Así, leemos sólo los datos de los ficheros .PitchTier y .Intensity, utilizando como guía este último, puesto que es el que contiene todas las muestras (tanto las catalogadas como "voiced", como las catalogadas como "unvoiced"). La forma de actuar es la siguiente: disponemos del valor x1 en el que se comienzan a tomar muestras, y también del paso, dx, con el que se avanza en el análisis. Se van leyendo datos del fichero .Intensity y acumulando en cada muestra el paso al tiempo en el que se toma la muestra. Si el nuevo tiempo no coincide con el siguiente valor temporal leído en el fichero .PitchTier, se asigna un valor de 0.0hz a la frecuencia en ese punto, y se cataloga a la muestra como "unvoiced" (un 0). Si por el contrario el nuevo tiempo es igual que el siguiente que aparece en el .PitchTier, se lee el valor de F0 en ese punto y se cataloga la muestra como "voiced" (un 1). Al terminar de procesar el fichero deberíamos tener todas las muestras consideradas en un fichero .Pitch, con los datos de la energía en todos esos puntos, los valores de F0 en las muestras "voiced" que aparecen en el .PitchTier colocadas en su tiempo correspondiente y 0.0 en resto de muestras. Además, aparecerá una columna con valor 1 o 0 según se considere a la muestra "voiced" o "unvoiced", respectivamente.

5.4.4. Obtención de información de los archivos de audio con Snack

Después de haber conseguido y almacenado correctamente los datos generados por Praat procedentes de las grabaciones de los ensayos de las frases del corpus, la intención es repetir la extracción de los mismos tipos de datos sobre el mismo conjunto de grabaciones (los ficheros de las frases aisladas obtenidos con Praat) para tener otros datos con los que poder contrastar la información que hemos adquirido, pero esta vez con otra herramienta diferente, Snack. Existen otros productos con el mismo cometido (SPTK, HTK...), pero se ha optado por la utilización de éste para poder reutilizar módulos ya implementados de los que se disponía con el fin de ahorrar en tiempo y esfuerzo.

Como veremos posteriormente disponemos de un script que invoca a varias funciones de Snack y que se intentó ejecutar desde el cliente SSH en el software Xshell 4 que se descargó en el equipo para trabajar con los archivos residentes en el servidor tauro.eca-simm.uva.es. Desde este tipo de terminal resultó imposible obtener resultado alguno, pues era imposible lanzar Snack. Se indagó en la web en busca de soluciones, entre las cuales se encontró una que sugería que se debía definir una variable de entorno "display". Aun así, los esfuerzos fueron inútiles. El error que se mostraba por pantalla era el siguiente:

```
Traceback (most recent call last):  
  File "getf0.py", line 10, in <module>  
    root = Tk()
```

```
File "/usr/lib/python2.6/lib-tk/Tkinter.py", line 1646, in __init__
    self.tk = _tkinter.create(screenName, baseName, className,
interactive, wantobjects, useTk, sync, use)
_tkinter.TclError: couldn't connect to display "localhost:10.0"
```

Se descubrió que para poder ejecutar Snack era necesario disponer de un cliente con Terminal X, de lo cual carecía el software Xshell 4 con el que se estaba trabajando hasta ese momento. Conociendo esto, se descargó el software Xming, que nos proporciona un Terminal X con el que poder lanzar Snack. Para utilizarlo, se deben seguir los siguientes pasos:

1. Para utilizar este software, debemos dirigirnos a la carpeta Xming del menú de inicio y seleccionar la opción XLaunch.

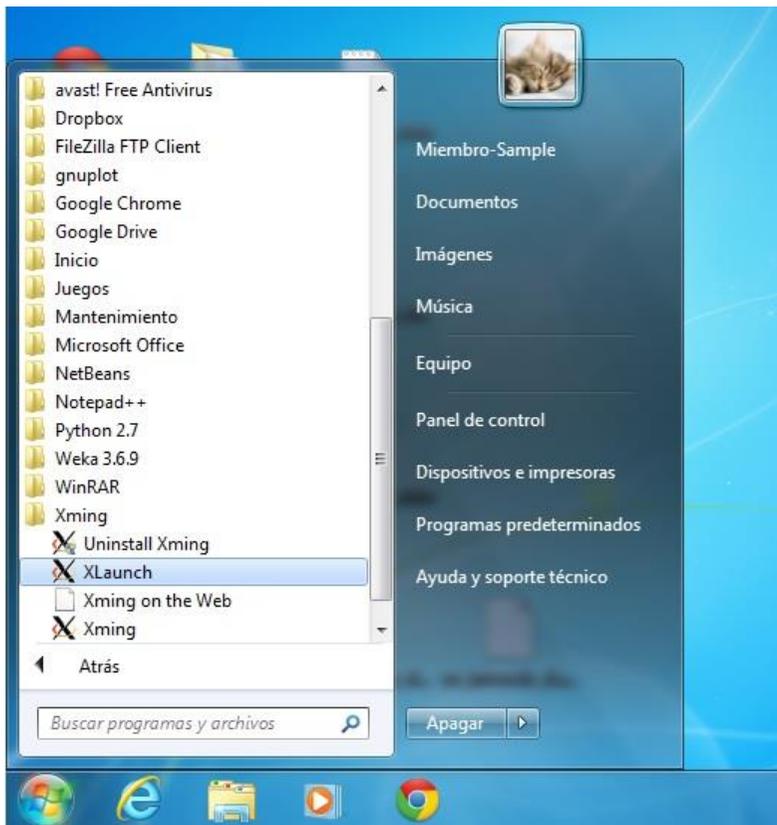


Ilustración 27: XLaunch en Menú de Inicio de Microsoft Windows

2. A continuación se abrirá una ventana con varias opciones entre las que debemos marcar "Multiple windows".

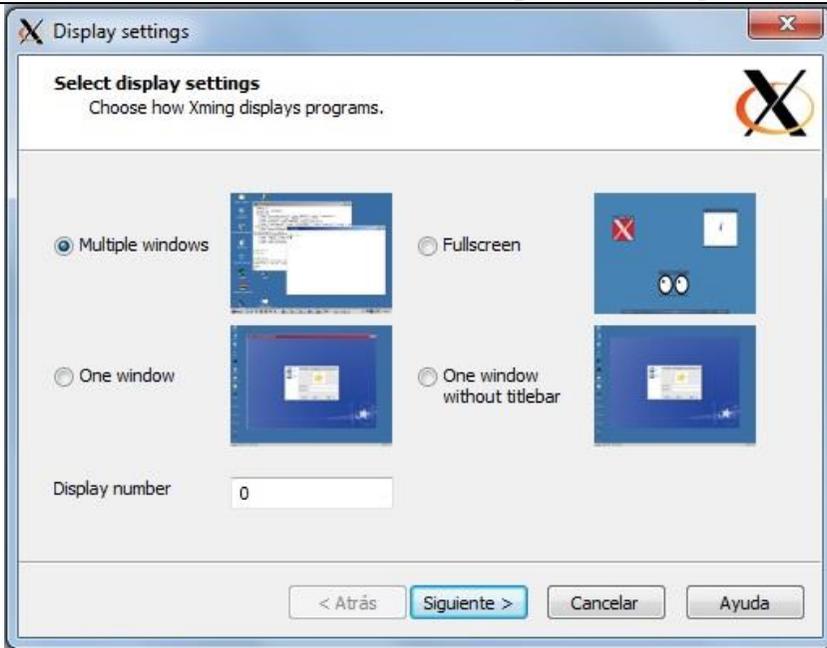


Ilustración 28: XLaunch - configuración del terminal

3. Se nos consultará cómo deseamos iniciar Xming y deberemos elegir la opción "Start a program".

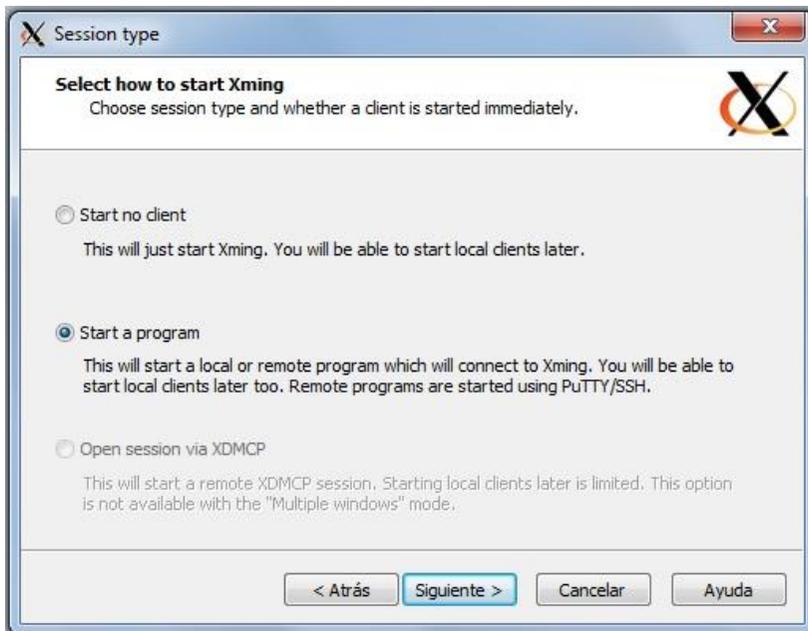


Ilustración 29: XLaunch - configuración de sesión

4. Después, indicamos que se desea la modalidad "Using PuTTY". Introducimos la dirección del servidor (`tauro.eca-simm.uva.es`) y el nombre y contraseña de usuario.

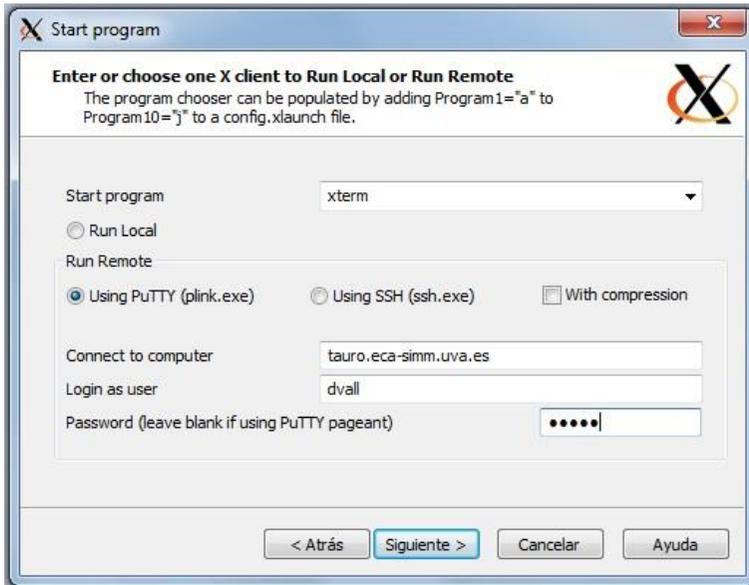


Ilustración 30: XLaunch - establecer login del cliente y servidor

5. El programa nos permite decidir si queremos añadir algún parámetro adicional, no hace falta que rellenemos nada.



Ilustración 31: XLaunch - Parámetros adicionales

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

6. *Antes de finalizar, podemos guardar la configuración, pero también podemos omitir este paso.*

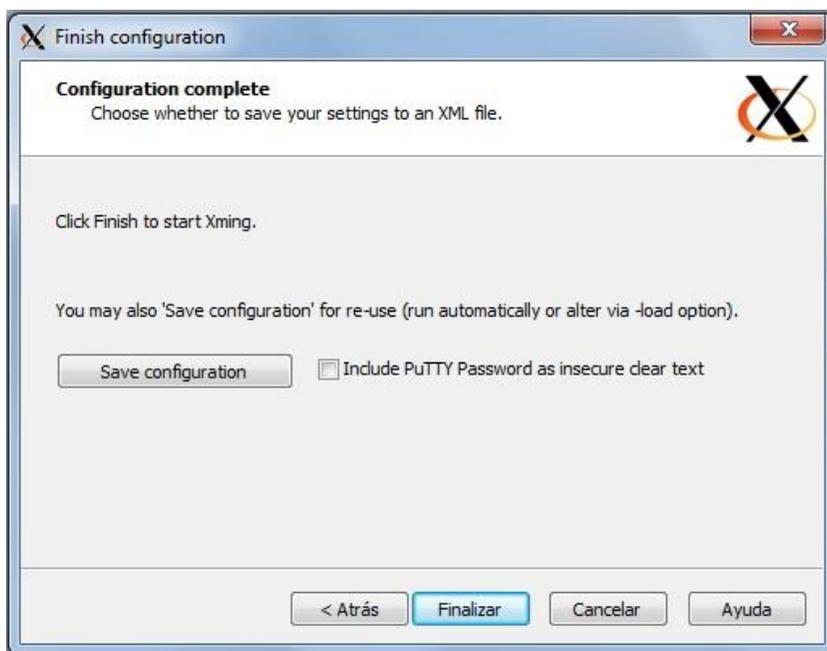


Ilustración 32: XLaunch – finalizar

7. *Tras pulsar "Finalizar", se abre el terminal y ya podemos trabajar con él.*

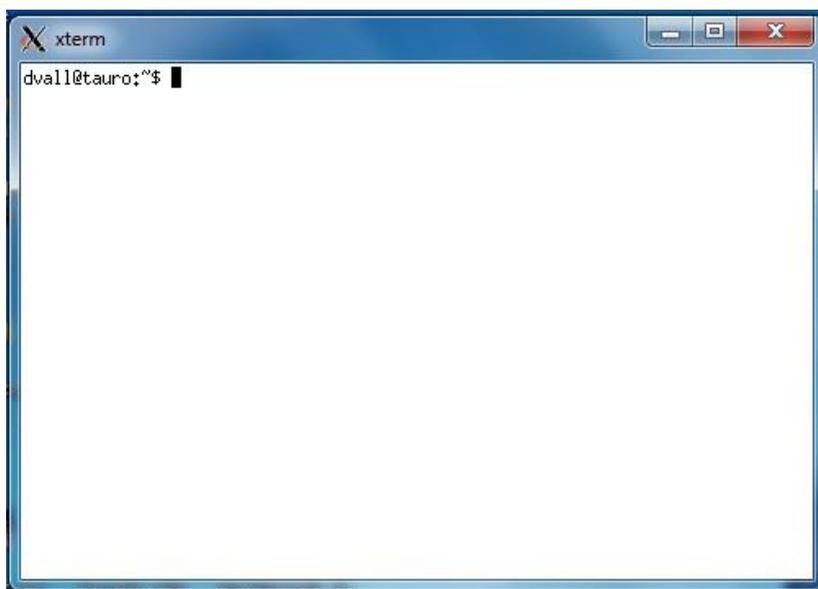


Ilustración 33: XLaunch - terminal

Para poder ejecutar Snack desde un script Python es necesario incluir las siguientes líneas en el código:

```
from Tkinter import *
root = Tk()

import tkSnack
tkSnack.initializeSnack(root)
```

A partir de aquí, podremos utilizar las funciones que necesitemos del API de Snack para operar sobre un sonido que carguemos con el comando `[variable] = tkSnack.Sound(load=[nombre_real])`, donde `[variable]` es el nombre que le damos al objeto de sonido en nuestro programa, y `[nombre_real]`, el nombre del archivo dentro de nuestro árbol de directorios.

Para nuestro propósito, se emplearán los siguientes métodos del API:

- `[variable].pitch(method='esps')`: valores de la frecuencia fundamental del objeto de sonido `[variable]`, y que son almacenados en un vector. Se pasa como parámetro el método de extracción de F0 elegido, en nuestro caso ESPS. Los valores por defecto de este comando son `method='esps'`, `framelength=0.01`, `windowlength=0.0075`, `maxpitch=550` y `minpitch=50`.
- `[variable].length()`: número de muestras del objeto de sonido `[variable]`.
- `[variable].info()`: información de resumen del objeto de sonido `[variable]`. Esta información incluye el número de muestras, la frecuencia de muestreo, máximo, mínimo, la codificación, el número de canales, el formato y el tamaño de la cabecera del objeto de sonido `[variable]`.
- `[variable].sample(ns)`: muestra `ns` del objeto de sonido `[variable]`.

5.4.5. Generación de los ficheros con formato .f0a

Dentro del directorio `/home/p/pfc/dvall/pcompare/corpus/scripts/pitch/snack` se encuentra el script `doGetF0E.sh`. Este script define dos paths, el directorio raíz en el que están ubicados los archivos de audio de las frases (`DPATH=/home/p/pfc/dvall/pcompare/corpus/data/frases/`), y el directorio desde el que ejecutamos los programas para procesar dichos archivos (`BPATH=/home/p/pfc/dvall/pcompare/corpus/scripts/pitch/snack`). El funcionamiento del script es sencillo: recorre todos los directorios-hijo del directorio raíz definido, y por cada archivo con formato ".wav" que encuentre, ejecuta una serie de programas que extraerán sus características mediante la invocación de las funciones pertinentes de Snack; y otros complementarios que darán como resultado unos ficheros con datos similares a los extraídos con Praat.

Los programas que lanza el script `doGetF0E.sh` son los que se detallan a continuación:

- `getf0.py`: desarrollado por César González Ferreras, extrae la frecuencia fundamental (F0) del archivo de audio recibido como parámetro de entrada, utilizando Snack. Se selecciona como método de extracción el algoritmo RAPT (ESPS).
- `getsamples.py`: desarrollado por César González Ferreras, y modificado por Valentín Cardeñoso Payo, extrae la información resumen del archivo de audio que recibe como parámetro de entrada (`orden sonido.info()`) y el conjunto de muestras obtenidas del mismo SIN normalizar (`orden sonido.sample(i)` para cada `i` del conjunto de muestras). Este archivo de audio es el mismo que se pasó como parámetro a `getf0.py`. La información resumen describe

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

los siguientes parámetros: número de muestras, frecuencia de muestreo, máximo, mínimo, codificación, número de canales, formato y tamaño de la cabecera.

- `getRMS.c`: desarrollado por Diego Vallejo Alonso, recibe como parámetro de entrada los datos de salida de `getsamples.py`. Este programa extrae los valores de la información resumen y lee las muestras del archivo de audio, para calcular a partir de ellos los datos referentes a la energía. El procedimiento para el cálculo de estos valores se basa en el realizado en el programa `getRMSEnergy.m` desarrollado por César González Ferreras en lenguaje Octave. Con la frecuencia de muestreo y el paso se computa el ancho de la ventana de análisis, con el que podremos deducir el número de ventanas utilizadas dividiendo el total de muestras por el ancho de ventana. Se identifican el principio y el final de cada intervalo analizado por la ventana de análisis (restando y sumando la mitad del ancho de la ventana al punto central), y se obtiene la media de las energías de las muestras del intervalo, asignando este valor al punto central del intervalo. El resultado es un conjunto de pares (tiempo, energía) que se almacenan en un fichero con formato ".ene" (sería el equivalente a los .Intensity en Praat). Para que se pueda ejecutar debe compilarse previamente e incluir el nombre del ejecutable elegido en el script `doGetFOE.sh`. Por simplicidad, se ha utilizado el mismo nombre que el archivo del programa, pero sin su extensión, es decir, `getRMS`.
- `fusionFOEne.py`: desarrollado por César González Ferreras, reúne los datos referentes a la frecuencia fundamental y a la energía del archivo de audio que sirvió de entrada desde la llamada a `getf0.py`, ordenados temporalmente y caracterizando al intervalo como "voiced" o "unvoiced" mediante un parámetro estadístico cuyo valor oscila entre 0.0 y 1.0. Cuanto más cercano se encuentre de 0.0, podremos clasificar a la muestra como "unvoiced", y cuando más próximo a 1.0 la catalogaremos como "voiced". Además, se incluyen dos columnas de coeficientes de correlación cruzada que de momento no son utilizados (todas las filas los tienen predeterminados a 0.0). La salida de este programa será redirigida a un fichero con extensión ".f0a", equivalente al ".f0p" obtenido de los parámetros de Praat.

5.4.6. Generación de información elaborada a partir de los datos de los ficheros .f0p y .f0a (ficheros .ppar y .spar)

Con la información reunida de las locuciones de cada frase generada por Praat y Snack correctamente organizada en ficheros de texto dentro del árbol de directorios, el siguiente paso es elaborar un poco más estos datos para poder estudiar los parámetros que caracterizan típicamente la señal. Además de la documentación analizada antes del inicio de la preparación del corpus, tomamos también como referencia el script `ProsodyPro`, que se comentará posteriormente; lo cual nos permitió seleccionar una serie de medidas que se incluirán en un vector de características que utilizaremos para cuantificar las distancias entre locuciones en las siguientes etapas. Estas medidas no son sólo las que nos proporcionan directamente los ficheros de texto que hemos generado hasta el momento, sino que también se incorporan otras, con carácter más estadístico, calculadas de forma indirecta mediante la operación de las primeras. Podemos destacar las siguientes:

- Tiempos de máximos y mínimos relativos y absolutos normalizados: los máximos y mínimos son valores que podemos lograr directamente simplemente comparando con el resto de muestras del intervalo (relativos) y con el resto de muestras del archivo completo (absolutos). En este caso además recogemos el tiempo que se tarda en alcanzar el máximo/mínimo dentro del intervalo y del fichero entero. El problema que nos encontramos con estas medidas, es que

al tratarse de una comparación de locuciones, el tiempo en sí no nos ofrece información suficiente, sino que nos interesa más una perspectiva que nos indique en qué posición con respecto al conjunto (ya sea el intervalo o el archivo completo), o a qué altura se alcanzan esos puntos máximos y mínimos de frecuencia y de energía. Es por ello que normalizamos los tiempos en los que se localizan dividiendo por el tiempo total del intervalo o del fichero entero, según corresponda. Los tiempos normalizados tendrán un valor real en $[0,1]$.

- Rango de frecuencias y energías: el rango (o recorrido) es una medida de dispersión estadística que equivale a la diferencia entre el valor máximo y mínimo del parámetro que estemos midiendo. Los rangos de frecuencias e intensidades nos podrían ayudar a descubrir si es un hombre o una mujer, o si es nativo o no nativo, por ejemplo. El punto negativo de esta medida es que valores espurios muy altos (o muy pequeños) pueden hacer que sea inútil.
- Medias de frecuencias e energías: al igual que ocurre con el rango, la media (aritmética en nuestro caso), es una medida muy sensible a los valores extremos. Sin embargo, si los datos de las muestras se comportan correctamente, puede ser de utilidad para identificar el "centro de gravedad" entre los valores recogidos, de modo que podamos comparar con otras medias de otros locutores para la misma frase, con otras frases del mismo locutor...
- Desviación típica de frecuencias e energías: la desviación típica es una medida de dispersión que nos da otra perspectiva en contraste con otras medidas de tendencia central, nos informa de la variabilidad que podemos esperar con respecto a la media. Esto nos puede ayudar a distinguir locutores que hablen de forma más "robótica".

El resultado de esta tarea son los ficheros de texto ".ppar" y ".spar", cuyo contenido se describe en el apartado "Ficheros de Datos" de la sección "Contenidos del Corpus" del presente documento.

Para la generación de estos ficheros, se ha utilizado un script al estilo del que se utilizó para conseguir los ficheros de datos ".f0a". La idea es recorrer todo el árbol de directorios, y por cada fichero wav que se encuentre, debería hacer en la misma carpeta un fichero ".f0a" y otro ".f0p", los cuales, por otra parte, se leen de forma distinta (los ficheros ".f0a" tienen 6 parámetros por línea, mientras que los ficheros ".f0p" tienen sólo 4). Por ello, en el script `doGetResumen.sh` ubicado en el directorio `/home/p/pfc/dvall/pcompare/corpus/scripts/pitch/snack`, además de definir los dos mismos paths sobre los que se recorren los directorios que en el script `doGetFOE.sh` (`DPATH=/home/p/pfc/dvall/pcompare/corpus/data/frases/` y `BPATH=/home/p/pfc/dvall/pcompare/corpus/scripts/pitch/snack`), se ejecutan dos programas implementados en lenguaje C que leen, cada uno, uno los datos del fichero ".f0p" (`getResumenP.c`), y el otro los del fichero ".f0a" (`getResumenS.c`) correspondientes al wav de turno. Al igual que sucedía con `getRMS.c`, para poder ejecutar estos dos programas al lanzar el script `doGetResumen.sh` se deben compilar previamente. La diferencia entre ambos es que `getResumenS.c` consume por cada línea, dos lecturas más (de los dos coeficientes de correlación que tenemos inicializados a 0.0 por el momento) por cada línea de un fichero ".f0a".

La manera de proceder es la misma: se abre el fichero de texto y se realiza la lectura del mismo, almacenando los datos en 4 vectores diferentes del mismo tamaño, uno para los tiempos, el segundo para las frecuencias, un tercero para las energías y el último para marcar si las muestras son sonoras o no. Los vectores tienen una asignación dinámica de espacio, pues debido a la extensión variable de los ficheros, no es de extrañar que sea necesario aumentar su tamaño varias veces (no perdemos datos en dicho aumento, puesto que nos ayudamos de otros vectores auxiliares que nos ayudan a mantener los datos). Después de la lectura, se opera con los valores, distinguiendo intervalos sonoros y sordos. Se han definido dos tipos de estructuras, una para recoger los parámetros relativos a los intervalos, y otra

para guardar los datos referentes al global del fichero. Una vez se han procesado al completo, se imprimen a un archivo ".ppar" o ".spar", según corresponda.

La elección de estas extensiones fue con motivo de la existencia de otras similares que son utilizadas por otros programas. La idea era expresar "praat-parámetros" o "snack-parámetros", con lo que la primera intención fue denominarlos con formato ".ppa" y ".spa", pero como se ha apuntado, ".ppa" es uno de los formatos que reconoce Microsoft PowerPoint. Así, se optó finalmente por usar ".ppar" y ".spar".

5.4.7. Cálculo de distancias con el algoritmo DTW

En este punto ya disponemos de toda la información de las locuciones de cada frase por parte de todos los locutores distribuida de forma organizada en los directorios pertinentes en /home/p/pfc/dvall/pcompare/corpus/data/frases/, y además tenemos los ficheros resumen ".ppar" y ".spar" que contienen los parámetros directos e indirectos que hemos obtenido de las estimaciones de Praat y snack, respectivamente.

A partir de aquí, lo que nos interesa son los datos contenidos en estos dos ficheros por cada frase, es decir, los 960 ficheros ".ppar" y los otros tantos ".spar". Tal como se especifica en el apartado "Ficheros de datos" de la sección "Contenidos" de este documento, dejando a un lado (de momento) los datos de las cabeceras, tenemos un total de 24 parámetros por cada intervalo. Estos parámetros constituirán nuestro vector de características, y disponemos uno por cada muestra de cada intervalo de cada frase ensayada por cada locutor. Estos vectores son los que utilizaremos para comparar todas las locuciones entre sí, calculando (o intentándolo) las distancias entre ellas.

El algoritmo que hemos escogido para computar estas distancias es DTW (Dynamic Time Warping). Podríamos haber pensado en una comparación punto a punto, pero en nuestro dominio, resulta lógico pensar que no nos servirá. La simple repetición de una misma frase por una misma persona en comparación con la primera que pronunció resultaría una distancia muy grande, puesto que por mucho que nos esforcemos, una persona no pronuncia dos veces nada de la misma forma. Si en lugar de pensar en esta distancia intra-persona nos fijáramos en las diferencias inter-persona, las magnitudes serían aún mayores. De un modo general, los segmentos sonoros que producimos al hablar tienen unas características (duraciones, intensidades, velocidades, entonaciones...) diferentes aunque intentemos decir lo mismo en dos o sucesivas ocasiones. Es por esto que necesitamos un algoritmo de alineamiento que nos ayude a relacionar los segmentos similares de dos locuciones, aunque estos se produzcan en instantes temporales diferentes, es decir, a encontrar la mínima distancia (o el camino más corto) entre ellos. Esto es lo que pretendemos del algoritmo DTW.

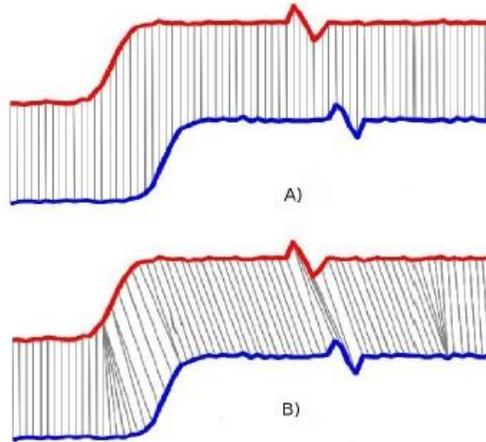


Ilustración 34: Distancia Euclídea punto a punto vs. DTW (Fuente: "Bases de Datos de Series Temporales: Representación y Consultas", Antonio Moreno García)

5.4.7.1. Algoritmo DTW

Como se ha sugerido, el planteamiento del algoritmo DTW surge de los problemas existentes en la comparación de dos realizaciones distintas de una misma locución, en las cuales podemos apreciar la carencia de sincronización. Esta variabilidad interna se produce de forma heterogénea, es decir, no se comporta siguiendo una ley fija, sino que puede producirse en diferentes momentos con diferentes duraciones, alargando o acortando el segmento que se analiza de forma impredecible. Por tanto, dadas estas condiciones, podemos estar comparando dos segmentos que no se refieran a la misma información. Es por ello que necesitamos alinear las muestras para poder computar la distancia entre las señales de las locuciones de una forma algo más "fiable".

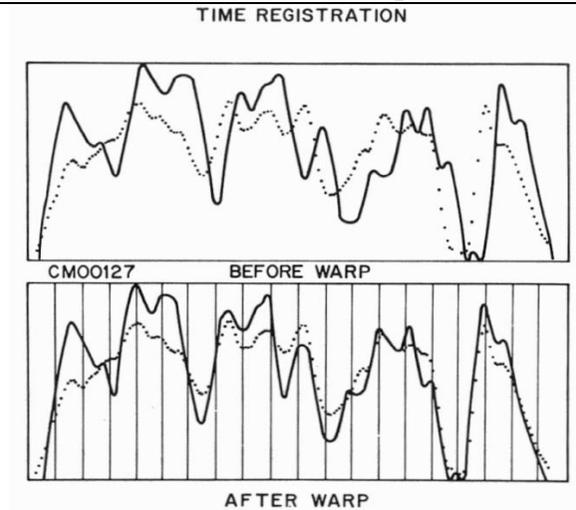


Ilustración 35: Alineamiento temporal de dos señales correspondientes a dos locuciones (Fuente: Universidad Autónoma de Madrid)

Durante el proceso de alineamiento temporal, el eje temporal (X), se comprime y se expande de manera no lineal para alinear los vectores de características entre la señal de la locución modelo y la de test. Como resultado, obtenemos el camino o función de alineamiento.

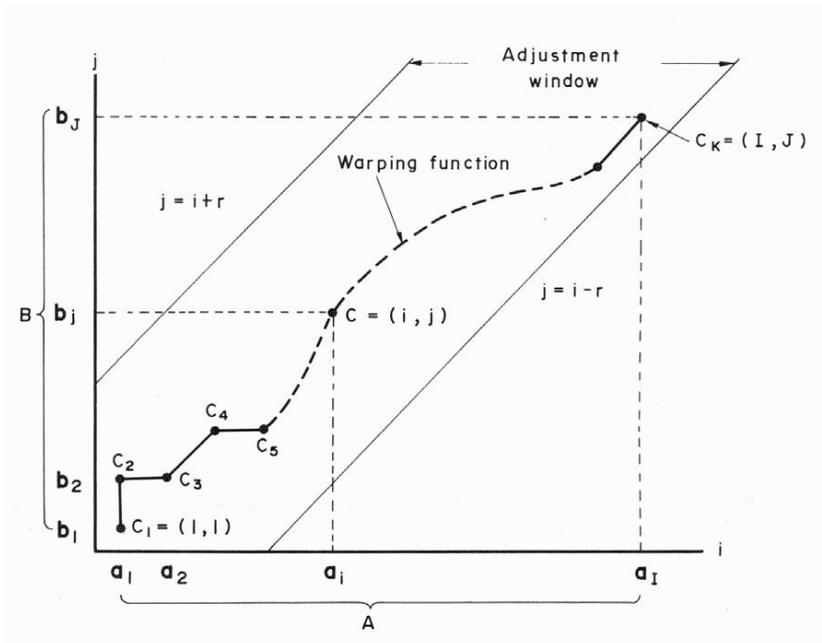


Ilustración 36: Función de alineamiento (Fuente: Universidad Autónoma de Madrid)

Un ejemplo de cómo sería su efecto sobre dos locuciones desalineadas sería el siguiente:

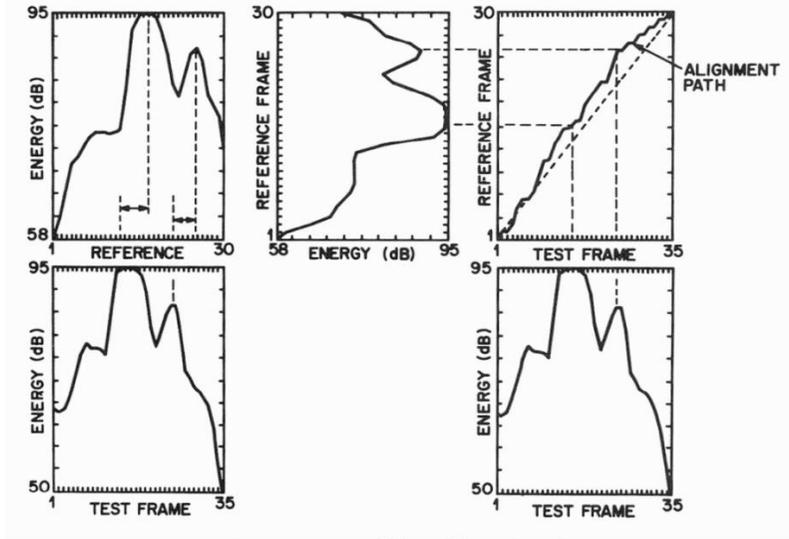


Ilustración 37: Efecto del alineamiento sobre dos señales sin alinear (Fuente: Universidad Autónoma de Madrid)

Si analizamos lo que hace, nos damos cuenta de que la función de alineamiento compara los vectores de características de las señales de las locuciones que pasan por cada punto de su trayectoria.

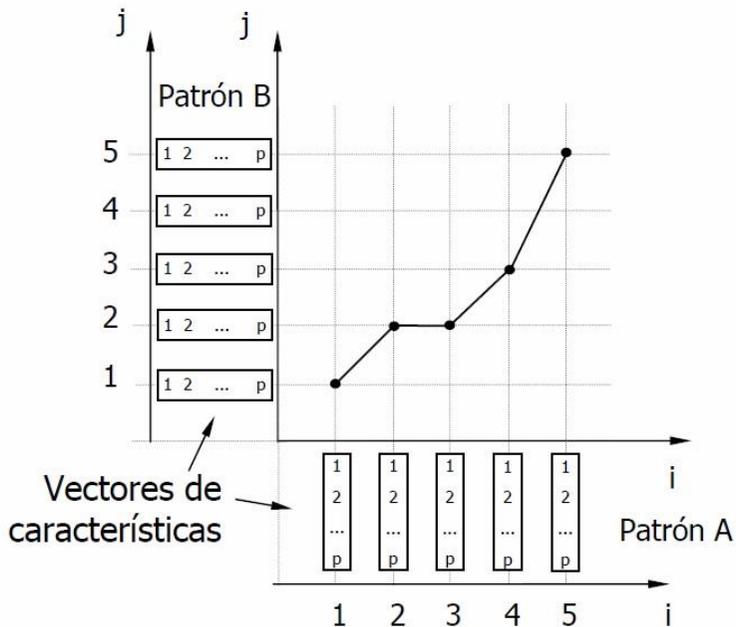


Ilustración 38: Comparación de vectores en los puntos de la trayectoria de la función de alineamiento (Fuente: Universidad Autónoma de Madrid)

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

Como se puede observar en la Ilustración 29, si la función de alineamiento pasa por (i,j) , entonces el vector de características del punto "i" de la muestra A, se compara con el mismo vector de características del punto "j" de la muestra B.

El objetivo es, como se viene adelantando, encontrar la función de alineamiento a partir de dos vectores correspondientes a las muestras (cada muestra con su vector de características, en nuestro caso los 24 parámetros de cada intervalo de un fichero ".ppar" o ".spar") de dos locuciones, $A = \{a_1, a_2, \dots, a_i, \dots, a_M\}$ y $B = \{b_1, b_2, \dots, b_j, b_N\}$.

- Llamaremos $C = \{c(1), c(2), \dots, c(k), \dots, c(K)\}$, a la función de alineamiento, donde $c(k) = [i(k), j(k)]$, esto es, un par de punteros a las muestras que se comparan entre sí.
- Para cada $c(k)$ tenemos una función de distancia:

$$d\{c(k)\} = \delta(a_{i(k)}, b_{j(k)})$$

- Una función de distancia típica (que es la que utilizaremos a priori), es la distancia euclídea:

$$d\{c(k)\} = (a_{i(k)} - b_{j(k)})^2$$

- La función de alineamiento será aquella que minimice la función de peso (distancia) final:

$$D(C) = \sqrt{\sum_{k=1}^K d\{c(k)\}}$$

- Además, la función de alineamiento debe satisfacer las siguientes condiciones:
 - Monotonicidad: El camino debe tener sentido de izquierda a derecha, y de abajo a arriba, viéndolo en forma de matriz. Es decir, se debe cumplir que $i_k \leq i_{k+1}$ y que $j_k \leq j_{k+1}$.
 - Continuidad: El paso de un nodo al siguiente debe hacerse sin saltar nodos intermedios: $i_k - i_{k-1} \leq 1$ y $j_k - j_{k-1} \leq 1$.
 - Frontera: $i_1 = j_1$, $i_N = n_x - 1$ y $j_N = n_y - 1$. Se establece que el principio y el final de los vectores deben coincidir con estos puntos independientes del camino seguido.
 - Ventana: $|i_k - j_k| = \Delta$, donde Δ es un entero positivo.

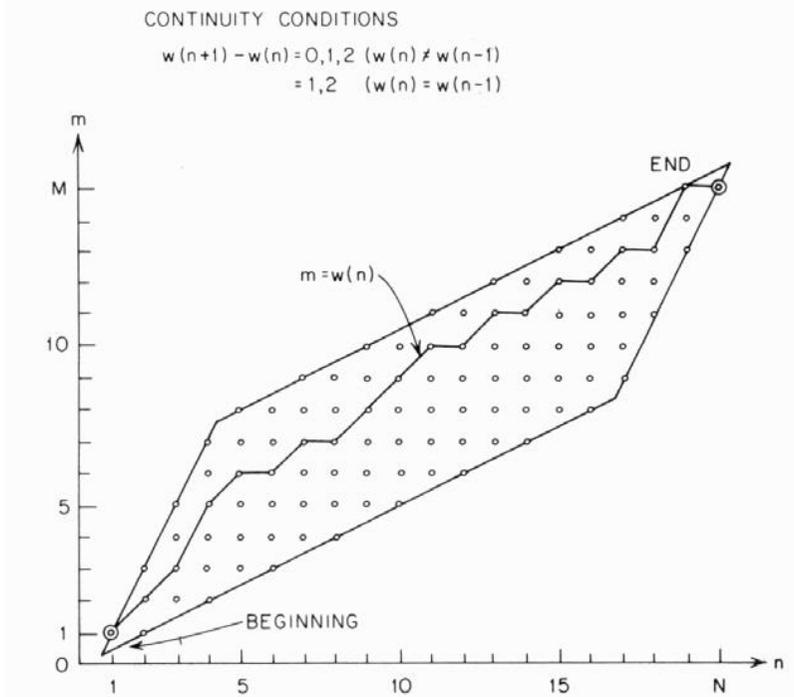


Ilustración 39: Algoritmo DTW - Restricción de Ventana (Fuente: Universidad Autónoma de Madrid)

- Pendiente: Se aplican restricciones sobre las pendientes del camino para evitar el avance hacia puntos donde el ajuste no sería bueno.

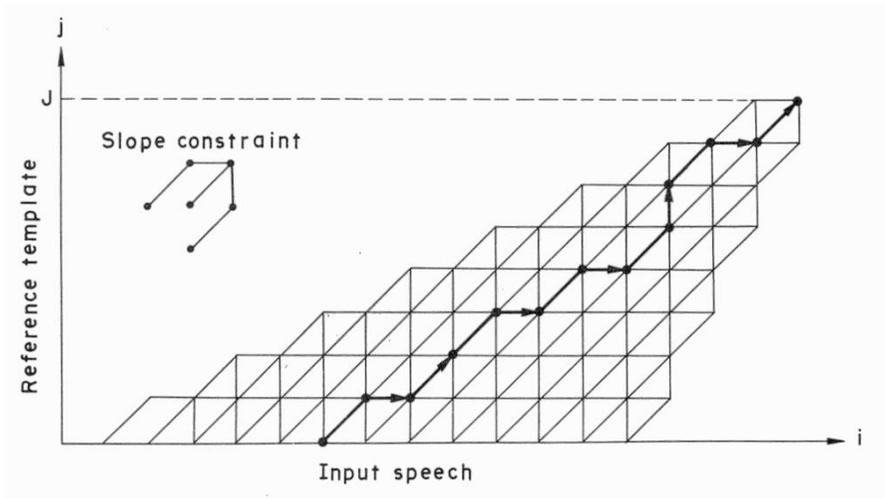


Ilustración 40: Algoritmo DTW - Restricción de Pendiente (Fuente: Universidad Autónoma de Madrid)

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

Algunas propuestas de restricción de pendiente son las que siguen. Nosotros utilizaremos de momento el primer caso.

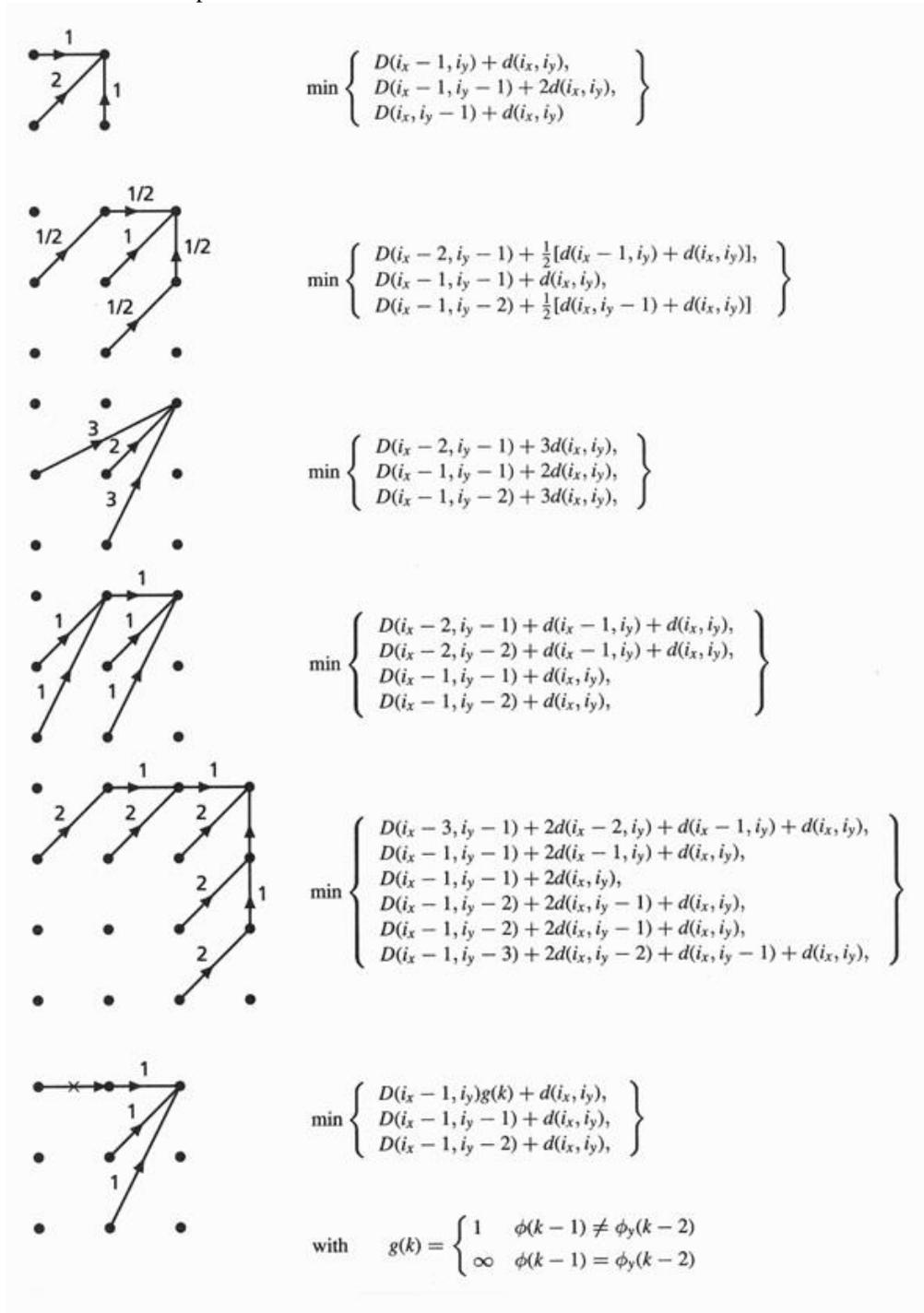


Ilustración 41: Algoritmo DTW - Tipos de Restricción de Pendiente (Fuente: Universidad Autónoma de Madrid)

Para resolver el problema de la alineación temporal podríamos pensar en calcular todos los pesos posibles (técnica de fuerza bruta) para posteriormente elegir el mínimo, pero el cómputo sería demasiado pesado. Es por eso que se suele resolver con técnicas de programación dinámica.

Según el método de programación dinámica, el peso entre el punto (1,1) y otro punto (i,j) es independiente de lo que ocurra después de (i,j), esto es, el peso de cualquier punto [i(k),j(k)] es el peso mínimo que se acumula hasta llegar a (i,j), más el propio peso de (i,j):

- $D(C_k) = d\{c(k)\} + \min_{\text{legal-}c(k-1)} \{D(C_{k-1})\}$, donde 'legal-c(k-1)' representa a los predecesores legales de C(k).
- Podemos describir el algoritmo de la siguiente forma:
 1. Inicialización:
 $D(C_1) = d\{c(1)\}$
 2. Recorremos el índice 'i':
para cada i desde i=1 hasta i=M
para cada j permitido
 Calcular la distancia acumulada según el método de restricción escogido y la posición del predecesor
fin para j
fin para i
 3. Final:
 $d(A,B) = D(C_K) / K$
 $c(K) = [i(M),j(N)]$
- Recursos – 2 arrays:
 - Uno de tamaño MxN para almacenar predecesores de cada punto.
 - Uno de tamaño 2xN para guardar los costes acumulados de la columna anterior y la presente, actualizados en cada paso de 'i'.
- Para obtener el camino de alineamiento, recorreremos hacia atrás ("backtracking") desde el punto final (M, N) hasta el inicial (1,1).

Después de esta introducción al algoritmo DTW, nos centraremos en cómo lo hemos utilizado para calcular las distancias entre las distintas locuciones de las diferentes frases. Tras una etapa de documentación, se encontraron en la web diferentes códigos que implementaban el algoritmo, tanto en lenguaje Java, como en C. Finalmente se optó por emplear la solución publicada por Interval Research Corporation, desarrollada en C. En el directorio /home/p/pfc/dvall/pcompare/corpus/scripts/pitch/, se ha adaptado esta rutina en el fichero dtw.c, la cual se sirve de las constantes y tipos de datos declarados en la biblioteca dtw.h, creada para los programas que pudieran hacer uso de ello.

En nuestro caso, en el mismo directorio se ha creado un programa dist.c que realiza una iteración por cada locutor, frase y repetición, calculando todas las distancias y escribiéndolas en un fichero de texto, imprimiendo cada dato con formato "DTW[utt1][utt2]=dist", donde [utt1] representa el fichero de datos (".ppar" o ".spar", según corresponda, sólo comparamos ficheros del mismo formato, esto es, nunca compararemos la distancia entre los datos de un fichero ".ppar" con los de otro ".spar") una locución realizada por un locutor de una frase (sea la repetición que sea) y [utt2] representa otra locución del mismo y distinto locutor sobre la misma u otra frase diferente (sea la repetición que sea); dist sería el dato numérico (un número real) que cuantifica la distancia existente entre ambas

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

locuciones. Por lógica, la comparación entre la locución de una frase consigo misma debería resultar una distancia igual a 0.

Por otra parte, debemos destacar que es muy importante realizar una buena gestión de memoria desde el código, dado que el programa puede generar un desbordamiento de la memoria de la máquina en que se ejecute, debido a la reserva que se realiza para almacenar los datos relativos a los ficheros de audio de las dos locuciones a comparar. Por tanto, no sólo debemos preocuparnos por los "malloc()" de la reserva, sino también de los "free()" pertinentes para liberar el espacio utilizado (y ponerlos en el orden correcto).

Para compilar este programa introducimos en la línea de comandos "cc -o dist dist.c dtw.c -lm", haciendo referencia a ambos códigos. Además, es imprescindible añadir "-lm" al final, puesto que utilizamos la librería matemática y en caso de omitirlo el compilador devuelve un error, impidiendo que se genere el archivo ejecutable.

Una vez compilado dist.c, para ejecutarlo debemos teclear en la línea de comandos "./dist [path]", donde [path] representa el directorio raíz en el que se encuentran las carpetas de todos los locutores y, dentro de éstas, las carpetas que contienen los datos referentes a cada locución de cada frase por cada locutor. Según la estructura que hemos definido previamente, [path] sería en este caso /home/p/pfc/dvall/pcompare/corpus/data/frases. Si se ha mantenido una buena gestión de memoria, la ejecución debería finalizar correctamente, aunque podría prologarse durante varios minutos por el gran número de cálculos a realizar.

Después de haber ejecutado correctamente dist.c, y tras haber observado los resultados obtenidos, nos damos cuenta de que los resultados no son como esperábamos. El valor de las distancias es demasiado elevado y además algunas se "disparan" con valor infinito o "casi infinito" para dos repeticiones diferentes de un mismo locutor de una misma frase, lo cual no es razonable. Por ello, pensamos en que es necesario ponderar los parámetros que intervienen en el cálculo de las distancias, puesto que es posible que se le esté dando más importancia de la que se debe a un parámetro cuyo valor puede alcanzar cifras de nivel de centenas frente a valores normalizados cuyo valor oscila entre 0 y 1. Este efecto además se agrava al elevar al cuadrado para calcular la distancia euclídea, que es la que utilizamos en dtw.c.

En vista de esta situación, utilizaremos la herramienta WEKA, con la que esperamos poder decidir un vector de pesos que minimice las distancias entre las locuciones comparadas para poder tomar decisiones acerca de los límites que estableceremos con el fin de evaluar un ensayo, dictaminando si la entonación ha sido tan buena como la de un nativo o no, además de otorgarle una puntuación (score) que cuantifique esa proximidad a la pronunciación de un español. En realidad lo que más nos interesa es que las distancias de las locuciones de los nativos entre sí sean lo más pequeñas posible, y que las distancias de los nativos en comparación con los extranjeros sean lo más grandes posible, para poder clasificarlos con mayor facilidad.

5.4.8. Generación de ficheros .csv

Para que WEKA pueda interpretar los datos que hemos generado a partir de las grabaciones de audio de las locuciones de las frases, necesitamos que la información esté organizada en un modo que WEKA pueda entenderla. Un formato de fichero que WEKA puede leer y que podemos generar fácilmente es ".csv"; para ello tan sólo debemos incluir ';' entre los valores de los parámetros obtenidos del procesamiento en Praat y en Snack, los cuales, como se comentó con anterioridad, están recogidos en los ficheros ".ppar" y ".spar".

Con este propósito se han creado dos nuevo programas "grp2.c" y "grs2.c" (get resumen praat-snack 2) en la carpeta /home/p/pfc/dvall/pcompare/corpus/scripts/pitch/snack. La forma de actuar de estos programas es la misma que getResumenP.c y getResumenS.c: reciben como entrada los valores de los ficheros ".f0p" y ".f0a", respectivamente, y leen sus valores, calculando las características que se describieron para los ficheros ".ppar" y ".spar", pero esta vez pudiendo imprimirlas a un fichero de texto con el formato deseado (datos separados por ','), además de mostrarlas por pantalla.

Ahora lo que pretendemos no es generar un fichero por cada frase como hicimos en etapas anteriores, sino que lo que queremos es crear un único fichero que incluya todos los datos de todas las repeticiones de todas las frases ensayadas por todos los locutores, es decir, los datos que están incluidos en cada fichero de cada ensayo, pero reunidos todos juntos en un solo fichero ".csv". Al final de cada línea se incluyen 5 campos más: locutor, frase, repetición, si es nativo o no (n/e) y su sexo (m/f). Para ser exactos, generaremos 4 ficheros: uno con los datos de las cabeceras de los ficheros con datos procedentes de Praat y otro con los de Snack (valores globales, HeadersP.csv y HeadersS.csv, respectivamente), y otros dos con los datos de las series temporales (o intervalos, valores locales) calculados también con la ayuda de estas herramientas como se expuso previamente (IntervalsP.csv e IntervalsS.csv).

Para ejecutar estos programas, se ha utilizado el mismo procedimiento que para generar los ficheros ".ppar" y ".spar", es decir, se ha modificado ligeramente el código del script doGetResumen.sh de modo que ahora ejecuta los programas grp2.c y grs2.c en lugar de getResumenP.c y getResumenS.c. El script recorre el árbol de directorios a partir del nodo raíz definido (/home/p/pfc/dvall/pcompare/corpus/data/frases), detectando en qué subdirectorios existe un archivo de audio con formato ".wav". Para cada archivo ".wav" que encuentre, ejecuta grp2.c y grs2.c, los cuales leen los correspondientes ".f0p" y ".f0a" y abren el fichero de intervalos o cabeceras conveniente en el que añaden los datos del intervalo o la cabecera pertinente, respectivamente. De esta forma, podríamos generar los ficheros ".ppar", ".spar" y ".csv" en una sola ejecución redirigiendo la salida estándar a los ficheros ".ppar" y ".spar", mientras que el programa escribe directamente en el fichero ".csv".

5.4.9. Estudio de las características extraídas con WEKA

Con los ficheros ".csv" ya generados, podemos utilizar WEKA para tratar de deducir qué vector de pesos es el más conveniente para aplicar en conjunto con el algoritmo DTW que tenemos programado para mejorar las distancias calculadas de modo que podamos clasificar las locuciones como pronunciadas por un nativo o un extranjero con el mayor grado de acierto posible. En este apartado se pretende explicar cuáles han sido las pruebas experimentadas con WEKA y los resultados obtenidos que posteriormente estudiaremos para decidir, según el porcentaje de aciertos en la clasificación, cuáles son los pesos más adecuados para incluirlos en nuestro sistema de evaluación.

Antes de comenzar a probar las funciones de WEKA, hay que destacar que debemos preparar los ficheros ".csv" para que puedan ser interpretados por la herramienta. Para ello, es importante que añadamos una cabecera a los ficheros con los nombres de los parámetros, también separados por ',', de modo que nos puedan servir como "clase" a la hora de clasificar, agrupar o seleccionar atributos, entre otras funciones.

De entre las 4 aplicaciones que WEKA propone al iniciar (Explorer, Experimenter, KnowledgeFlow, Simple CLI), nos centraremos en la opción "Explorer".

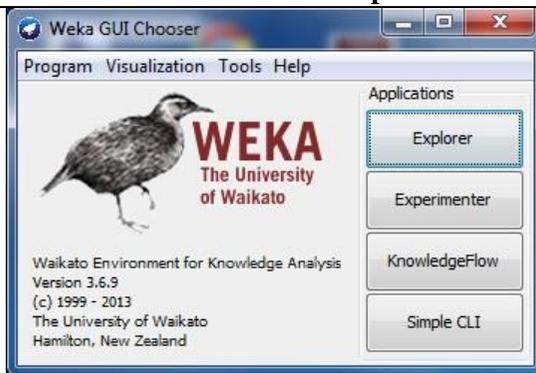


Ilustración 42: Pantalla de inicio de WEKA

Una vez que hemos accedido al contenido de la aplicación "Explorer", podemos observar que disponemos de un conjunto de pestañas con diferentes funcionalidades. Inicialmente WEKA nos sitúa en la pestaña "Preprocess", con la que podemos manipular, seleccionar e incluso añadir atributos e instancias de los datos de entrada para poder explotar al máximo el resto de funciones que nos ofrece, además de visualizar información del atributo seleccionado en un momento determinado. De las opciones que WEKA pone a nuestra disposición, nos interesan especialmente dos:

- Selección de atributos (Select Attributes)
- Clasificación (Classify)

En primer lugar, debemos cargar uno de los ficheros ".csv" (con la cabecera con los nombre de los parámetros incluida). Para ello, en la misma pestaña "Preprocess" pulsamos sobre el botón "Open file...", que nos permitirá elegir en la típica ventana de selección de archivos el fichero que deseemos.

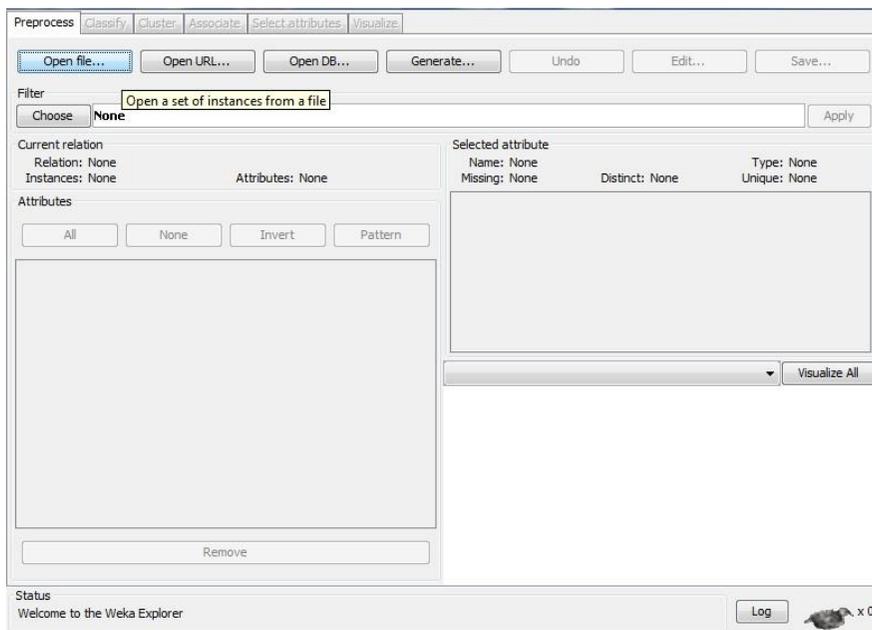


Ilustración 43: Abrir fichero en WEKA

Si el fichero se ha cargado correctamente, aparecerá en el panel izquierdo de la interfaz la lista con los nombres de los atributos (los parámetros que hemos calculado y almacenado en los ficheros) y en la derecha un gráfico que ilustra con colores las cantidades que puede haber de los distintos valores que puede adoptar el atributo que esté seleccionado en un instante concreto. Por defecto, WEKA selecciona el primer atributo de la lista.

Por otra parte, WEKA muestra el total de instancias y el número de atributos sobre el panel de la lista de atributos, y sobre el gráfico del panel de la derecha información referente al atributo seleccionado (nombre, tipo, cantidad de valores únicos, cantidad de valores distintos). Además, entre esta información y el gráfico de la parte inferior podemos observar una tabla con valores como el máximo, el mínimo, la media y la desviación de los valores de un atributo en caso de que éste sea numérico; y la etiqueta y la cantidad de muestras que se corresponden con cada etiqueta en caso de que el atributo en cuestión sea nominal.

Concretamente, nuestros ficheros de cabeceras tienen 31 atributos, 26 numéricos y 5 nominales, mientras que los de intervalos presentan 29 atributos, 24 numéricos y los mismos 5 nominales. Cabe destacar que uno de estos atributos nominales, nativo o no nativo, será el que utilicemos como clase para clasificar y seleccionar atributos.

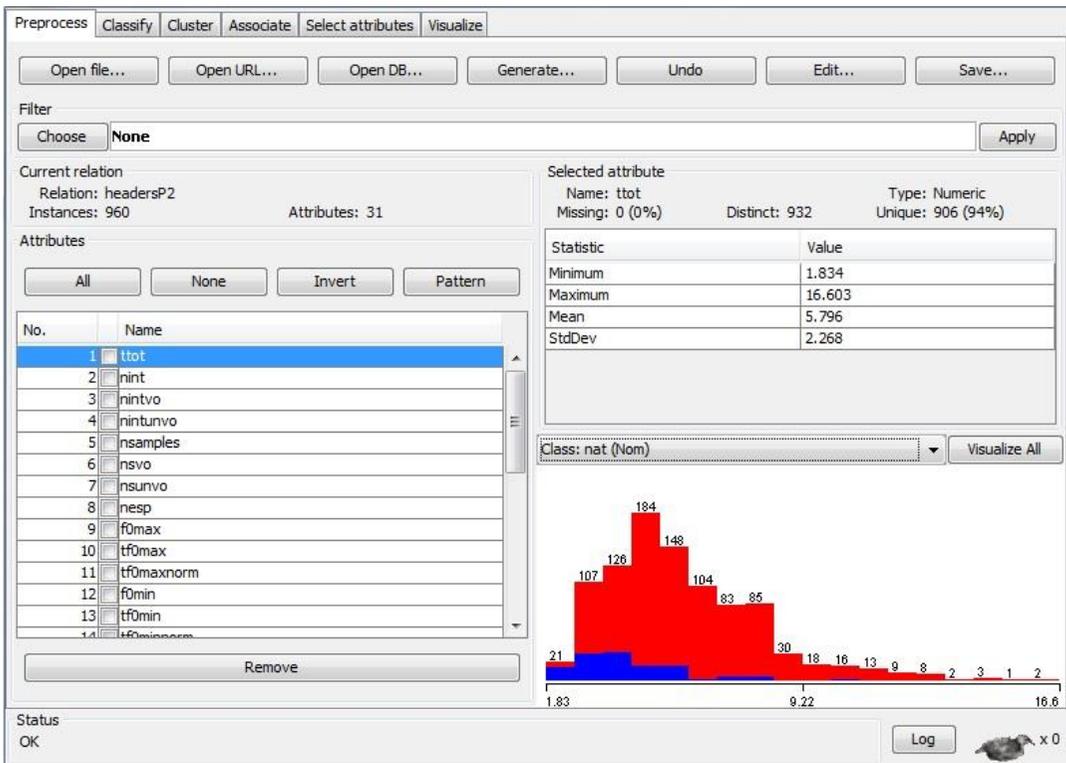


Ilustración 44: Preprocess con fichero cargado en WEKA

Desde la opción "Preprocess" también podemos seleccionar qué atributos o instancias consideramos para el análisis de los datos de las muestras, para ello WEKA pone a nuestra disposición una serie de "filtros" que podemos aplicar en cascada para quedarnos con los valores que estimemos oportuno. Estos "filtros" se escogen presionando sobre el botón "Choose". En este trabajo operaremos

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

con filtros no supervisados ("unsupervised"), que son independientes del algoritmo de análisis posterior. Si utilizáramos filtros supervisados, podríamos vernos obligados a utilizar determinados algoritmos que trabajan en asociación con unos filtros específicos.

Podemos aplicar filtros a nivel de atributo o a nivel de instancia. Los filtros de atributos operan verticalmente, si lo viéramos en forma de tabla veríamos que incide sobre las columnas. No se ocupan del valor concreto de las muestras, sino que actúan de forma genérica sobre todo el conjunto de muestras independientemente de sus valores. Por otra parte, los filtros de instancia intervienen sobre el conjunto de las muestras teniendo en cuenta sus valores, es decir, seleccionaría las muestras a considerar para el análisis fijándose en el valor que adquiere una instancia para un atributo determinado. Volviendo al símil con la representación de una tabla, este filtro operaría horizontalmente, sobre las filas.

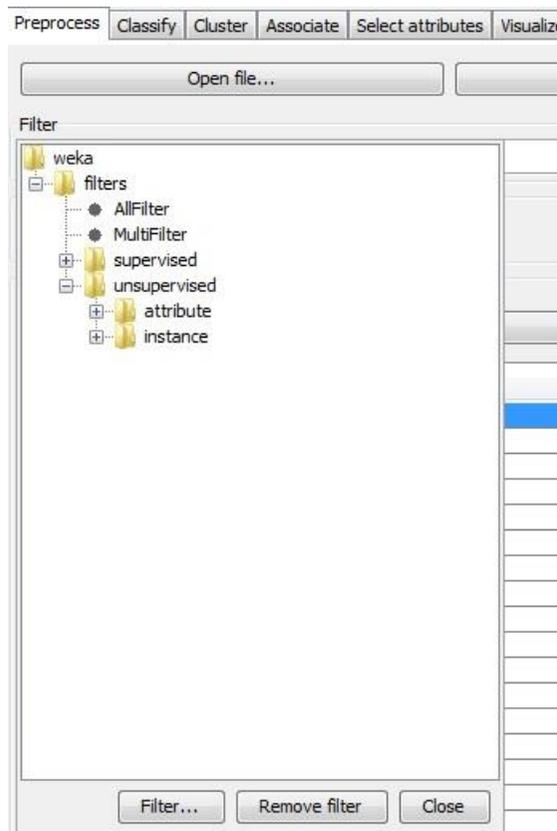


Ilustración 45: Escoger Filtro en Preprocess

Nos interesa estudiar las posibles selecciones de pesos en función del análisis de las muestras según criterios de distinción de sonoridad, sexos, frases o la combinación de varios de ellos, y el comportamiento de los modelos resultantes en lo referente a la clasificación de los ensayos como nativos o extranjeros. En vista de esto, necesitaremos aplicar varios de estos filtros según el criterio que escojamos. esencialmente, utilizaremos los filtros "Remove" y "NominalToBinary" de los filtros no supervisados de atributos; y el filtro "RemoveWithValues" de los filtros no supervisados de instancias.

- El filtro no supervisado de atributos "Remove", elimina del análisis los atributos que se le pasen como parámetro. La aplicación de este filtro tiene el mismo efecto sobre el conjunto de

muestras que si marcáramos los checkbox situados a la izquierda de los atributos en la lista del panel izquierdo de la interfaz de la pestaña "Preprocess" y a continuación pulsáramos el botón "Remove" de la parte inferior del panel.

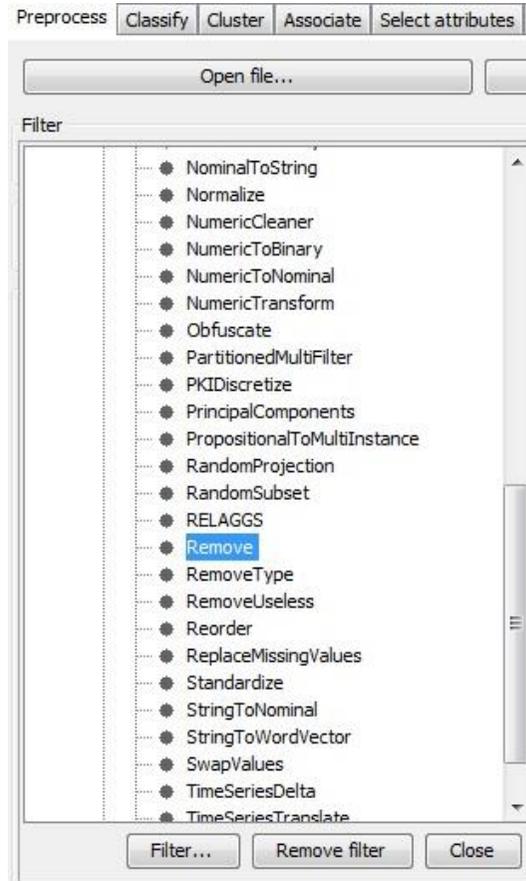


Ilustración 46: Filtro de atributos no supervisado "Remove"

Este filtro recibe exclusivamente como parámetro los índices de los atributos que deseamos desechar para el análisis. Podemos operar de forma inversa y pasarle los índices de los atributos que deseamos mantener, marcando como verdadero (True) el parámetro "invertSelection".

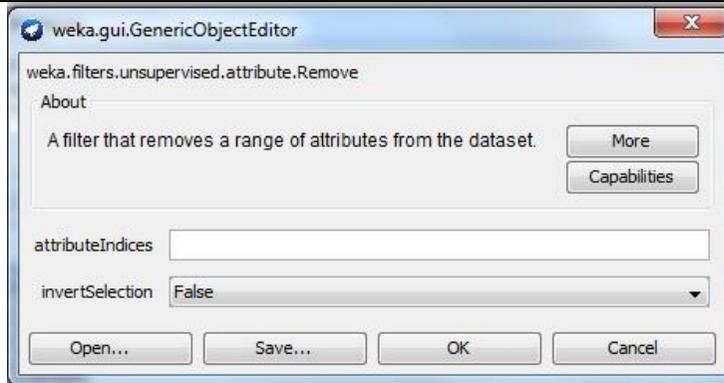


Ilustración 47: Parámetros del filtro de atributos no supervisado "Remove"

- El filtro no supervisado de atributos "NominalToBinary" transforma un atributo nominal con k valores posibles en k atributos binarios (representados como 0's y 1's). Nosotros lo utilizaremos para analizar los datos referentes a cada frase por separado. El id de cada frase es de la forma $s[0-9][0-9]$, y tenemos un total de 15 frases ($s01, s02, \dots, s15$), con lo cual se trata de un atributo nominal. Podríamos pensar en el filtro que veremos a continuación ("RemoveWithValues") como una clara opción para filtrar por frase, pero por sus características no es aplicable a los atributos nominales.

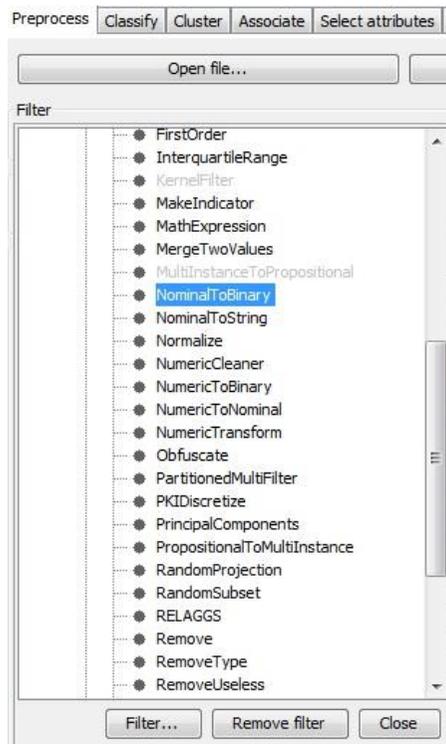


Ilustración 48: Filtro no supervisado de atributos "NominalToBinary"

Este filtro recibe como parámetros los índices de los atributos nominales que deseamos "descomponer". Al igual que en el filtro anterior podemos, además, introducir los índices de los atributos que deseamos mantener íntegros activando la opción "invertSelection". También disponemos de otras opciones que no utilizaremos.

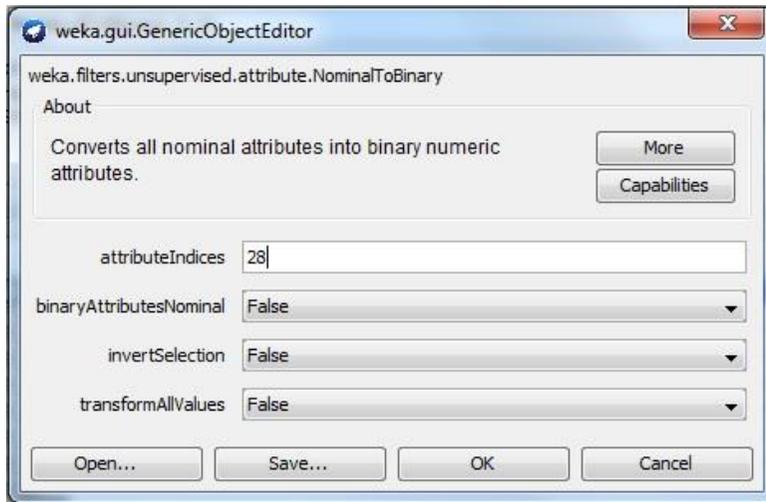


Ilustración 49: Parámetros del filtro de atributos no supervisado "NominalToBinary"

Como hemos anticipado, su efecto es dividir un atributo nominal con k valores posibles en k nuevos atributos.

28	<input type="checkbox"/>	sent=s15
29	<input type="checkbox"/>	sent=s04
30	<input type="checkbox"/>	sent=s10
31	<input type="checkbox"/>	sent=s09
32	<input type="checkbox"/>	sent=s07
33	<input type="checkbox"/>	sent=s03
34	<input type="checkbox"/>	sent=s11
35	<input type="checkbox"/>	sent=s14
36	<input type="checkbox"/>	sent=s13
37	<input type="checkbox"/>	sent=s05
38	<input type="checkbox"/>	sent=s08
39	<input type="checkbox"/>	sent=s02
40	<input type="checkbox"/>	sent=s06
41	<input type="checkbox"/>	sent=s01
42	<input type="checkbox"/>	sent=s12

Ilustración 50: Efecto de aplicar el filtro no supervisado de atributos "NominalToBinary" al atributo nominal "sent", con 15 valores posibles

- El filtro de instancias "RemoveWithValues" elimina del análisis las instancias cuyo valor para un atributo/s cuyo índice se ha pasado como parámetro supera o queda por debajo de un umbral establecido, también fijado a nuestra elección. En caso de que dispusiéramos de un

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

atributo nominal con dos valores posibles también podríamos utilizarlo para marginar un conjunto de muestras para el análisis.

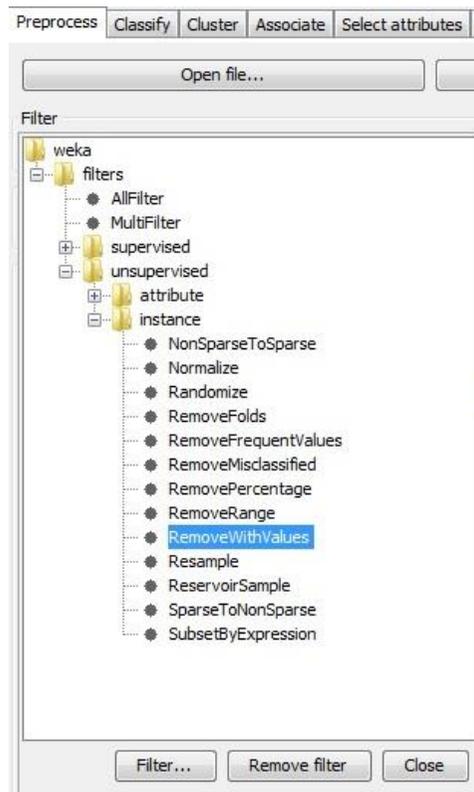


Ilustración 51: Filtro no supervisado de instancias "RemoveWithValues"

Este filtro recibe como parámetros el índice/s numéricos que se someterán al filtrado tomando como referencia el valor umbral (splitPoint). Las instancias cuyos valores para los atributos cuyo/s índice/s se ha seleccionado (AttributeIndex) son inferiores a este límite, son eliminados para el análisis y los superiores se mantienen. En este caso, la activación de la opción "invertSelection" hace que las muestras consideradas para el análisis sean las que tengan los valores de los atributos indicados por debajo del umbral establecido. En caso de que dispongamos de un atributo nominal con sólo dos valores posibles, podríamos indicar su índice en el parámetro "nominalIndices".

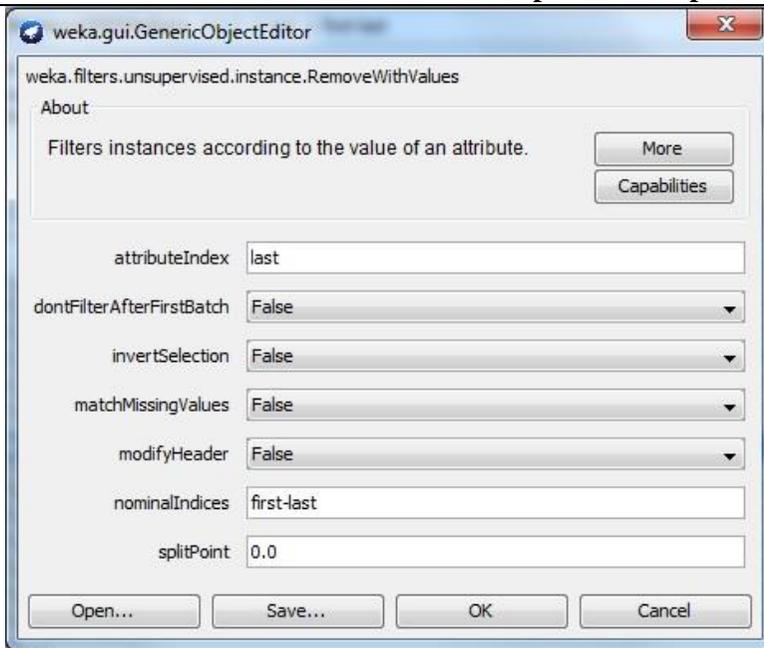


Ilustración 52: Parámetros del filtro no supervisado de instancias "RemoveWithValues"

Éstos son los filtros que aplicaremos para el análisis de los datos de las locuciones recogidos en los ficheros ".csv" que elaboramos en la etapa anterior. Los diferentes experimentos que se han realizado mediante la aplicación de varios de estos filtros de forma combinada sobre los datos globales y locales serán explicados posteriormente. A continuación se describe la manera de proceder en cuanto a la selección de atributos y a la clasificación sobre la totalidad de las muestras (sin aplicar filtros).

Como adelantamos, estamos especialmente en dos de las características de WEKA, la selección de atributos y la clasificación. Estamos buscando un modelo que nos ayude a clasificar con el mayor acierto posible las locuciones de las frases como pronunciadas por un nativo o por un extranjero. Las distancias que obtuvimos a través del algoritmo DTW no nos han satisfecho en primera vuelta, y razonando, hemos decidido ponderar nuestro vector de características para tratar de conseguir unas distancias más correctas. La función de selección de atributos de WEKA nos ayudará en este sentido.

Por regla general, antes de acceder a esta opción realizaríamos un "preprocesado" del fichero ".csv" aplicando los filtros que consideráramos necesarios para lograr nuestro objetivo. Sin embargo, puesto que ya se ha explicado cómo utilizar los filtros y dado que se va a exponer un ejemplo que ilustre los pasos que se han seguido, evitaremos repetirlo.

Después de haber preprocesado el fichero, nos dirigimos a la pestaña "Select attributes". Pretendemos saber qué atributos son los más importantes para clasificar las muestras como nativo o extranjero, así que necesitamos un evaluador de atributos que nos pueda dar esta información. Los evaluadores de atributos pueden trabajar con varios métodos de búsqueda. En nuestro caso, hemos escogido el evaluador "InfoGainAttributeEval", que evalúa la valía del atributo midiendo la ganancia de información con respecto a la case. La utilización de este evaluador nos obliga a emplear el método de búsqueda "Ranker", el cual ordena los atributos según sus evaluaciones individuales.

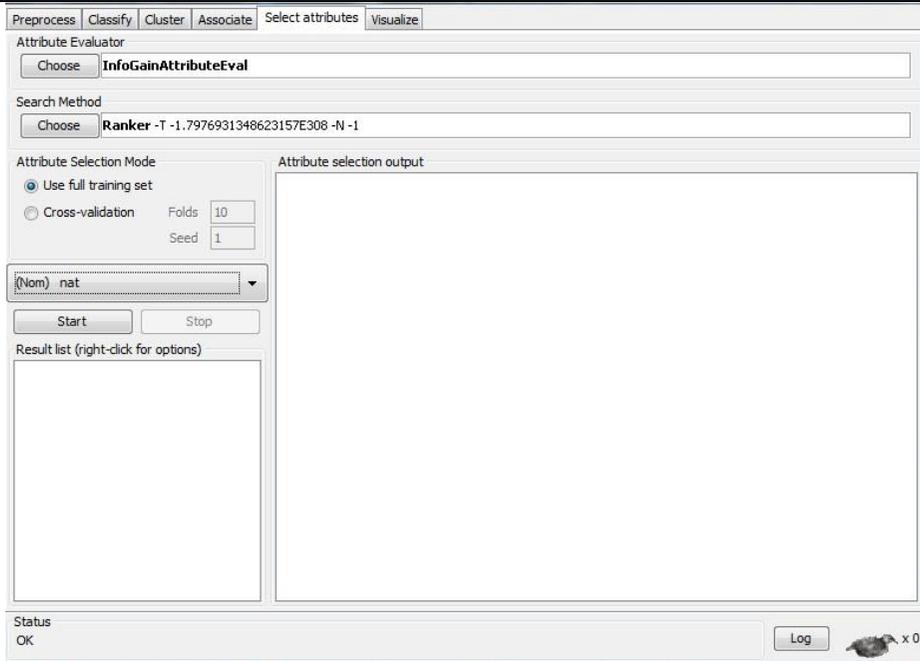


Ilustración 53: Select attributes en WEKA

En cuanto al modo de selección de atributos, se ha optado por emplear el conjunto de entrenamiento completo (Use full training set), frente a una validación cruzada. Como clase, seleccionamos el atributo "nat", de modo que se evaluará a los atributos según su grado de importancia para clasificar como nativo o no nativo. A continuación se pulsa el botón "Start" y se inicia la evaluación.

El resultado de la evaluación podemos verlo en el panel de la derecha, "Attribute selection output", en donde se nos muestra la orden ejecutada, con los atributos considerados y debajo, una lista ordenada en la que cada fila la componen el id del atributo, el nombre del atributo y su peso estimado. En los sucesivos experimentos, estos pesos serán los que utilizaremos para ponderar nuestro vector de características con el fin de obtener unas distancias que nos permitan distinguir lo mejor posible entre nativos y extranjeros. Además, este "ranking" de atributos nos permite analizar cuáles son los más determinantes para dicha clasificación.

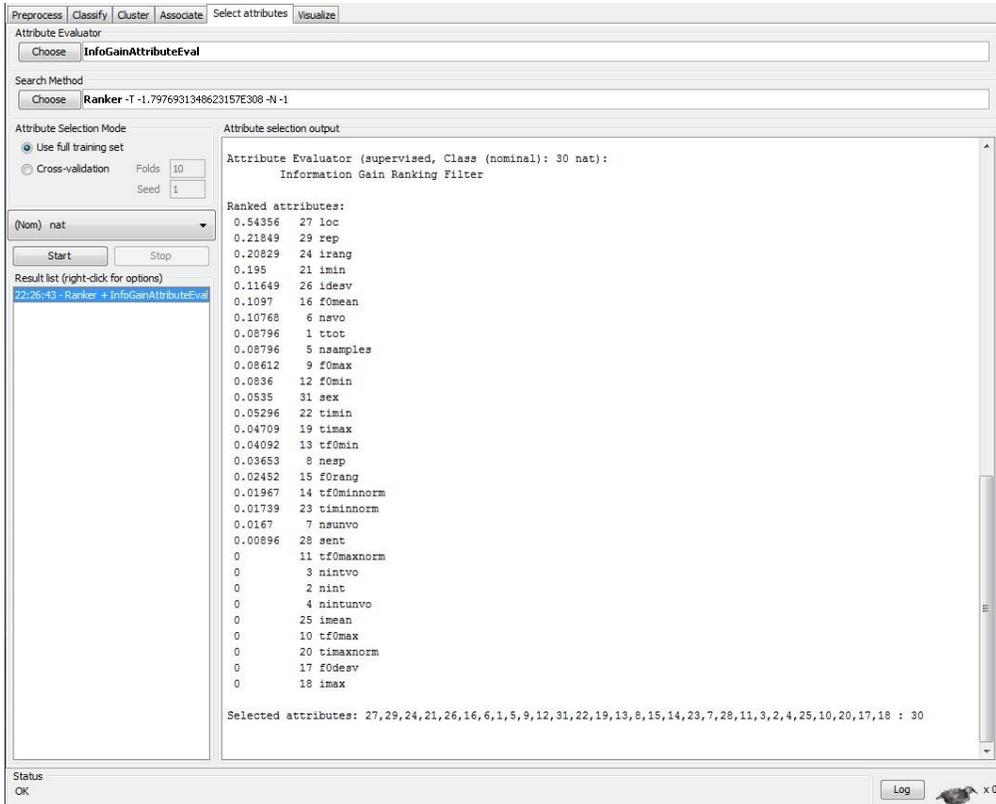


Ilustración 54: Resultados de la selección de atributos con WEKA

Después de haber mantenido los atributos que más nos interesan según los pesos deducidos de la fase de selección de atributos y de haber eliminado el resto, pasamos a comprobar cómo se comporta el modelo con los parámetros que nos quedan. Nos situamos en la pestaña de "Classify", donde podremos realizar una simulación de clasificación según el conjunto de de datos que le proporcionemos a WEKA.

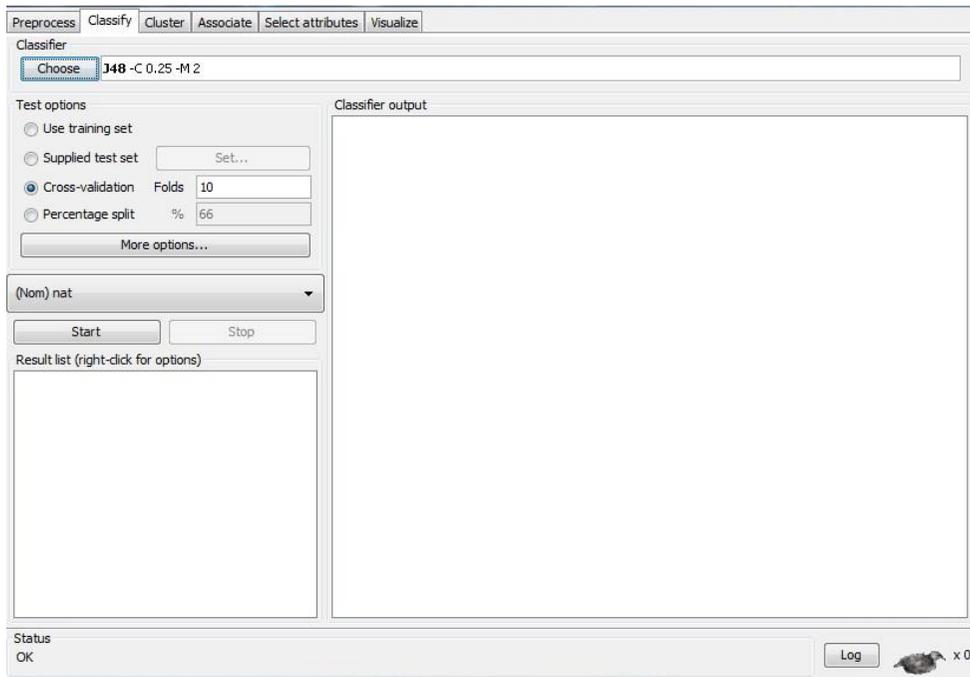


Ilustración 55: Classify en WEKA

En primer lugar, debemos escoger un clasificador. Para los experimentos que realizaremos más adelante, se ha decidido utilizar como clasificador lo que WEKA ha denominado como árbol de decisión J48, el cual se genera mediante el algoritmo C4.5. Lo siguiente que hacemos es elegir alguna de las opciones de prueba que se nos ofrecen:

- Use training set: se entrena con todos los valores disponibles, esto es, se utiliza la misma prueba para entrenar y probar, y después se evalúa. Los resultados son positivos pero no son reales.
- Supplied test set: permite aportar un nuevo archivo de datos (debe tener el mismo formato y tener los mismos atributos que el fichero con el que se ha entrenado), generalmente distintos a los de entrenamiento, con los que se evalúa el modelo.
- Cross-validation: se realiza la evaluación con la técnica de la evaluación cruzada, que consiste en dividir la muestra inicial en k partes, y para cada una de las k partes, entrenar el sistema con las $k-1$ restantes y evaluar con la parte k actual.
- Porcentaje split: se establece el porcentaje de muestras que se emplearán para entrenar al modelo, dejando al porcentaje restante para evaluarlo.

Como se puede observar, indicamos a WEKA qué muestra se utiliza para entrenar al sistema (training set), y sobre cuál se ejecuta para probarlo (test set). Se ha decidido que cualquiera de las dos últimas opciones, "Cross-validation", o "Porcentaje Split" puede ser la que más nos convenga. Obviamente la primera ha quedado descartada por la falta de realidad de los resultados procedentes de la evaluación, mientras que la segunda se descarta porque no disponemos de otro conjunto de datos con el que podamos probar el modelo. Así, finalmente hemos optado por el método de "Cross-validation" que divide el conjunto de muestras en las partes que le indiquemos y prueba el modelo tantas veces

como partes ha quedado dividido el conjunto. Una vez escogido el método, volvemos a marcar el atributo "nat" como clase y pulsamos sobre el botón "Start".

Al igual que ocurría en la pestaña de selección de atributos, los resultados de la evaluación se muestran en el panel de la derecha, donde podemos encontrar en primera instancia la orden ejecutado con los atributos considerados y ya debajo el árbol de decisión generado y un resumen de la evaluación junto con la matriz de confusión.

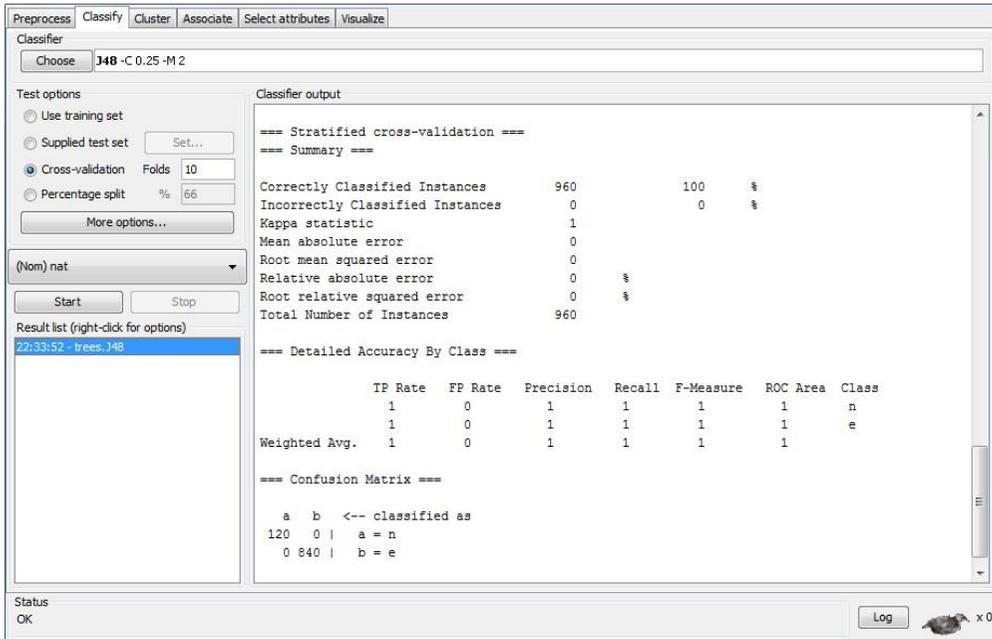


Ilustración 56: Resultados de clasificación en WEKA

Entre los resultados de la evaluación del modelo, obtenemos el porcentaje de instancias correcta e incorrectamente clasificadas, el error medio, el error cuadrático medio, el porcentaje de verdaderos positivos (TP) y falsos positivos (FP), y el área ROC (Receiver Operating Characteristic), que es el área debajo de la curva que representa la razón de verdaderos positivos frente a la razón de falsos positivos. A partir de estos valores podremos determinar si el modelo es lo suficientemente bueno o malo como para considerarlo y probarlo en nuestro sistema de valuación.

5.4.9.1. Experimentos realizados para la selección de atributos

Una vez se han explicado las opciones que utilizaremos para analizar nuestro conjunto de datos procedentes de las locuciones de las frases, y que están recogidos en los ficheros ".csv" que hemos abierto con WEKA, vamos a describir cuáles han sido los experimentos que se han llevado a cabo para deducir los pesos con los que ponderaremos el vector de características para calcular las distancias a la hora de aplicar el algoritmo DTW.

Antes de comenzar, aclarar que antes de las pruebas que se exponen a continuación, se ha decidido realizar una primera elección de características, excluyendo parámetros que a podríamos considerar inútiles o redundantes. Des este modo, nos quedarían los siguientes vectores:

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

- {f0max, tf0maxnorm, f0min, tf0minnorm, f0rang, f0mean, f0desv, imax, timaxnorm, imin, timinnorm, irang, imean, idesv, sent, nat, sex}, para los vectores de características globales (procedentes de los archivos de cabeceras).
- {voi, f0max, tf0maxnorm, f0min, tf0minnorm, f0rang, f0mean, f0desv, imax, timaxnorm, imin, timinnorm, irang, imean, idesv, sent, nat, sex}, para los vectores de características locales (procedentes de los archivos de intervalos).

Tanto el atributo numérico "voi" en el caso de las características locales, como los atributos nominales "sent", "nat" y "sex" en caso de ambos tipos de características, serán lo que tomemos como referencia para las pruebas que realizaremos. Cada prueba supondrá una modalidad de experimento. Cada modalidad representa una combinación de casos que pueden darse según consideremos los diferentes valores que pueden tomar estos atributos o no. Podemos estudiar los efectos de estimar los distintos valores de los atributos en combinación con los efectos de los distintos valores de otros, o por el contrario, evaluarlos de forma aislada.

5.4.9.1.1. MODALIDAD 1: Pesos deducidos del análisis del fichero de datos globales (estáticos) contenidos con Praat considerando todas las muestras

- Se carga el fichero de datos globales (cabeceras de los ficheros ".ppar") obtenidos con Praat.
- Se utilizar el filtro de atributos no supervisado "Remove" y se eliminan los atributos que nos sobran (los que no aparecen en el vector descrito anteriormente).
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:

1.	0.2241	1	f0max
2.	0.2171	12	irang
3.	0.2076	10	imin
4.	0.1514	3	f0min
5.	0.1248	6	f0mean
6.	0.0752	14	idesv
7.	0.0488	4	tf0minnorm
8.	0.031	5	f0rang
9.	0.0182	11	timinnorm
10.	0.0	2	tf0maxnorm
11.	0.0	8	imax
12.	0.0	13	imean
13.	0.0	9	timaxnorm
14.	0.0	7	f0desv

- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:

i)	Instancias clasificadas correctamente	904	94.1667%
ii)	Instancias clasificadas incorrectamente	56	5.8333%
iii)	Índice Kappa	0.7172	
iv)	Error absoluto medio	0.0673	
v)	Raíz del error cuadrático medio	0.2356	
vi)	Error absoluto relativo	30.6661%	
vii)	Raíz del error cuadrático relativo	71.226%	
viii)	TP nativos	0.7	

- ix) FP nativos 0.024
- x) TP extranjeros 0.976
- xi) FP extranjeros 0.3
- xii) Matriz de confusión:

Clasificados como	Nativos	Extranjeros
A=Nativos	84	36
B=Extranjeros	20	820

Tabla 5: Matriz de confusión modalidad 1

5.4.9.1.2. MODALIDAD 2: Pesos deducidos del análisis del fichero de datos globales (estáticos) obtenidos con Praat distinguiendo sexos

- Se carga el fichero de datos globales (cabeceras de los ficheros ".ppar") obtenidos con Praat.
- Se utiliza el filtro de atributos no supervisado "Remove" y se eliminan los atributos que nos sobran (los que no aparecen en el vector descrito anteriormente).

A. *LOCUTORES FEMENINOS*

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "sex". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "last". Ponemos como "splitPoint" el valor "f" (nos da igual cual pongamos, mientras no marquemos "invertSelection" como verdadero se quedará con las instancias cuyo valor para sexo es femenino).
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:
 1. 0.1401 12 irang
 2. 0.0999 10 imin
 3. 0.0973 6 f0mean
 4. 0.0412 14 idesv
 5. 0.0266 4 tf0minnorm
 6. 0.0231 11 timinnorm
 7. 0.0 2 tf0maxnorm
 8. 0.0 3 f0min
 9. 0.0 5 f0rang
 10. 0.0 16 sex
 11. 0.0 8 imax
 12. 0.0 7 f0desv
 13. 0.0 9 timaxnorm
 14. 0.0 13 imean
 15. 0.0 1 f0max
- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

i)	Instancias clasificadas correctamente	746	95.641%
ii)	Instancias clasificadas incorrectamente	34	4.359%
iii)	Índice Kappa	0.6503	
iv)	Error absoluto medio	0.0568	
v)	Raíz del error cuadrático medio	0.2038	
vi)	Error absoluto relativo	39.7364%	
vii)	Raíz del error cuadrático relativo	76.493%	
viii)	TP nativos	0.583	
ix)	FP nativos	0.013	
x)	TP extranjeros	0.988	
xi)	FP extranjeros	0.417	
xii)	Matriz de confusión:		

Clasificados como	Nativos	Extranjeros
A=Nativos	35	25
B=Extranjeros	9	711

Tabla 6: Matriz de confusión modalidad 2 – Locutores Femeninos

B. LOCUTORES MASCULINOS

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "sex". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "last". Ponemos como "splitPoint" el valor "f" y marcamos "invertSelection" como verdadero. En este caso se quedará con las instancias cuyo valor para sexo sea masculino.
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:
 1. 0.7767 12 irang
 2. 0.6877 3 f0min
 3. 0.5366 10 imin
 4. 0.5253 14 idesv
 5. 0.2876 8 imax
 6. 0.2471 1 f0max
 7. 0.2129 5 f0rang
 8. 0.182 13 imean
 9. 0.0849 7 f0desv
 10. 0.0 4 tf0minnorm
 11. 0.0 2 tf0maxnorm
 12. 0.0 9 timaxnorm
 13. 0.0 11 timinnorm
 14. 0.0 6 f0mean
 15. 0.0 16 sex

- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:
 - i) Instancias clasificadas correctamente 178 98.8889%
 - ii) Instancias clasificadas incorrectamente 2 1.111%
 - iii) Índice Kappa 0.975
 - iv) Error absoluto medio 0.0111
 - v) Raíz del error cuadrático medio 0.1054
 - vi) Error absoluto relativo 2.4962%
 - vii) Raíz del error cuadrático relativo 22.3605%
 - viii) TP nativos 0.983
 - ix) FP nativos 0.008
 - x) TP extranjeros 0.992
 - xi) FP extranjeros 0.017
 - xii) Matriz de confusión:

Clasificados como	Nativos	Extranjeros
A=Nativos	59	1
B=Extranjeros	1	119

Tabla 7: Matriz de confusión modalidad 2 – Locutores Masculinos

5.4.9.1.3. MODALIDAD 3: Pesos deducidos del análisis del fichero de datos locales (dinámicos) obtenidos con Praat considerando todas las muestras

- Se carga el fichero de datos locales (series temporales de los ficheros ".ppar") obtenidos con Praat.
- Se utilizar el filtro de atributos no supervisado "Remove" y se eliminan los atributos que nos sobran (los que no aparecen en el vector descrito anteriormente).
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:
 - 1. 0.0605 4 tf0minnorm
 - 2. 0.06038 3 f0min
 - 3. 0.0467 6 f0mean
 - 4. 0.04629 1 f0max
 - 5. 0.04356 11 timinnorm
 - 6. 0.04105 9 timaxnorm
 - 7. 0.03736 2 tf0maxnorm
 - 8. 0.0241 14 idesv
 - 9. 0.01819 7 f0desv
 - 10. 0.01729 10 imin
 - 11. 0.01437 13 imean
 - 12. 0.01212 12 irang
 - 13. 0.01024 5 f0rang
 - 14. 0.00761 8 imax
- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

i)	Instancias clasificadas correctamente	24536	90.743%
ii)	Instancias clasificadas incorrectamente	2503	9.257%
iii)	Índice Kappa	0.4581	
iv)	Error absoluto medio	0.154	
v)	Raíz del error cuadrático medio	0.2848	
vi)	Error absoluto relativo	70.7274%	
vii)	Raíz del error cuadrático relativo	86.3167%	
viii)	TP nativos	0.377	
ix)	FP nativos	0.017	
x)	TP extranjeros	0.983	
xi)	FP extranjeros	0.623	
xii)	Matriz de confusión:		

Clasificados como	Nativos	Extranjeros
A=Nativos	1266	2096
B=Extranjeros	407	23270

Tabla 8: Matriz de confusión modalidad 3

5.4.9.1.4. MODALIDAD 4: Pesos deducidos del análisis del fichero de datos locales (dinámicos) obtenidos con Praat distinguiendo sexos

- Se carga el fichero de datos locales (series temporales de los ficheros ".ppar") obtenidos con Praat.
- Se utiliza el filtro de atributos no supervisado "Remove" y se eliminan los atributos que nos sobran (los que no aparecen en el vector descrito anteriormente).

A. *LOCUTORES FEMENINOS*

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "sex". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "last". Ponemos como "splitPoint" el valor "f" (nos da igual cual pongamos, mientras no marquemos "invertSelection" como verdadero se quedará con las instancias cuyo valor para sexo es femenino).
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:
 1. 0.05618 4 f0min
 2. 0.0452 5 tf0minnorm
 3. 0.03796 10 timaxnorm
 4. 0.03663 2 f0max
 5. 0.03572 12 timinnorm
 6. 0.0349 7 f0mean
 7. 0.03006 3 tf0maxnorm
 8. 0.01679 15 idesv

9.	0.01098	11	imin
10.	0.01037	8	f0desv
11.	0.00829	13	irang
12.	0.00815	14	imean
13.	0.00596	16	sent
14.	0.0048	6	f0rang
15.	0.00239	9	imax
16.	0.0	18	sex
17.	0.0	1	voi

- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:

i)	Instancias clasificadas correctamente	20060	93.0427%
ii)	Instancias clasificadas incorrectamente	1500	6.9573%
iii)	Índice Kappa	0.2808	
iv)	Error absoluto medio	0.121	
v)	Raíz del error cuadrático medio	0.2488	
vi)	Error absoluto relativo	81.471%	
vii)	Raíz del error cuadrático relativo	93.3167%	
viii)	TP nativos	0.187	
ix)	FP nativos	0.004	
x)	TP extranjeros	0.996	
xi)	FP extranjeros	0.813	
xii)	Matriz de confusión:		

Clasificados como	Nativos	Extranjeros
A=Nativos	326	1415
B=Extranjeros	85	19734

Tabla 9: Matriz de confusión modalidad 4 – Locutores Femeninos

B. LOCUTORES MASCULINOS

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "sex". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "last". Ponemos como "splitPoint" el valor "f" y marcamos "invertSelection" como verdadero. En este caso se quedará con las instancias cuyo valor para sexo sea masculino.
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:

1.	0.1389	8	f0desv
2.	0.1388	4	f0min
3.	0.1295	5	tf0minnorm
4.	0.1102	10	timaxnorm
5.	0.1064	3	tf0maxnorm

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

6.	0.1061	7	f0mean
7.	0.1049	12	timinnorm
8.	0.1031	14	imean
9.	0.1027	2	f0max
10.	0.0991	9	imax
11.	0.079	6	f0rang
12.	0.0733	15	idesv
13.	0.0645	11	imin
14.	0.0386	13	irang
15.	0.0147	16	sent
16.	0.0	18	sex
17.	0.0	1	voi

- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:

i)	Instancias clasificadas correctamente	4729	86.3114%
ii)	Instancias clasificadas incorrectamente	750	13.6886%
iii)	Índice Kappa	0.6417	
iv)	Error absoluto medio	0.2047	
v)	Raíz del error cuadrático medio	0.3301	
vi)	Error absoluto relativo	49.1376%	
vii)	Raíz del error cuadrático relativo	72.3204%	
viii)	TP nativos	0.626	
ix)	FP nativos	0.037	
x)	TP extranjeros	0.963	
xi)	FP extranjeros	0.374	
xii)	Matriz de confusión:		

Clasificados como	Nativos	Extranjeros
A=Nativos	1014	607
B=Extranjeros	143	3715

Tabla 10: Matriz de confusión modalidad 4 – Locutores Masculinos

5.4.9.1.5. MODALIDAD 5: Pesos deducidos del análisis del fichero de datos locales (dinámicos) obtenidos con Praat distinguiendo intervalos sonoros y sordos

- Se carga el fichero de datos locales (series temporales de los ficheros ".ppar") obtenidos con Praat.
- Se utiliza el filtro de atributos no supervisado "Remove" y se eliminan los atributos que nos sobran (los que no aparecen en el vector descrito anteriormente).

A. INTERVALOS VOICED

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "voi". Si hemos eliminado los

Capítulo 5: Preparación del Corpus

atributos que hemos desestimado correctamente, este índice debería ser "first". Ponemos como "splitPoint" el valor "0.5", de modo que mientras el parámetro "invertSelection" se quedará con las instancias cuyo valor para "voi" es 1 (sonoro).

- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:

1.	0.12537	4	f0min
2.	0.12126	5	tf0minnorm
3.	0.09635	7	f0mean
4.	0.09538	2	f0max
5.	0.08305	10	timaxnorm
6.	0.07488	3	tf0maxnorm
7.	0.06276	12	timinnorm
8.	0.04931	11	imin
9.	0.04074	14	imean
10.	0.03807	8	f0desv
11.	0.02195	6	f0rang
12.	0.01975	13	irang
13.	0.0195	9	imax
14.	0.0034	15	idesv
15.	0.0	1	voi

- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:

i)	Instancias clasificadas correctamente	12049	92.2871%
ii)	Instancias clasificadas incorrectamente	1007	7.7129%
iii)	Índice Kappa	0.593	
iv)	Error absoluto medio	0.1288	
v)	Raíz del error cuadrático medio	0.257	
vi)	Error absoluto relativo	58.8757%	
vii)	Raíz del error cuadrático relativo	77.7391%	
viii)	TP nativos	0.536	
ix)	FP nativos	0.022	
x)	TP extranjeros	0.978	
xi)	FP extranjeros	0.464	
xii)	Matriz de confusión:		

Clasificados como	Nativos	Extranjeros
A=Nativos	874	757
B=Extranjeros	250	11175

Tabla 11: Matriz de confusión modalidad 5 – Intervalos Voiced

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

B. INTERVALOS UNVOICED

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "voi". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "first". Ponemos como "splitPoint" el valor "0.5" y marcamos "invertSelection" como verdadero. En este caso se quedará con las instancias cuyo valor para "voi" sea 0 (intervalo sordo).
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:

1.	0.06584	12	timinnorm
2.	0.05037	15	idesv
3.	0.03615	13	irang
4.	0.03192	11	imin
5.	0.02659	10	timaxnorm
6.	0.02216	14	imean
7.	0.00715	9	imax
8.	0.00404	3	tf0maxnorm
9.	0.00404	5	tf0minnorm
10.	0.0	4	f0min
11.	0.0	2	f0max
12.	0.0	8	f0desv
13.	0.0	6	f0rang
14.	0.0	7	f0mean
15.	0.0	1	voi

- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:
 - i) Instancias clasificadas correctamente 12461 89.1154%
 - ii) Instancias clasificadas incorrectamente 1522 10.8846%
 - iii) Índice Kappa 0.3137
 - iv) Error absoluto medio 0.1762
 - v) Raíz del error cuadrático medio 0.3068
 - vi) Error absoluto relativo 81.193%
 - vii) Raíz del error cuadrático relativo 93.1471%
 - viii) TP nativos 0.247
 - ix) FP nativos 0.018
 - x) TP extranjeros 0.982
 - xi) FP extranjeros 0.753
 - xii) Matriz de confusión:

Clasificados como	Nativos	Extranjeros
A=Nativos	428	1303
B=Extranjeros	219	12033

Tabla 12: Matriz de confusión modalidad 5 – Intervalos Unvoiced

5.4.9.1.6. MODALIDAD 6: Pesos deducidos del análisis del fichero de datos locales (dinámicos) obtenidos con Praat distinguiendo intervalos sonoros y sordos y distinguiendo sexos

- Se carga el fichero de datos locales (series temporales de los ficheros ".ppar") obtenidos con Praat.
- Se utiliza el filtro de atributos no supervisado "Remove" y se eliminan los atributos que nos sobran (los que no aparecen en el vector descrito anteriormente).

A. *LOCUTORES FEMENINOS E INTERVALOS VOICED*

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "sex". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "last". Ponemos como "splitPoint" el valor "f" (nos da igual cual pongamos, mientras no marquemos "invertSelection" como verdadero se quedará con las instancias cuyo valor para sexo es femenino).
- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "voi". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "first". Ponemos como "splitPoint" el valor "0.5", de modo que mientras el parámetro "invertSelection" se quedará con las instancias cuyo valor para "voi" es 1 (sonoro).
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:

1.	0.11498	4	f0min
2.	0.0904	5	tf0minnorm
3.	0.07562	2	f0max
4.	0.0732	10	timaxnorm
5.	0.07091	7	f0mean
6.	0.06052	3	tf0maxnorm
7.	0.04868	12	timinnorm
8.	0.02713	11	imin
9.	0.02209	14	imean
10.	0.02066	8	f0desv
11.	0.01437	13	irang
12.	0.01052	6	f0rang
13.	0.01035	9	imax
14.	0.0038	15	idesv
15.	0.0	17	sex
16.	0.0	1	voi

- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:
 - i) Instancias clasificadas correctamente 9771 93.9881%
 - ii) Instancias clasificadas incorrectamente 625 6.0119%
 - iii) Índice Kappa 0.4748
 - iv) Error absoluto medio 0.1043
 - v) Raíz del error cuadrático medio 0.2322
 - vi) Error absoluto relativo 69.7194%

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

vii) Raíz del error cuadrático relativo	84.9337%
viii) TP nativos	0.374
ix) FP nativos	0.01
x) TP extranjeros	0.99
xi) FP extranjeros	0.626
xii) Matriz de confusión:	

Clasificados como	Nativos	Extranjeros
A=Nativos	316	530
B=Extranjeros	95	9455

Tabla 13: Matriz de confusión modalidad 6 – Locutores Femeninos e Intervalos Voiced

B. LOCUTORES FEMENINOS E INTERVALOS UNVOICED

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "sex". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "last". Ponemos como "splitPoint" el valor "f" (nos da igual cual pongamos, mientras no marquemos "invertSelection" como verdadero se quedará con las instancias cuyo valor para sexo es femenino).
- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "voi". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "first". Ponemos como "splitPoint" el valor "0.5" y marcamos "invertSelection" como verdadero. En este caso se quedará con las instancias cuyo valor para "voi" sea 0 (intervalo sordo).
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:
 1. 0.05165 12 timinnorm
 2. 0.03151 15 idesv
 3. 0.02443 10 timaxnorm
 4. 0.02425 11 imin
 5. 0.02179 13 irang
 6. 0.01534 14 imean
 7. 0.00495 9 imax
 8. 0.00328 3 tf0maxnorm
 9. 0.00328 5 tf0minnorm
 10. 0.0 2 f0max
 11. 0.0 4 f0min
 12. 0.0 17 sex
 13. 0.0 8 f0desv
 14. 0.0 6 f0rang
 15. 0.0 7 f0mean
 16. 0.0 1 voi

- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:
 - i) Instancias clasificadas correctamente 10305 92.3056%
 - ii) Instancias clasificadas incorrectamente 859 7.6944%
 - iii) Índice Kappa 0.1682
 - iv) Error absoluto medio 0.1292
 - v) Raíz del error cuadrático medio 0.2581
 - vi) Error absoluto relativo 87.5655%
 - vii) Raíz del error cuadrático relativo 95.0352%
 - viii) TP nativos 0.112
 - ix) FP nativos 0.006
 - x) TP extranjeros 0.994
 - xi) FP extranjeros 0.888
 - xii) Matriz de confusión:

Clasificados como	Nativos	Extranjeros
A=Nativos	100	795
B=Extranjeros	64	10205

Tabla 14: Matriz de confusión modalidad 6 – Locutores Femeninos e Intervalos Unvoiced

C. LOCUTORES MASCULINOS E INTERVALOS VOICED

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "sex". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "last". Ponemos como "splitPoint" el valor "f" y marcamos "invertSelection" como verdadero. En este caso se quedará con las instancias cuyo valor para sexo sea masculino.
- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "voi". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "first". Ponemos como "splitPoint" el valor "0.5", de modo que mientras el parámetro "invertSelection" se quedará con las instancias cuyo valor para "voi" es 1 (sonoro).
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:
 - 1. 0.2912 8 f0desv
 - 2. 0.283 4 f0min
 - 3. 0.2815 14 imean
 - 4. 0.2701 5 tf0minnorm
 - 5. 0.2259 10 timaxnorm
 - 6. 0.2221 3 tf0maxnorm
 - 7. 0.2113 7 f0mean
 - 8. 0.205 9 imax

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

9.	0.2023	2	f0max
10.	0.1821	11	imin
11.	0.1664	6	f0rang
12.	0.1465	12	timinnorm
13.	0.0428	15	idesv
14.	0.0239	13	irang
15.	0.0	17	sex
16.	0.0	1	voi

- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:

i)	Instancias clasificadas correctamente	2497	93.8722%
ii)	Instancias clasificadas incorrectamente	163	6.1278%
iii)	Índice Kappa	0.8501	
iv)	Error absoluto medio	0.1024	
v)	Raíz del error cuadrático medio	0.2356	
vi)	Error absoluto relativo	24.6164%	
vii)	Raíz del error cuadrático relativo	51.6469%	
viii)	TP nativos	0.866	
ix)	FP nativos	0.031	
x)	TP extranjeros	0.969	
xi)	FP extranjeros	0.134	
xii)	Matriz de confusión:		

Clasificados como	Nativos	Extranjeros
A=Nativos	680	105
B=Extranjeros	58	1817

Tabla 15: Matriz de confusión modalidad 6 – Locutores Masculinos e Intervalos Voiced

D. LOCUTORES MASCULINOS E INTERVALOS UNVOICED

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "sex". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "last". Ponemos como "splitPoint" el valor "f" y marcamos "invertSelection" como verdadero. En este caso se quedará con las instancias cuyo valor para sexo sea masculino.
- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "voi". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "first". Ponemos como "splitPoint" el valor "0.5" y marcamos "invertSelection" como verdadero. En este caso se quedará con las instancias cuyo valor para "voi" sea 0 (intervalo sordo).
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:

1.	0.15873	12	timinnorm
----	---------	----	-----------

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:

1.	0.1988	10	imin
2.	0.1249	6	f0mean
3.	0.0898	1	f0max
4.	0.0441	5	f0rang
5.	0.0431	4	tf0minnorm
6.	0.0406	3	f0min
7.	0.0365	14	idesv
8.	0.0238	11	timinnorm
9.	0.0177	13	imean
10.	0.0	9	timaxnorm
11.	0.0	2	tf0maxnorm
12.	0.0	12	irang
13.	0.0	8	imax
14.	0.0	7	f0desv

- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:

i)	Instancias clasificadas correctamente	909	94.6875%
ii)	Instancias clasificadas incorrectamente	51	5.3125%
iii)	Índice Kappa	0.7395	
iv)	Error absoluto medio	0.0679	
v)	Raíz del error cuadrático medio	0.2258	
vi)	Error absoluto relativo	30.9365%	
vii)	Raíz del error cuadrático relativo	68.2722%	
viii)	TP nativos	0.708	
ix)	FP nativos	0.019	
x)	TP extranjeros	0.981	
xi)	FP extranjeros	0.292	
xii)	Matriz de confusión:		

Clasificados como	Nativos	Extranjeros
A=Nativos	85	35
B=Extranjeros	16	824

Tabla 17: Matriz de confusión modalidad 7

5.4.9.1.8. MODALIDAD 8: Pesos deducidos del análisis del fichero de datos globales (estáticos) obtenidos con Snack distinguiendo sexos

- Se carga el fichero de datos globales (cabeceras de los ficheros ".spar") obtenidos con Snack.
- Se utiliza el filtro de atributos no supervisado "Remove" y se eliminan los atributos que nos sobran (los que no aparecen en el vector descrito anteriormente).

A. LOCUTORES FEMENINOS

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "sex". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "last". Ponemos como "splitPoint" el valor "f" (nos da igual cual pongamos, mientras no marquemos "invertSelection" como verdadero se quedará con las instancias cuyo valor para sexo es femenino).
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:

1.	0.1248	10	imin
2.	0.1053	6	f0mean
3.	0.0332	1	f0max
4.	0.0328	11	timinnorm
5.	0.0233	4	tf0minnorm
6.	0.0164	3	f0min
7.	0.0	2	tf0maxnorm
8.	0.0	5	f0rang
9.	0.0	16	sex
10.	0.0	7	f0desv
11.	0.0	14	idesv
12.	0.0	13	imean
13.	0.0	12	irang
14.	0.0	9	timaxnorm
15.	0.0	8	imax

- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:
 - i) Instancias clasificadas correctamente 737 94.4872%
 - ii) Instancias clasificadas incorrectamente 43 5.5128%
 - iii) Índice Kappa 0.5762
 - iv) Error absoluto medio 0.0688
 - v) Raíz del error cuadrático medio 0.2311
 - vi) Error absoluto relativo 48.0837%
 - vii) Raíz del error cuadrático relativo 86.7091%
 - viii) TP nativos 0.55
 - ix) FP nativos 0.022
 - x) TP extranjeros 0.978
 - xi) FP extranjeros 0.45
 - xii) Matriz de confusión:

Clasificados como	Nativos	Extranjeros
A=Nativos	33	27
B=Extranjeros	16	704

Tabla 18: Matriz de confusión modalidad 8 – Locutores Femeninos

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

B. LOCUTORES MASCULINOS

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "sex". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "last". Ponemos como "splitPoint" el valor "f" y marcamos "invertSelection" como verdadero. En este caso se quedará con las instancias cuyo valor para sexo sea masculino.
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:

1.	0.7067	14	idesv
2.	0.4593	10	imin
3.	0.4485	13	imean
4.	0.3304	12	irang
5.	0.3124	8	imax
6.	0.285	7	f0desv
7.	0.2848	5	f0rang
8.	0.2783	3	f0min
9.	0.1677	1	f0max
10.	0.1574	4	tf0minnorm
11.	0.0624	2	tf0maxnorm
12.	0.0	6	f0mean
13.	0.0	11	timinnorm
14.	0.0	9	timaxnorm
15.	0.0	16	sex

- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:

i)	Instancias clasificadas correctamente	168	93.3333%
ii)	Instancias clasificadas incorrectamente	12	6.6667%
iii)	Índice Kappa	0.8487	
iv)	Error absoluto medio	0.0753	
v)	Raíz del error cuadrático medio	0.2534	
vi)	Error absoluto relativo	16.9055%	
vii)	Raíz del error cuadrático relativo	53.7598%	
viii)	TP nativos	0.883	
ix)	FP nativos	0.042	
x)	TP extranjeros	0.958	
xi)	FP extranjeros	0.117	
xii)	Matriz de confusión:		

Clasificados como	Nativos	Extranjeros
A=Nativos	53	7
B=Extranjeros	5	115

Tabla 19: Matriz de confusión modalidad 8 – Locutores Masculinos

5.4.9.1.9. MODALIDAD 9: Pesos deducidos del análisis del fichero de datos locales (dinámicos) obtenidos con Snack considerando todas las muestras

- Se carga el fichero de datos locales (series temporales de los ficheros ".spar") obtenidos con Snack.
- Se utilizar el filtro de atributos no supervisado "Remove" y se eliminan los atributos que nos sobran (los que no aparecen en el vector descrito anteriormente).
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:

1.	0.07743	4	tf0minnorm
2.	0.0598	9	timaxnorm
3.	0.05946	11	timinnorm
4.	0.05409	1	f0max
5.	0.05402	6	f0mean
6.	0.04953	3	f0min
7.	0.04844	2	tf0maxnorm
8.	0.02348	7	f0desv
9.	0.01415	10	imin
10.	0.0095	14	idesv
11.	0.00761	5	f0rang
12.	0.00748	13	imean
13.	0.00248	8	imax
14.	0.00108	12	irang

- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:
 - i) Instancias clasificadas correctamente 17577 85.5703%
 - ii) Instancias clasificadas incorrectamente 2964 14.4297%
 - iii) Índice Kappa 0.0176
 - iv) Error absoluto medio 0.2451
 - v) Raíz del error cuadrático medio 0.3502
 - vi) Error absoluto relativo 98.9072%
 - vii) Raíz del error cuadrático relativo 99.485%
 - viii) TP nativos 0.012
 - ix) FP nativos 0.001
 - x) TP extranjeros 0.999
 - xi) FP extranjeros 0.988
 - xii) Matriz de confusión:

Clasificados como	Nativos	Extranjeros
A=Nativos	1266	2096
B=Extranjeros	407	23270

Tabla 20: Matriz de confusión modalidad 9

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

5.4.9.1.10. MODALIDAD 10: Pesos deducidos del análisis del fichero de datos locales (dinámicos) obtenidos con Snack distinguiendo sexos

- Se carga el fichero de datos locales (series temporales de los ficheros ".spar") obtenidos con Snack.
- Se utiliza el filtro de atributos no supervisado "Remove" y se eliminan los atributos que nos sobran (los que no aparecen en el vector descrito anteriormente).

A. LOCUTORES FEMENINOS

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "sex". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "last". Ponemos como "splitPoint" el valor "f" (nos da igual cual pongamos, mientras no marquemos "invertSelection" como verdadero se quedará con las instancias cuyo valor para sexo es femenino).
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:

1.	0.06207	5	tf0minnorm
2.	0.05223	10	timaxnorm
3.	0.04674	12	timinnorm
4.	0.04167	4	f0min
5.	0.04037	3	tf0maxnorm
6.	0.03716	2	f0max
7.	0.03491	7	f0mean
8.	0.01138	11	imin
9.	0.01102	8	f0desv
10.	0.0064	16	sent
11.	0.0054	15	idesv
12.	0.00259	6	f0rang
13.	0.00125	14	imean
14.	0.00111	13	irang
15.	0.0	18	sex
16.	0.0	9	imax
17.	0.0	1	voi

- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:
 - i) Instancias clasificadas correctamente 15022 90.8113%
 - ii) Instancias clasificadas incorrectamente 1520 9.1887%
 - iii) Índice Kappa 0
 - iv) Error absoluto medio 0.1669
 - v) Raíz del error cuadrático medio 0.2889
 - vi) Error absoluto relativo 99.9732%
 - vii) Raíz del error cuadrático relativo 100%
 - viii) TP nativos 0
 - ix) FP nativos 0
 - x) TP extranjeros 1
 - xi) FP extranjeros 1
 - xii) Matriz de confusión:

Clasificados como	Nativos	Extranjeros
A=Nativos	0	1520
B=Extranjeros	0	15022

Tabla 21: Matriz de confusión modalidad 10 – Locutores Femeninos

B. LOCUTORES MASCULINOS

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "sex". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "last". Ponemos como "splitPoint" el valor "f" y marcamos "invertSelection" como verdadero. En este caso se quedará con las instancias cuyo valor para sexo sea masculino.
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:

1.	0.2145	8	f0desv
2.	0.1842	5	tf0minnorm
3.	0.1787	10	timaxnorm
4.	0.1684	12	timinnorm
5.	0.1531	15	idesv
6.	0.1339	3	tf0maxnorm
7.	0.1124	7	f0mean
8.	0.0939	4	f0min
9.	0.0915	14	imean
10.	0.0751	2	f0max
11.	0.0658	9	imax
12.	0.0588	13	irang
13.	0.0562	6	f0rang
14.	0.026	11	imin
15.	0.0208	16	sent
16.	0.0	18	sex
17.	0.0	1	voi
- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:

i)	Instancias clasificadas correctamente	3435	85.8965%
ii)	Instancias clasificadas incorrectamente	564	14.1035%
iii)	Índice Kappa	0.6967	
iv)	Error absoluto medio	0.1951	
v)	Raíz del error cuadrático medio	0.3338	
vi)	Error absoluto relativo	42.1223%	
vii)	Raíz del error cuadrático relativo	69.3622%	
viii)	TP nativos	0.816	
ix)	FP nativos	0.116	
x)	TP extranjeros	0.884	

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

xi) FP extranjeros 0.184

xii) Matriz de confusión:

Clasificados como	Nativos	Extranjeros
A=Nativos	1188	268
B=Extranjeros	296	2247

Tabla 22: Matriz de confusión modalidad 10 – Locutores Masculinos

5.4.9.1.11. MODALIDAD 11: Pesos deducidos del análisis del fichero de datos locales (dinámicos) obtenidos con Snack distinguiendo intervalos sonoros y sordos

- Se carga el fichero de datos locales (series temporales de los ficheros ".spar") obtenidos con Snack.
- Se utiliza el filtro de atributos no supervisado "Remove" y se eliminan los atributos que nos sobran (los que no aparecen en el vector descrito anteriormente).

A. *INTERVALOS VOICED*

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "voi". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "first". Ponemos como "splitPoint" el valor "0.5", de modo que mientras el parámetro "invertSelection" se quedará con las instancias cuyo valor para "voi" es 1 (sonoro).
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:

1.	0.16071	5	tf0minnorm
2.	0.11983	10	timaxnorm
3.	0.11304	2	f0max
4.	0.1126	7	f0mean
5.	0.10573	12	timinnorm
6.	0.10222	3	tf0maxnorm
7.	0.10213	4	f0min
8.	0.03198	8	f0desv
9.	0.02341	15	idesv
10.	0.01656	6	f0rang
11.	0.01353	14	imean
12.	0.01229	11	imin
13.	0.0039	9	imax
14.	0.0	13	irang
15.	0.0	1	voi

- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:

i)	Instancias clasificadas correctamente	8705	88.6998%
ii)	Instancias clasificadas incorrectamente	1109	11.3002%

iii) Índice Kappa	0.5022
iv) Error absoluto medio	0.1681
v) Raíz del error cuadrático medio	0.305
vi) Error absoluto relativo	67.0222%
vii) Raíz del error cuadrático relativo	86.162%
viii) TP nativos	0.501
ix) FP nativos	0.047
x) TP extranjeros	0.953
xi) FP extranjeros	0.499
xii) Matriz de confusión:	

Clasificados como	Nativos	Extranjeros
A=Nativos	723	719
B=Extranjeros	390	7982

Tabla 23: Matriz de confusión modalidad 11 – Intervalos Voiced

B. INTERVALOS UNVOICED

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "voi". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "first". Ponemos como "splitPoint" el valor "0.5" y marcamos "invertSelection" como verdadero. En este caso se quedará con las instancias cuyo valor para "voi" sea 0 (intervalo sordo).
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:

1.	0.0562	12	timinnorm
2.	0.03982	10	timaxnorm
3.	0.02866	11	imin
4.	0.019	8	f0desv
5.	0.00893	14	imean
6.	0.00529	15	idesv
7.	0.00222	9	imax
8.	0.0018	13	irang
9.	0.00126	3	tf0maxnorm
10.	0.00126	5	tf0minnorm
11.	0.0	6	f0rang
12.	0.0	2	f0max
13.	0.0	7	f0mean
14.	0.0	4	f0min
15.	0.0	1	voi
- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:

i) Instancias clasificadas correctamente	9317	86.8556%
--	------	----------

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

ii) Instancias clasificadas incorrectamente	1410	13.1444%
iii) Índice Kappa	0.2555	
iv) Error absoluto medio	0.2097	
v) Raíz del error cuadrático medio	0.3318	
vi) Error absoluto relativo	85.537%	
vii) Raíz del error cuadrático relativo	94.792%	
viii) TP nativos	0.205	
ix) FP nativos	0.021	
x) TP extranjeros	0.979	
xi) FP extranjeros	0.795	
xii) Matriz de confusión:		

Clasificados como	Nativos	Extranjeros
A=Nativos	314	1220
B=Extranjeros	190	9003

Tabla 24: Matriz de confusión modalidad 11 – Intervalos Unvoiced

5.4.9.1.12. MODALIDAD 12: Pesos deducidos del análisis del fichero de datos locales (dinámicos) obtenidos con Snack distinguiendo intervalos sonoros y sordos y distinguiendo sexos

- Se carga el fichero de datos locales (series temporales de los ficheros ".spar") obtenidos con Snack.
- Se utiliza el filtro de atributos no supervisado "Remove" y se eliminan los atributos que nos sobran (los que no aparecen en el vector descrito anteriormente).

A. *LOCUTORES FEMENINOS E INTERVALOS VOICED*

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "sex". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "last". Ponemos como "splitPoint" el valor "f" (nos da igual cual pongamos, mientras no marquemos "invertSelection" como verdadero se quedará con las instancias cuyo valor para sexo es femenino).
- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "voi". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "first". Ponemos como "splitPoint" el valor "0.5", de modo que mientras el parámetro "invertSelection" se quedará con las instancias cuyo valor para "voi" es 1 (sonoro).
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:
 1. 0.1285 5 tf0minnorm
 2. 0.10171 10 timaxnorm
 3. 0.08949 4 f0min
 4. 0.08505 3 tf0maxnorm

5.	0.08232	12	timinnorm
6.	0.07899	2	f0max
7.	0.07425	7	f0mean
8.	0.01848	8	f0desv
9.	0.01444	15	idesv
10.	0.00822	11	imin
11.	0.006	6	f0rang
12.	0.00483	14	imean
13.	0.0	17	sex
14.	0.0	9	imax
15.	0.0	13	irang
16.	0.0	1	voi

- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:
 - i) Instancias clasificadas correctamente 7182 90.9459%
 - ii) Instancias clasificadas incorrectamente 715 9.0541%
 - iii) Índice Kappa 0.402
 - iv) Error absoluto medio 0.1391
 - v) Raíz del error cuadrático medio 0.2654
 - vi) Error absoluto relativo 82.1204%
 - vii) Raíz del error cuadrático relativo 91.226%
 - viii) TP nativos 0.398
 - ix) FP nativos 0.038
 - x) TP extranjeros 0.962
 - xi) FP extranjeros 0.602
 - xii) Matriz de confusión:

Clasificados como	Nativos	Extranjeros
A=Nativos	293	444
B=Extranjeros	271	9889

Tabla 25: Matriz de confusión modalidad 12 – Locutores Femeninos e Intervalos Voiced

B. LOCUTORES FEMENINOS E INTERVALOS UNVOICED

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "sex". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "last". Ponemos como "splitPoint" el valor "f" (nos da igual cual pongamos, mientras no marquemos "invertSelection" como verdadero se quedará con las instancias cuyo valor para sexo es femenino).
- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "voi". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "first". Ponemos como

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

"splitPoint" el valor "0.5" y marcamos "invertSelection" como verdadero. En este caso se quedará con las instancias cuyo valor para "voi" sea 0 (intervalo sordo).

- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:

1.	0.04331	12	timinnorm
2.	0.03221	10	timaxnorm
3.	0.02153	11	imin
4.	0.00893	8	f0desv
5.	0.00545	14	imean
6.	0.00313	15	idesv
7.	0.00235	13	irang
8.	0.00142	3	tf0maxnorm
9.	0.00142	5	tf0minnorm
10.	0.0	2	f0max
11.	0.0	4	f0min
12.	0.0	17	sex
13.	0.0	6	f0rang
14.	0.0	7	f0mean
15.	0.0	9	imax
16.	0.0	1	voi

- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:

i)	Instancias clasificadas correctamente	7866	90.989%
ii)	Instancias clasificadas incorrectamente	779	9.011%
iii)	Índice Kappa	0.0174	
iv)	Error absoluto medio	0.1636	
v)	Raíz del error cuadrático medio	0.286	
vi)	Error absoluto relativo	99.2764%	
vii)	Raíz del error cuadrático relativo	99.6524%	
viii)	TP nativos	0.01	
ix)	FP nativos	0.001	
x)	TP extranjeros	0.999	
xi)	FP extranjeros	0.99	
xii)	Matriz de confusión:		

Clasificados como	Nativos	Extranjeros
A=Nativos	8	775
B=Extranjeros	4	7858

Tabla 26: Matriz de confusión modalidad 12 – Locutores Femeninos e Intervalos Unvoiced

C. LOCUTORES MASCULINOS E INTERVALOS VOICED

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "sex". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "last". Ponemos como "splitPoint" el valor "f" y marcamos "invertSelection" como verdadero. En este caso se quedará con las instancias cuyo valor para sexo sea masculino.
- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "voi". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "first". Ponemos como "splitPoint" el valor "0.5", de modo que mientras el parámetro "invertSelection" se quedará con las instancias cuyo valor para "voi" es 1 (sonoro).
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:

1.	0.395	8	f0desv
2.	0.3857	5	tf0minnorm
3.	0.3241	10	timaxnorm
4.	0.3129	15	idesv
5.	0.2846	3	tf0maxnorm
6.	0.2566	12	timinnorm
7.	0.2305	7	f0mean
8.	0.2245	14	imean
9.	0.2032	4	f0min
10.	0.1787	9	imax
11.	0.1258	13	irang
12.	0.1257	2	f0max
13.	0.1135	6	f0rang
14.	0.0341	11	imin
15.	0.0	17	sex
16.	0.0	1	voi
- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:

i)	Instancias clasificadas correctamente	1737	90.6103%
ii)	Instancias clasificadas incorrectamente	180	9.3897%
iii)	Índice Kappa	0.7997	
iv)	Error absoluto medio	0.1348	
v)	Raíz del error cuadrático medio	0.2855	
vi)	Error absoluto relativo	28.9803%	
vii)	Raíz del error cuadrático relativo	59.2066%	
viii)	TP nativos	0.892	
ix)	FP nativos	0.086	
x)	TP extranjeros	0.914	
xi)	FP extranjeros	0.108	
xii)	Matriz de confusión:		

Clasificados como	Nativos	Extranjeros
-------------------	---------	-------------

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

A=Nativos	629	76
B=Extranjeros	104	1108

Tabla 27: Matriz de confusión modalidad 12 – Locutores Masculinos e Intervalos Voiced

D. LOCUTORES MASCULINOS E INTERVALOS UNVOICED

- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "sex". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "last". Ponemos como "splitPoint" el valor "f" y marcamos "invertSelection" como verdadero. En este caso se quedará con las instancias cuyo valor para sexo sea masculino.
- Se utiliza el filtro de instancias no supervisado "RemoveWithValues". En los parámetros "attributeIndices" y "nominalIndices" se pone el índice del atributo "voi". Si hemos eliminado los atributos que hemos desestimado correctamente, este índice debería ser "first". Ponemos como "splitPoint" el valor "0.5" y marcamos "invertSelection" como verdadero. En este caso se quedará con las instancias cuyo valor para "voi" sea 0 (intervalo sordo).
- Realizamos la selección de atributos poniendo como clasificador el atributo "nat". Obtenemos el siguiente ranking de atributos:
 1. 0.2099225 8 f0desv
 2. 0.1731165 12 timinnorm
 3. 0.1314893 10 timaxnorm
 4. 0.0714221 15 idesv
 5. 0.0286583 11 imin
 6. 0.0193973 13 irang
 7. 0.0184887 9 imax
 8. 0.0082767 14 imean
 9. 0.0000345 3 tf0maxnorm
 10. 0.0000345 5 tf0minnorm
 11. 0.0 4 f0min
 12. 0.0 2 f0max
 13. 0.0 17 sex
 14. 0.0 6 f0rang
 15. 0.0 7 f0mean
 16. 0.0 1 voi
- Evaluamos el modelo en la pestaña "Classify". Obtenemos los siguientes resultados:
 - i) Instancias clasificadas correctamente 1739 83.5255%
 - ii) Instancias clasificadas incorrectamente 343 16.4745%
 - iii) Índice Kappa 0.6372
 - iv) Error absoluto medio 0.2265
 - v) Raíz del error cuadrático medio 0.369
 - vi) Error absoluto relativo 49.1001%

vii) Raíz del error cuadrático relativo	76.8447%
viii) TP nativos	0.736
ix) FP nativos	0.109
x) TP extranjeros	0.891
xi) FP extranjeros	0.264
xii) Matriz de confusión:	

Clasificados como	Nativos	Extranjeros
A=Nativos	553	198
B=Extranjeros	145	1186

Tabla 28: Matriz de confusión modalidad 12 – Locutores Masculinos e Intervalos Unvoiced

5.4.10. Repetición del cálculo de distancias con el algoritmo DTW

Después de haber analizado con WEKA los datos procedentes de las locuciones de las frases obtenidos con Praat y Snack, hemos deducido una serie de ponderaciones para nuestro vector de características, que aplicaremos en el cálculo de la matriz de distancias al ejecutar el algoritmo DTW.

Se ha modificado el código de dtw.c incluyendo la definición de todos los vectores de pesos deducidos por cada modalidad de experimento llevada a cabo, asignando un índice a cada una, el cual utilizaremos como identificador en los ficheros con los datos de las distancias para identificar con qué vectores de pesos se han calculado. Además, se ha añadido a la función dtw dos parámetros, uno para indicar la modalidad y otro para indicar el sexo. No podemos conocer directamente de los datos que leemos de los ficheros si el locutor es de género femenino o masculino, por lo que necesitamos recibir esta información como parámetro desde la función main() que invoca a dtw() en dist.c. Al recorrer cada uno de los locutores, tan sólo sabiendo su índice o leyendo la primera letra de su identificador podremos inferir su sexo.

El programa dist.c también ha sido modificado, añadiendo un bucle exterior que haga una iteración por cada modalidad. De este modo, se crearían tantos ficheros de datos con distancias como modalidades de prueba hemos realizado en la etapa anterior, cada uno conteniendo las distancias calculadas con el vector o los vectores de pesos deducidos de la modalidad de experimento correspondiente. La forma de compilarlo y ejecutarlo es exactamente igual a como se indicó previamente.

La siguiente etapa será comprobar qué vector de pesos funciona mejor para clasificar las locuciones en comparación con las distancias que se dan entre los locutores nativos entre sí, las distancias existentes entre las locuciones de los extranjeros entre sí, y las distancias entre ambos tipos de locutores.

Trataremos de calcular la distancia a ambos grupos, determinando la pertenencia de la locución a un conjunto o a otro por un criterio de cercanía. Si una locución está más cerca del grupo de nativos que del de extranjeros, será clasificada como "nativo" y viceversa. Según el grado de acierto que se dé en dicha clasificación, podremos decidir qué vector de pesos es el que mejor se ajusta a lo que pretendemos conseguir.

5.4.11. Prueba del módulo de evaluación con los pesos deducidos en las pruebas y elección de la ponderación más adecuada

Una vez ya tenemos los ficheros con las distancias existentes entre todas las frases y repeticiones, debemos decidir qué pesos, de los que hemos deducido a partir de las pruebas que hemos llevado a cabo con WEKA, son los que hacen que el sistema clasifique con más cierto si las locuciones son procedentes de un hablante nativo o extranjero. La situación ideal sería que las distancias de los locutores nativos con sus iguales y la de los extranjeros con sus iguales fuera lo más pequeña posible, y que las distancias entre nativos y extranjeros se maximizaran para poder distinguir ambos conjuntos fácilmente.

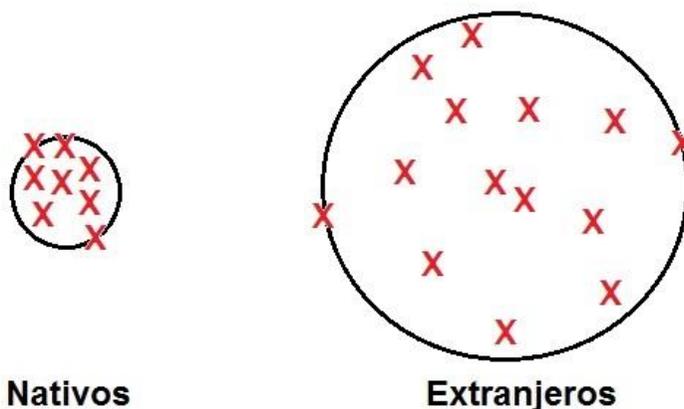


Ilustración 57: Situación ideal - dos conjuntos bien diferenciados

Sin embargo, sabemos que esta situación ideal podría no darse y que ambos conjuntos se solaparan, dificultando la clasificación, por lo que necesitamos definir los criterios que nos permitan decidir la pertenencia al grupo que corresponda.

Podríamos pensar en calcular la distancia con respecto al centroide (el promedio de todos los componentes del grupo) de ambos conjuntos para determinar la puntuación del ensayo en función de la distancia desde el intento (cuanto más cercana al centroide, más próxima al "10" de nota, por ejemplo) pero, al tratarse de series temporales, no nos sería útil. Además, al no disponer del centroide, tampoco podremos calcular un radio para el clúster a partir del cual podamos establecer la frontera para decidir si un elemento pertenece al conjunto o no. No obstante, tenemos otras alternativas para abordar el problema. La que nosotros emplearemos se basa en un criterio de proximidad. Tiene sentido pensar que una locución pertenecerá al grupo del que más cerca se encuentre; puede ocurrir que un elemento no sea igual que los que le rodean, pero se entiende que esta situación se dará en un número de casos pequeño en relación con el total.

En primer lugar, calcularemos por cada frase la distancia promedio entre los elementos de cada conjunto por separado utilizando los pesos deducidos en el análisis de los datos de grabaciones con WEKA, es decir, tendremos hasta 12 posibles promedios por cada frase, uno por cada modalidad de prueba. El diámetro de los conjuntos será el valor máximo de las distancias entre sus elementos (excluyendo infinitos). En este cálculo no consideraremos las distancias cuyo valor sea 0, infinito ni "casi infinito" por la influencia que pueden tener en el promedio, al tratarse de valores extremos. Los elementos que consideraremos para el grupo de los nativos serán las locuciones de cada frase de cada

Capítulo 5: Preparación del Corpus

uno de los ocho locutores de referencia (sin distinguir sexos), mientras que para el grupo de los extranjeros tomaremos las primeras repeticiones de cada frase de los catorce locutores extranjeros. La idea es poder probar el porcentaje de acierto en la clasificación con las repeticiones restantes, de forma que al final seleccionaremos la ponderación correspondiente a la modalidad que maximice el grado de corrección. Tenemos el inconveniente de que sólo disponemos de una única repetición de cada frase de los locutores de referencia, por lo que no podremos probar el sistema tan bien como se debería. De hecho, la ausencia de repeticiones de los locutores nativos seguramente haya influido en el análisis de WEKA y haya reducido la efectividad en la selección de atributos y la clasificación.

A continuación, por cada elemento (repetición de las frases) con el que probaremos, computamos la distancia a los elementos representativos los dos conjuntos por separado (los que indicamos previamente), y calculamos el promedio con respecto ambos grupos. Si la distancia promedio del elemento que probamos con respecto a los que hemos tomado como referencia de un grupo es menor que la distancia promedio entre los mismos elementos de ese mismo grupo, entonces ese elemento pertenece a ese conjunto. Razonamos igual para el otro conjunto, pudiendo darse la circunstancia de que un elemento pueda pertenecer al conjunto de nativos y al de extranjeros a la vez. Esto querrá decir que ambos conjuntos se solapan y que existe una zona de incertidumbre en la que el pronóstico de la clasificación puede ser erróneo. La clasificación de la entonación de la frase como nativa o extranjera dependerá de si la distancia promedio al grupo de los nativos es menor que la distancia promedio al conjunto de los extranjeros, y viceversa.

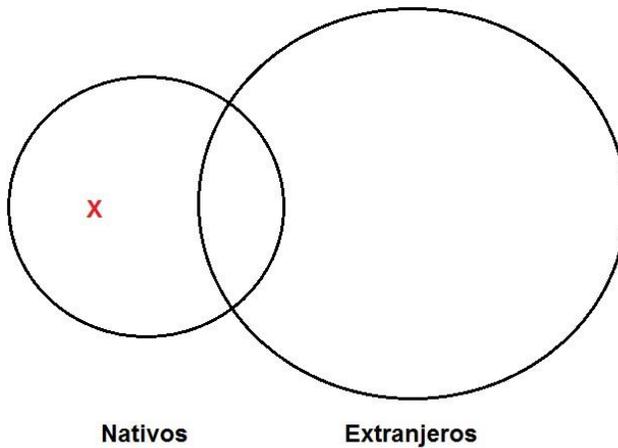


Ilustración 58: Locución pertenece al conjunto de los nativos

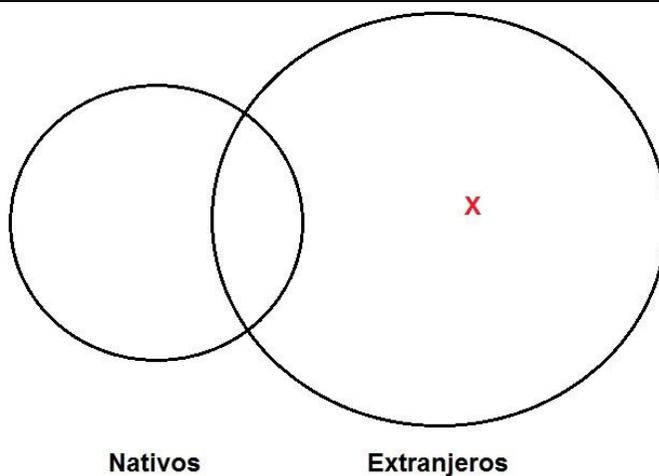


Ilustración 59: Locución pertenece al conjunto de los extranjeros

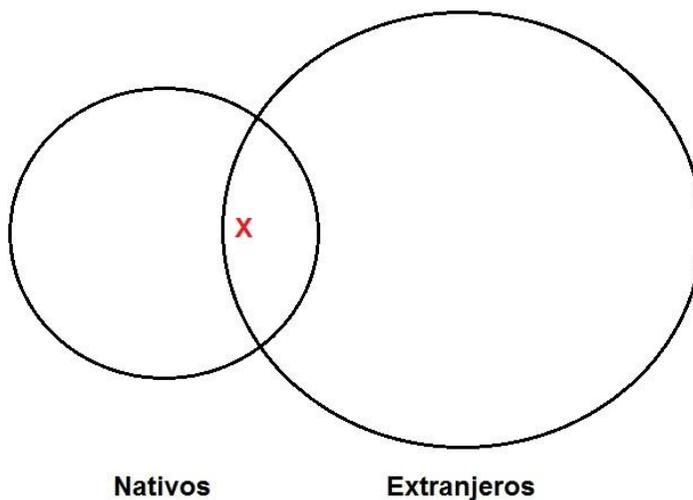


Ilustración 60: Locución perteneciente a ambos conjuntos

Obviamente, también puede suceder que un elemento que probemos no pertenezca ni al conjunto de nativos ni al conjunto de extranjeros. En estos casos, se clasificará al elemento dentro del grupo que tenga más próximo, como en el caso anterior.

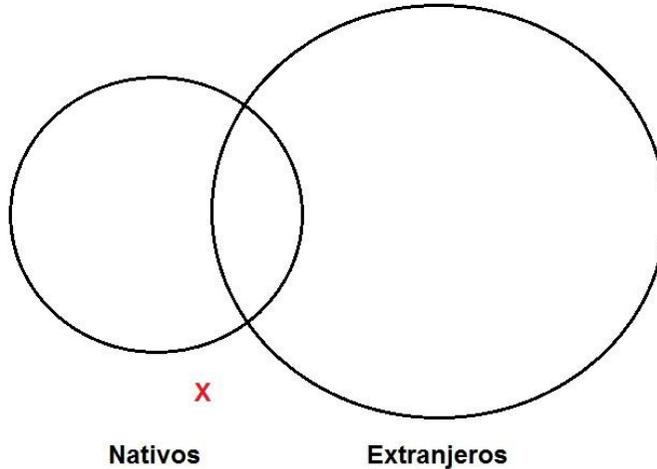


Ilustración 61: Locución no pertenece a ninguno de los dos conjuntos

En cuanto a la asignación de una puntuación a un ensayo, podríamos hacerlo en función de un locutor considerado "Golden Speaker" escogido como referencia por parte de especialistas; sin embargo como no hemos consultado a ninguno, escogeremos uno al azar entre los ocho de los que disponemos (también podríamos escoger el que más se aproximara a un supuesto centroide del conjunto de nativos).

Después de haber explicado la teoría de cómo pretendemos solucionar el problema, ya podemos exponer cómo lo hemos resuelto en la práctica. Como hemos comentado, el primer paso se trataba de calcular las distancias promedio entre los elementos tomados como referencia de cada conjunto por separado (la única repetición de cada frase de todos los locutores en el caso de los nativos, y las primeras repeticiones de cada frase de todos los locutores para los extranjeros). Con tal motivo, se ha desarrollado el programa `dist_prom.c` en el directorio `/home/p/pfc/dvall/pcompare/corpus/scripts/pitch`, el cual lee el fichero `all_[modalidad].data`, que contiene las distancias entre todas las locuciones para la modalidad `[modalidad]`, y lee los elementos pertinentes para calcular el promedio de las distancias y las imprime de forma etiquetada en un fichero de texto, `dist_comp.txt`. También se ha implementado otro programa, `dist_prom_data.c` en la misma carpeta que realiza la misma función pero escribe los promedios en bruto (sin etiquetas) en el fichero `dist_prom.data`.

Posteriormente, se ha realizado la prueba de cuáles son los pesos que hacen que el sistema califique con un mayor porcentaje de acierto con el programa `class_comp.c` en el mismo directorio `/home/p/pfc/dvall/pcompare/corpus/scripts/pitch`. Ese programa indica si el elemento probado (repetición de una frase), está dentro del conjunto de extranjeros, si está dentro del grupo de los nativos, y lo clasifica dentro de uno u otro, según su proximidad. Podríamos haber obrado igual que con el programa anterior leyendo los datos del fichero `all_[modalidad].data`, pero ya pensando en la rutina que se debe ejecutar cuando integremos el módulo en un servidor y las grabaciones lleguen a nuestro sistema, se ha optado por desarrollarlo de manera simule lo que se ejecutará en ese momento. Para cada modalidad, se calculan las distancias a los puntos que tomamos como referencia, y según éstas, decidimos la pertenencia o no a ambos conjuntos y clasificamos el intento como nativo o extranjero. Después de haber procesado todos los elementos de prueba, calculamos el porcentaje de aciertos de la clasificación realizada con los pesos deducidos de cada modalidad de prueba realizada durante el análisis de los datos con WEKA, teniendo en cuenta que todos los elementos son extranjeros e imprimimos un resumen con el que podremos contrastar los datos. Todo ello además queda guardado en un fichero de texto `class_comp.txt`.

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

Los porcentajes de acierto en la clasificación resultantes de la ejecución de estos programas han sido los siguientes:

Total: 630 locuciones de prueba			
Modalidad	Clasificados Extranjeros	Clasificados Nativos	Porcentaje de Acierto
Modalidad 1	249	381	39.523810
Modalidad 2	272	358	43.174603
Modalidad 3	247	383	39.206349
Modalidad 4	251	379	39.841270
Modalidad 5	410	220	65.079365
Modalidad 6	412	218	65.396825
Modalidad 7	315	315	50.000000
Modalidad 8	341	289	54.126984
Modalidad 9	308	322	48.888889
Modalidad 10	323	307	51.269841
Modalidad 11	476	154	75.555556
Modalidad 12	485	145	76.984127

Tabla 29: Resultados de las pruebas del sistema: Porcentajes de Acierto en la clasificación de las locuciones como nativo o extranjero

En vista de estos resultados, el sistema tiene un porcentaje de acierto mayor (aproximadamente un 77%) en la clasificación con los datos extraídos de las grabaciones de las locuciones de las frases con Snack, aplicando para el cálculo de la matriz de distancias que sirve de entrada al algoritmo DTW los pesos deducidos del análisis del fichero de los datos locales distinguiendo intervalos sonoros y sordos y distinguiendo sexos con WEKA.

Por tanto, para el módulo de evaluación de la entonación utilizaremos la herramienta Snack para calcular los valores que caractericen las nuevas grabaciones que lleguen al sistema. Se calculará la distancia promedio a los elementos del conjunto de nativos y del conjunto de extranjeros que se utilizaron durante la prueba de corrección utilizando el algoritmo DTW, el cual recibirá como entrada una matriz de distancias cuyos valores serán ponderados con los pesos deducidos del análisis del fichero de datos locales distinguiendo intervalos sonoros y sordos y distinguiendo sexos con WEKA. Finalmente, la distancia promedio a ambos conjuntos será comparada con las distancias promedio obtenidas de la misma forma con el fin de clasificar el ensayo como nativo o no nativo, y de comprobar su pertenencia a al grupo de los nativos, al de los extranjeros o a ambos.

5.4.12. *Sistema de Puntuación (Scoring)*

El último paso que nos queda para definir completamente nuestro módulo de evaluación de la entonación es decidir qué criterio vamos a seguir para asignar una puntuación (“score”) a los ensayos realizados por los usuarios. Como los usuarios potenciales de nuestro sistema serán personas extranjeras que deseen mejorar sus aptitudes con el español, parece lógico pensar que esperen los resultados en comparación con las locuciones de un usuario nativo. Por ello, y puesto que ya tenemos calculadas las distancias promedio entre los nativos, la tomaremos como el diámetro. Nosotros consideraremos el radio como la distancia límite para un “aprobado”, esto es, si la distancia promedio del ensayo de un usuario (d_N) a las locuciones que tenemos como referencia es exactamente igual que la distancia promedio entre estas locuciones, el intento será calificado con un 5. Si la distancia fuera 0 (lo cual parece imposible), el intento obtendría una puntuación de 10. Para los ensayos que “suspenden”, se les asignará una calificación de 0 a los que estén alejados de los locutores de referencia a una de infinito (nosotros asignaremos un valor lo suficientemente grande para facilitar los cálculos). Siguiendo estas indicaciones, la nota seguiría esta función:

$$\begin{aligned} S(d_N) &= 10, \text{ si } d_N = 0; \\ S(d_N) &= 5, \text{ si } d_N = \text{radio}; \\ S(d_N) &= 0, \text{ si } d_N = \infty \end{aligned}$$

Por consiguiente, la función que utilizaremos para asignar una nota será la siguiente:

$$S(d_N) = \frac{1}{1 + \frac{d_N}{\text{radio}}} \cdot 10$$

Aplicación Final

Hasta ahora hemos trabajado con los datos procedentes del corpus para tomar decisiones de cara al diseño del módulo de evaluación de entonación que queremos desarrollar. Una vez ya tenemos claro qué herramienta vamos a utilizar y qué parámetros son los que vamos a comparar, implementaremos el módulo de manera que evalúe una entonación en función de las decisiones que hemos tomado durante la fase previa. Finalmente, integraremos nuestro módulo de evaluación de la entonación en una aplicación de modo que podamos ilustrar su funcionamiento.

La finalidad de esta aplicación no será otra que ayudar al aprendizaje de la segunda lengua mediante la mejora de la pronunciación y la entonación. Como ya apuntamos, las tecnologías del habla resultan interesantes para esta tarea puesto que pueden complementar el estudio en el aula. Así pues, nuestra aplicación UVaPronunciación estará compuesta por un conjunto de ejercicios basado en la audición y repetición de las frases que hemos considerado en el corpus. A continuación presentaremos la especificación técnica de sus características.

6.1. Análisis

6.1.1. *Requisitos del sistema*

6.1.1.1. **Requisitos funcionales**

- Rq. 1. La aplicación deberá reproducir en voz alta cadenas de texto.
- Rq. 2. La aplicación deberá reconocer las palabras habladas por un usuario.
- Rq. 3. La aplicación deberá permitir conservar la grabación del ensayo después del reconocimiento.
- Rq. 4. La aplicación deberá calificar la pronunciación del usuario.
- Rq. 5. La aplicación deberá extraer los parámetros de duración, frecuencia fundamental y energía de los ensayos y guardarlos: tiempo, número de muestras, máximos y mínimos de frecuencia y energía, tiempo en el que suceden los máximos y mínimos, tiempo normalizado en que se dan los máximos y los mínimos, rangos de frecuencia y energía, promedios de frecuencia y energía y desviación típica de frecuencia y energía; tanto para el global de la grabación, como para cada una de las series temporales (intervalos sonoros y sordos) identificadas dentro de la misma.
- Rq. 6. La aplicación deberá calificar la entonación del usuario.
- Rq. 7. La aplicación deberá guardar la información referente a los ensayos: usuario que la realizó, frase que ensayó, fecha del ensayo, lugar en el que se almacena, puntuación de la pronunciación, palabras reconocidas para la evaluación de la pronunciación, puntuación de la entonación, distancia considerada para evaluar la entonación, clasificación como nativo o no nativo del ensayo y si el usuario realizó la audición previamente.
- Rq. 8. La aplicación deberá permitir saber al usuario la información de su ensayo: puntuación de la pronunciación, palabras reconocidas para la evaluación de la pronunciación, puntuación de la entonación, clasificación como nativo o no nativo del ensayo y nota global.
- Rq. 9. La aplicación deberá permitir el registro de varios usuarios.

6.1.1.2. Requisitos No Funcionales

-Facilidad de uso

- Rq-NF. 1. La aplicación deberá poder ser manejada en base a Activities, botones y cuadros de texto.
- Rq-NF. 2. La aplicación deberá poder ser utilizada correctamente por cualquier miembro de la comunidad universitaria sin necesidad de un curso de preparación.

-Restricciones de diseño

- Rq-NF. 3. La el cliente móvil de la aplicación deberá estar desarrollado según las directrices de la arquitectura Android.
- Rq-NF. 4. El cliente móvil de la aplicación deberá estar desarrollado en lenguaje Java.
- Rq-NF. 5. El componente servidor de la aplicación estará albergado en el servidor "aries" del grupo ECA-SIMM.
- Rq-NF. 6. El componente servidor de la aplicación deberá ser un servidor web HTTP Apache.
- Rq-NF. 7. Las rutinas encargadas de la gestión de usuarios y cargas y descargas de archivos de audio en el servidor deberán estar desarrolladas en PHP.
- Rq-NF. 8. El componente servidor albergará sistema de gestión de bases de datos MySQL.
- Rq-NF. 9. La cliente móvil de la aplicación deberá integrar un conjunto de frases que leerá de un fichero con extensión ".txt".
- Rq-NF. 10. El cliente móvil de la aplicación deberá descargar los audios correspondientes a las frases que ha leído en el fichero de texto del componente servidor de la aplicación.
- Rq-NF. 11. El cliente móvil de la aplicación deberá reproducir los archivos de audio descargados desde el componente servidor.
- Rq-NF. 12. El cliente móvil de la aplicación deberá mostrar las frases que componen el ejercicio al usuario siguiendo un orden secuencial o aleatorio.
- Rq-NF. 13. El cliente móvil de la aplicación deberá subir al servidor las grabaciones correspondientes a las locuciones de los ensayos realizados por los usuarios.
- Rq-NF. 14. La aplicación extraerá los valores que caractericen la locución del ensayo utilizando la herramienta Snack.
- Rq-NF. 15. La aplicación puntuará la pronunciación de una frase por parte del usuario con una calificación entre 0 y 10.
- Rq-NF. 16. La aplicación puntuará la pronunciación de una frase mediante un criterio basado en el cálculo de la distancia de Levenshtein.
- Rq-NF. 17. La aplicación deberá utilizar el servicio de reconocimiento de voz proporcionado por Google Voice.
- Rq-NF. 18. La aplicación puntuará la entonación de una frase por parte del usuario con una calificación entre 0 y 10.
- Rq-NF. 19. La aplicación calificará un ensayo como nativo o extranjero.
- Rq-NF. 20. La aplicación deberá crear un directorio para cada usuario que se registre en el componente servidor.
- Rq-NF. 21. La aplicación creará un directorio para cada ensayo dentro del directorio del usuario que lo realizó.
- Rq-NF. 22. La aplicación deberá almacenar la grabación del ensayo y los ficheros de datos que se generen a partir de ella en el directorio que se creó a raíz de su realización en el componente servidor.

- Rq-NF. 23. La aplicación deberá generar por cada ensayo ficheros con formato ".ene", ".f0", ".f0a", ".spar" con las características descritas en la sección "Ficheros de datos" del apartado "Contenidos" del capítulo anterior.
- Rq-NF. 24. La aplicación deberá generar dos ficheros codificados en formato JSON con los datos del reconocimiento de voz y los de la entonación. Los ficheros tendrán formato ".ret" y ".json", respectivamente.

-Interfaces

a) Interfaces de usuario

- Rq-NF. 25. La aplicación deberá ser manejada en dispositivos con pantalla táctil.
- Rq-NF. 26. La aplicación deberá permitir la introducción de datos a través de la entrada de voz para el reconocimiento de habla.
- Rq-NF. 27. La aplicación deberá permitir al usuario controlar el volumen de los contenidos auditivos que escuchará.

b) Interfaces software

- Rq-NF. 28. La aplicación deberá poder ser ejecutada en cualquier dispositivo móvil que integre el sistema operativo Android con versión 2.2 o superior.

c) Interfaces de comunicación

- Rq-NF. 29. El cliente móvil de la aplicación deberá comunicarse con el componente servidor mediante peticiones HTTP.
- Rq-NF. 30. La aplicación deberá tener permiso para acceder a internet a través de una red de datos HSPA o Wi-Fi.

6.1.2. Modelo de Casos de Uso

6.1.2.1. Descripción General de los Actores

ACT-0001	Usuario
Descripción	Este actor representa a la persona que va a manejar la aplicación.

Tabla 30: Actores que intervienen en los Casos de Uso

A pesar de que podemos identificar hasta cuatro casos de uso, nos centraremos exclusivamente en el caso de estudio "Hacer Ejercicio", que es el que realmente nos interesa para este trabajo. En él es donde se encuentra la funcionalidad principal de la aplicación.

6.1.2.2. Diagrama del Modelo de Casos de Uso

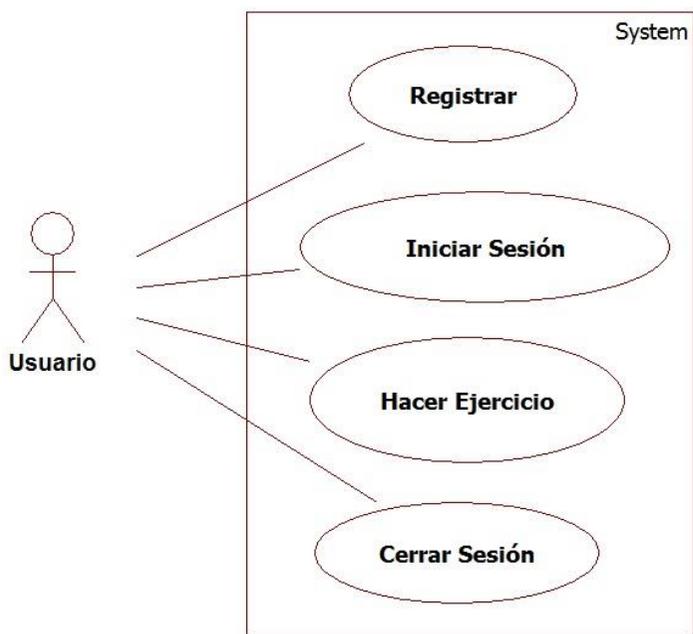


Ilustración 62: Modelo de Casos de Uso

6.1.2.3. Descripción de los Casos de Uso

UC-0001		Hacer Ejercicio
Descripción	El sistema deberá comportarse tal como se describe en el siguiente caso de uso cuando el actor Usuario (ACT-0001) desee realizar un ejercicio.	
Precondición	El actor Usuario (ACT-0001) está registrado en aplicación y ha iniciado sus sesión. Existen frases cargadas en el sistema.	
Postcondición	Los datos referentes a los intentos realizados por el actor Usuario (ACT-0001) sobre las frases correspondientes quedan guardados.	
Secuencia normal	Paso	Acción
	1	El caso de uso comienza cuando el actor Usuario (ACT-0001) selecciona la opción "Ir a Ejercicio".
	2	El sistema muestra una frase.
	3	El actor Usuario (ACT-0001) selecciona "Escuchar".
	4	El sistema reproduce en voz alta la frase mostrada.
	5	El actor Usuario (ACT-0001) selecciona "Intentar".
	6	El actor Usuario (ACT-0001) pronuncia las palabras de la frase mostrada.
	7	El sistema reconoce las palabras habladas por el actor Usuario (ACT-0001) y analiza su entonación.
	8	El sistema muestra las puntuaciones del último intento.
9	El actor Usuario (ACT-0001) selecciona "Volver".	

	10	El sistema guarda el ensayo y sus datos y el caso de uso finaliza.
Flujo alternativo	Paso	Acción
	3	Si el actor Usuario (ACT-0001) selecciona "Siguiete", el caso de uso vuelve al paso 2 y continua. Flujo alternativo: Si el actor Usuario (ACT-0001) selecciona "Intentar", el caso de uso continua en el paso 5. Flujo alternativo: Si el actor Usuario (ACT-0001) selecciona "Volver" el caso de uso queda sin efecto.
	5	Si el actor Usuario (ACT-0001) selecciona "Escuchar", el caso de uso vuelve al paso 4 y continua. Flujo alternativo: Si el actor Usuario (ACT-0001) selecciona "Siguiete", el caso de uso vuelve al paso 2 y continua. Flujo alternativo: Si el actor Usuario (ACT-0001) selecciona "Volver" el caso de uso queda sin efecto.
	9	Si el actor Usuario (ACT-0001) selecciona "Siguiete", el caso de uso vuelve al paso 2 y continua. Flujo alternativo Si el actor Usuario (ACT-0001) selecciona "Escuchar", el caso de uso vuelve al paso 4 y continua. Flujo alternativo: Si el actor Usuario (ACT-0001) selecciona "Intentar", el caso de uso vuelve al paso 6 y continua.
Comentarios	Rq.1, Rq.2, Rq.3, Rq.4, Rq.5, Rq.6, Rq.7 ,Rq.8.	

Tabla 31: Caso de Uso - Hacer Ejercicio (Análisis)

6.1.3. Realización de los casos de Uso

6.1.3.1. Diagrama de Clases

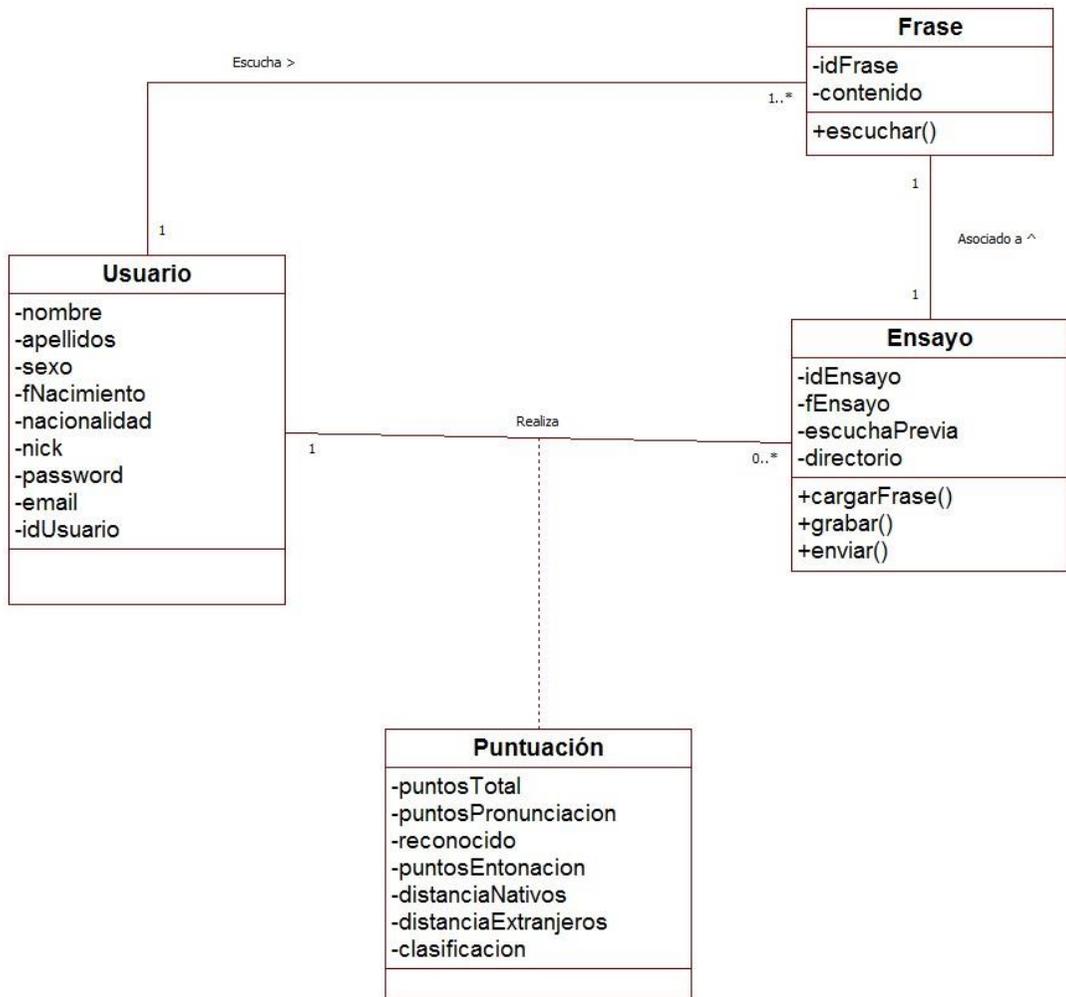


Ilustración 63: Diagrama de Clases de Análisis

6.1.3.2. Descripción de las Clases

Usuario	
Responsabilidad	Encapsula los datos relacionados con el usuario guardados en el sistema.
Atributos	-nombre: Nombre del usuario -apellidos: Apellidos del usuario. -sexo: Sexo del usuario (hombre o mujer). -fNacimiento: Fecha de nacimiento del usuario.

	-nacionalidad: País de origen del usuario. -nick: Nombre con el que el usuario se identifica en el sistema. -password: Contraseña del usuario. -email: Dirección de correo electrónico del usuario. -idUserio: Identificador del usuario en el sistema.
Operaciones	

Tabla 32: Descripción de la clase Usuario

Frase	
Responsabilidad	Encapsula los datos relacionados con las frases incluidas en la aplicación.
Atributos	-idFrase: Identificador de la frase en el sistema. -contenido: Texto de la frase.
Operaciones	+escuchar(): Reproducción en voz alta de la frase.

Tabla 33: Descripción de la clase Frase

Ensayo	
Responsabilidad	Encapsula los datos relacionados con los ensayos realizados por los usuarios de la aplicación.
Atributos	-idEnsayo: Identificador del ensayo en el sistema. -fEnsayo: Fecha en la que se realiza el ensayo. -escuchaPrevia: Si el usuario ha escuchado previamente la frase o no. -directorio: carpeta en la que se aloja en el sistema el ensayo y sus datos.
Operaciones	+cargarFrase: Cargar la frase del ejercicio. +grabar(): Grabación del ensayo por parte del usuario. +enviar(): Enviar la grabación del ensayo para su evaluación.

Tabla 34: Descripción de la clase Ensayo

Puntuación	
Responsabilidad	Encapsula los datos relacionados con los resultados de los ensayos realizados por los usuarios de la aplicación.
Atributos	-puntosTotal: Puntuación global del ensayo teniendo en cuenta pronunciación y entonación. -puntosPronunciacion: Puntuación del ensayo considerando sólo la pronunciación. -reconocido: Palabras reconocidas del ensayo del usuario. -puntosEntonacion: Puntuación del ensayo considerando sólo la entonación. -distanciaNativos: Distancia promedio del ensayo con respecto al conjunto de los nativos. -distanciaExtranjeros: Distancia promedio del ensayo con respecto al conjunto de los extranjeros. -clasificacion: Clasificación del ensayo como nativo o extranjero.
Operaciones	

Tabla 35: Descripción de la clase Puntuación

6.1.3.3. Diagramas de Secuencia

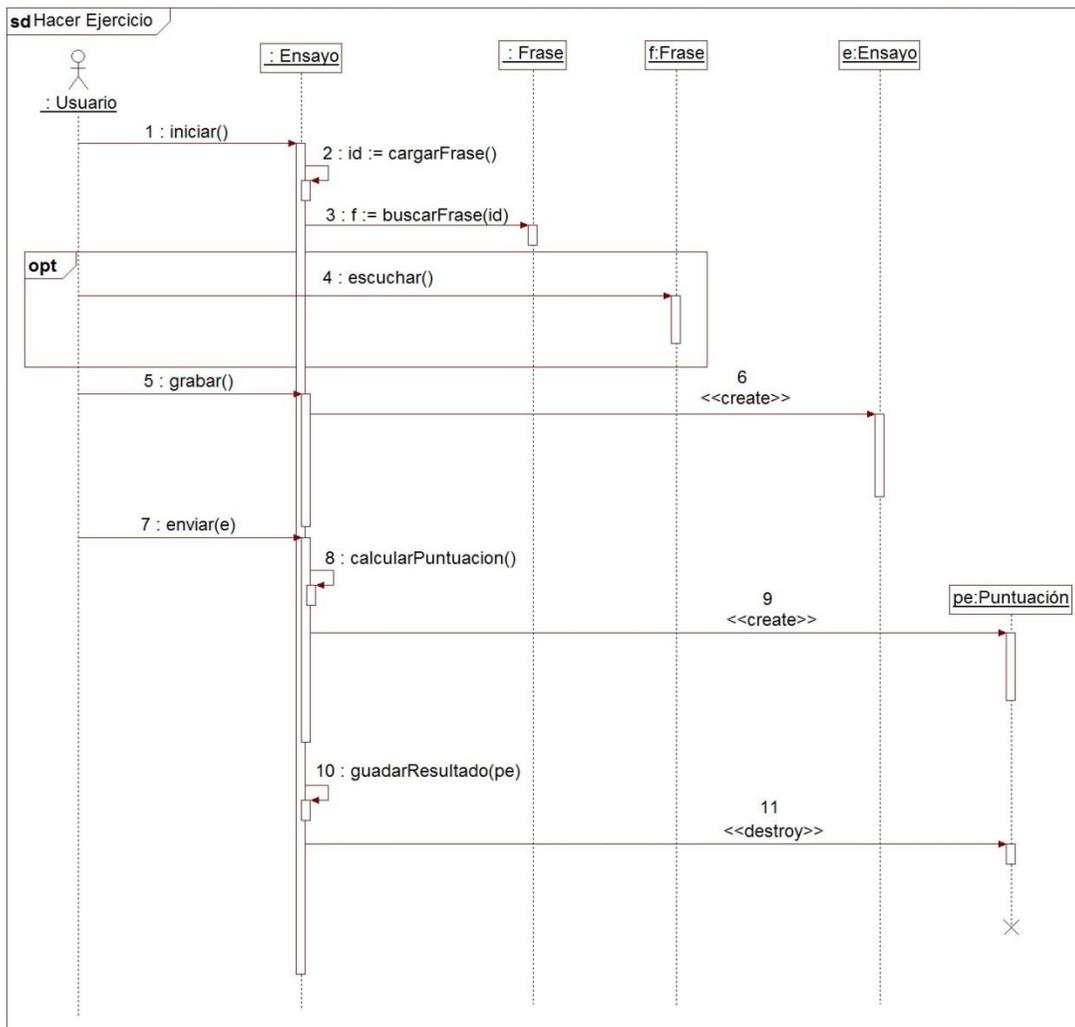


Ilustración 64: Diagrama de secuencia Hacer Ejercicio - Análisis

6.2. Diseño

6.2.1. Arquitectura propuesta

6.2.1.1. Diseño de la Arquitectura

6.2.1.1.1. Topología del Sistema (Diagrama de Despliegue)

La aplicación está diseñada para su uso mediante un cliente móvil Android, el cual interactuará mediante peticiones HTTP con un servidor web que será el que realice la totalidad de los cálculos para el tratamiento de la entonación, utilizando ffmpeg y Snack. Además, los datos de los usuarios y los

ensayos quedarán registrados en una base de datos MySQL. El módulo del cliente se sirve de una base de datos SQLite local en la cual se realizan inserciones ,actualizaciones y borrados mediante sentencias SQL. El servidor, para reconocer las palabras del ensayo del usuario, hace uso de la tecnología de Google Voice, la cual reside en los servidores de Google y al que accedemos realizando peticiones HTTP a través de internet. Todos los cálculos correspondientes a la lógica del reconocimiento se efectúan en los servidores de Google. De este modo, el diagrama de despliegue sería el siguiente:

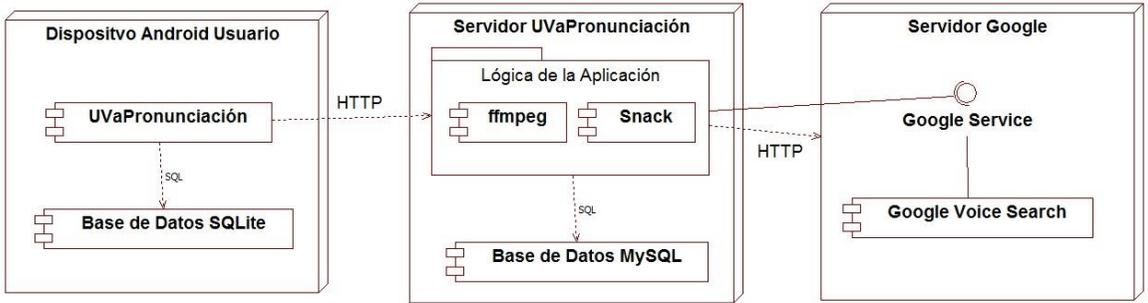


Ilustración 65: Diagrama de Despliegue del Sistema

6.2.1.2. Visión Global

Las tareas principales del sistema son la carga y descarga de archivos de audio, el tratamiento de las grabaciones mediante extracción de parámetros típicos para la caracterización del habla y la transmisión de los resultados al usuario. Por ello, la arquitectura propuesta se basará en un patrón de tres capas: Presentación, Lógica de la Aplicación y Persistencia de Datos. El sistema de capas será relajado, la capa superior podrá acceder a su capa inmediatamente inferior o a la más baja. Como utilizamos tecnología Android, la cual se fundamenta en Java, necesitaremos una capa más en la que se encuentren los paquetes y las clases importadas que se emplean para el correcto funcionamiento de la aplicación. También incluiremos en esta nueva capa la biblioteca que hace que podamos utilizar Snack en el sistema. A esta capa la denominaremos Bibliotecas:

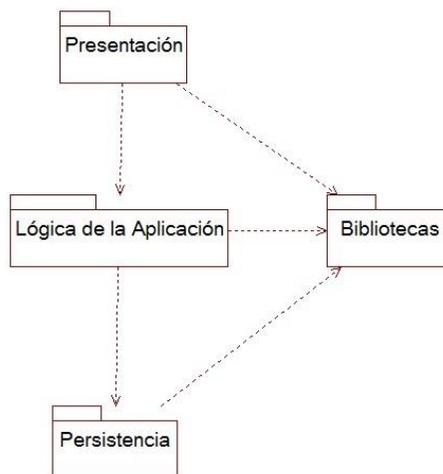


Ilustración 66: Arquitectura del sistema (3 capas + Bibliotecas)

6.2.1.2.1. Diseño de las capas del Sistema

a) Capa de Presentación

La capa de Presentación contiene los elementos necesarios para que el usuario puede interactuar con la aplicación. Los controles de esta capa recogen los eventos disparados por el usuario y la información que éste introduce para que el programa pueda enviarla al servidor, donde será tratada hasta emitir de vuelta al cliente móvil los resultados pertinentes, los cuales también serán mostrados por los componentes de esta capa. Aquí encontraremos las Activities y los controles definidos en sus interfaces.

b) Capa de Lógica de la Aplicación

Esta capa es la que recibe las peticiones del usuario y la que se encarga de procesar la información. Se comunica con la capa de Presentación recogiendo las solicitudes de ejecución de las operaciones correspondientes y efectuándolas para devolver los resultados de las acciones realizadas.

A su vez, esta capa contiene cuatro paquetes: Gestión de Usuarios, Pronunciación, Entonación y Recepción de Ensayos. El primero contiene todo lo necesario para gestionar los registros de usuarios y su identificación en el sistema al iniciar sesión; el segundo y el tercero se ocuparán del análisis del ensayo desde el punto de vista de la pronunciación y la entonación, respectivamente. Estos dos paquetes, a su vez, dependen del cuarto, que se ocupa de la recepción de las grabaciones, el almacenamiento de las mismas en las carpetas adecuadas dentro del sistema, y de su conversión en formatos que puedan ser entendido por los servidores que darán soporte al análisis de la pronunciación y la entonación.

c) Capa de Persistencia

Ésta es la capa en la que residen los datos que se almacenan en el sistema y donde se encuentran los elementos necesarios para acceder a ellos y recuperarlos. El almacenamiento de los datos en el dispositivo móvil se realizará en una base de datos SQLite. Para facilitar la tarea de introducción, extracción y actualización de información en la base de datos se ha implementado un patrón DAO, lo cual permite reducir el acoplamiento y habilitar una interfaz única para tratar con la base. Por otra parte, en el servidor, utilizaremos un esquema similar, pero esta vez con una base de datos MySQL.

d) Bibliotecas

La capa adicional de Bibliotecas incluye los paquetes y clases que importamos al principio del código de nuestras clases, los cuales son necesarios para implementar muchos componentes y que la aplicación se comporte de la manera esperada. Los paquetes que utilizamos en nuestra aplicación son los siguientes:

1. *java.io*: proporciona todas las operaciones de entrada/salida. Incorpora interfaces, clases y excepciones para acceder a todo tipo de ficheros.
2. *java.net*: proporciona las clases para implementar aplicaciones de servicios en la red.
3. *java.text*: proporciona clases de apoyo adicionales para la internacionalización de las aplicaciones.

4. *java.util*: proporciona una clases genéricas de colecciones (Dictionary, Vector, Hashable..).
5. *org.apache.http*: Interfaces y clases de los componentes HTTP. Tratan con los aspectos fundamentales requeridos para utilizar el protocolo HTTP, como representar un mensaje incluyendo sus cabeceras y una entidad opcional, además de conexiones sobre qué mensajes se envían.
6. *org.json*: conjunto de métodos para tratar con objetos JSON.
7. *libsnack*: contiene todo el conjunto de funciones del Snack Sound Toolkit.
8. *android.app*: contiene clases de alto nivel que encapsulan el modelo de aplicación de Android por completo.
9. *android.content*: contiene clases para compartir recursos entre aplicaciones, gestionar paquetes y gestionar recursos.
10. *android.database.sqlite*: contiene las clases para gestionar una base de datos SQLite que la aplicación utilizaría para gestionar su propia base de datos.
11. *android.media*: proporciona clases que gestionan varias interfaces multimedia en audio y video.
12. *android.os*: proporciona servicios del sistema operativo, paso de mensajes y comunicación entre procesos en el dispositivo.
13. *android.view*: proporciona clases que exponen las clases básicas de interfaz de usuario que se encargan del diseño de la pantalla y la interacción con el usuario.
14. *android.widget*: contiene elementos típicos de UI para utilizar en la pantalla de la aplicación.

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

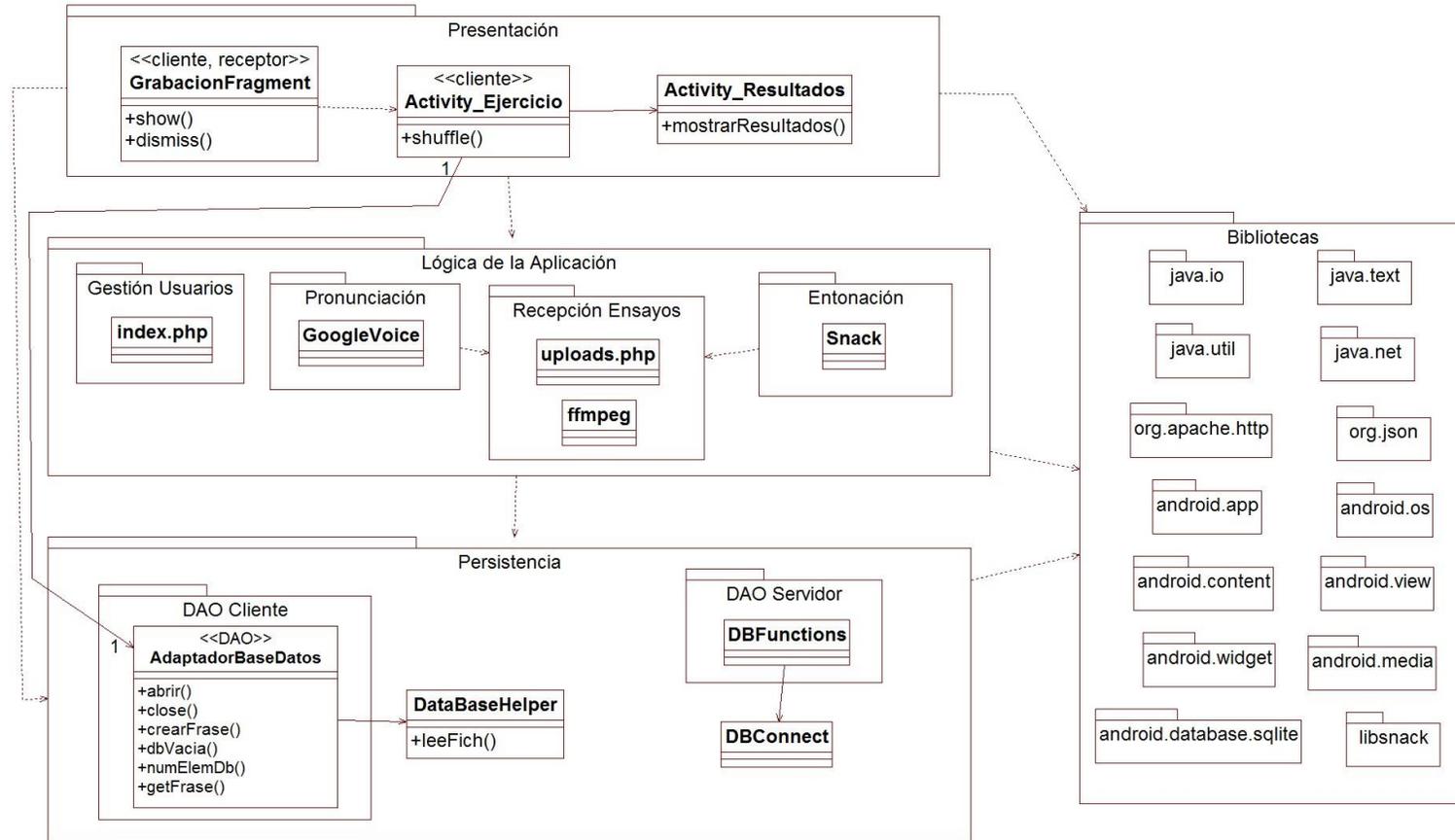


Ilustración 67: Arquitectura de capas detallado (3 capas + bibliotecas)

6.2.2. Modelo de Casos de Uso de Diseño

6.2.2.1. Descripción de los Casos de Uso

UC-0001		Hacer Ejercicio
Descripción	El sistema deberá comportarse tal como se describe en el siguiente caso de uso cuando el actor Usuario (ACT-0001) desee realizar un ejercicio.	
Precondición	El actor Usuario (ACT-0001) está registrado y ha iniciado sesión en el sistema. El actor Usuario (ACT-0001) desea realizar un ejercicio. Hay frases cargadas en el sistema. Existe una base de datos creada en el sistema.	
Postcondición	Los datos referentes a los intentos realizados por el actor Usuario (ACT-0001) sobre las frases correspondientes quedan almacenados en la base de datos.	
Secuencia normal	Paso	Acción
	1	El caso de uso comienza cuando el actor Usuario (ACT-0001) pulsa el botón "Ir al Ejercicio" en Activity_MenuPrincipal.
	2	La aplicación móvil carga la interfaz de Activity_Ejercicio.
	3	La aplicación móvil abre la base de datos.
	4	La aplicación móvil accede a una frase de la base de datos.
	5	La aplicación móvil muestra por pantalla la frase a la que ha accedido.
	6	La aplicación móvil descarga la grabación modelo correspondiente a la frase en pantalla
	7	El actor Usuario (ACT-0001) pulsa el botón "Escuchar".
	8	La aplicación móvil reproduce en voz alta la grabación de la frase mostrada en pantalla.
	9	El actor Usuario (ACT-0001) pulsa el botón "Intentar".
	10	La aplicación móvil abre un cuadro de diálogo.
	11	El usuario pulsa el botón "Rec".
	12	El usuario habla a la entrada de voz del dispositivo pronunciando las palabras de la frase mostrada en pantalla.
	13	El usuario pulsa el botón "Stop".
	14	La aplicación móvil pregunta al usuario si desea enviar la grabación del ensayo.
	15	El usuario pulsa "Enviar".
	16	La aplicación móvil envía la grabación al servidor.
	17	El servidor recibe la grabación y crea una carpeta para el ensayo en la carpeta del usuario que lo ha realizado.
	18	El servidor convierte la grabación a formato ".flac" y ".wav" con ffmpeg.
	19	El servidor envía la grabación en formato ".flac" a Google Voice para reconocer las palabras de la locución.
	20	El servidor crea un archivo con formato ".ret" con los resultados del reconocimiento de voz.
21	El servidor realiza el análisis de la entonación con Snack, generando los ficheros con formato ".f0", ".f0a", ".ene" y ".spar"	

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

		descritos en el capítulo 5 de esta memoria.
	22	El servidor calcula las puntuaciones del ensayo.
	23	El servidor genera el fichero con los resultados con formato ".json".
	24	El servidor guarda los resultados en la base de datos.
	24	El servidor devuelve los resultados de la evaluación.
	26	La aplicación móvil recibe los resultados, los muestra en la Activity_ Resultados y el caso de uso finaliza.
Flujo alternativo	Paso	Acción
	6	Si se produce un fallo en la descarga, la aplicación da la oportunidad al usuario de reintentar.
	7	Si el actor Usuario (ACT-0001) pulsa el botón "Intentar", el caso de uso continúa en el paso número 10.
	15	Si el usuario pulsa "Cancelar", el cuadro de diálogo se cierra y el caso de uso continúa en paso número 7.
	16	Si se produce un fallo al subir la grabación al servidor, la aplicación da al usuario la oportunidad de reintentar.
	17	Si el servidor no recibe la grabación, se sobrepasa el timeout, se muestra un aviso al usuario y se le permite reintentar.
	21	Si se produce un error en este paso, la aplicación móvil no recibe los resultados del ensayo y el caso de uso queda sin efecto.
	26	Si la aplicación móvil no recibe los resultados, el caso de uso queda sin efecto.
Comentarios		Rq.1, Rq.2, Rq.3, Rq.4, Rq.5, Rq.6, Rq.7 , Rq.8. Rq-NF.3, Rq-NF.9, Rq-NF.10, Rq-NF.11, Rq-NF.13, Rq-NF.14, Rq-NF.15, Rq-NF.16, Rq-NF.17, Rq-NF.18, Rq-NF.19, Rq-NF.20, Rq-NF.22, Rq-NF.23, Rq-NF.24,

Tabla 36: Caso de Uso - Hacer Ejercicio (Diseño)

6.2.3. Realización de los Casos de Uso de Diseño

6.2.3.1. Diagrama de Clases

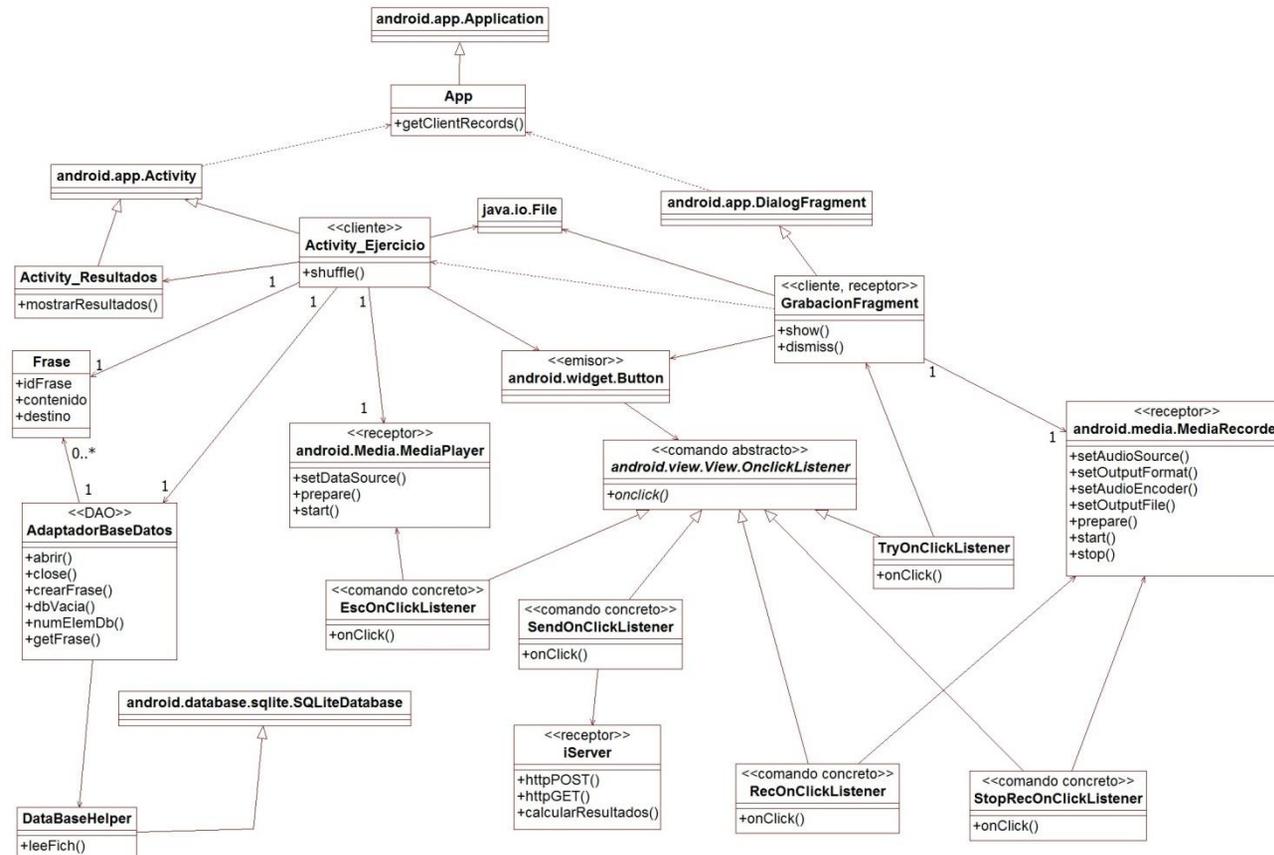


Ilustración 68: Diagrama de Clases de Diseño

6.2.3.2. Diagramas de Secuencia

Para hacer más cómoda la lectura del documento, se ha optado por dividir el único caso de uso en cuatro diagramas que podríamos juntar en uno si los ubicamos secuencialmente en el orden exposición. Los diagramas son: Cargar Frase, Escuchar Frase, Grabar Ensayo y Evaluar Ensayo.

6.2.3.2.1. Diagrama de Secuencia de Diseño 1: Cargar Frase

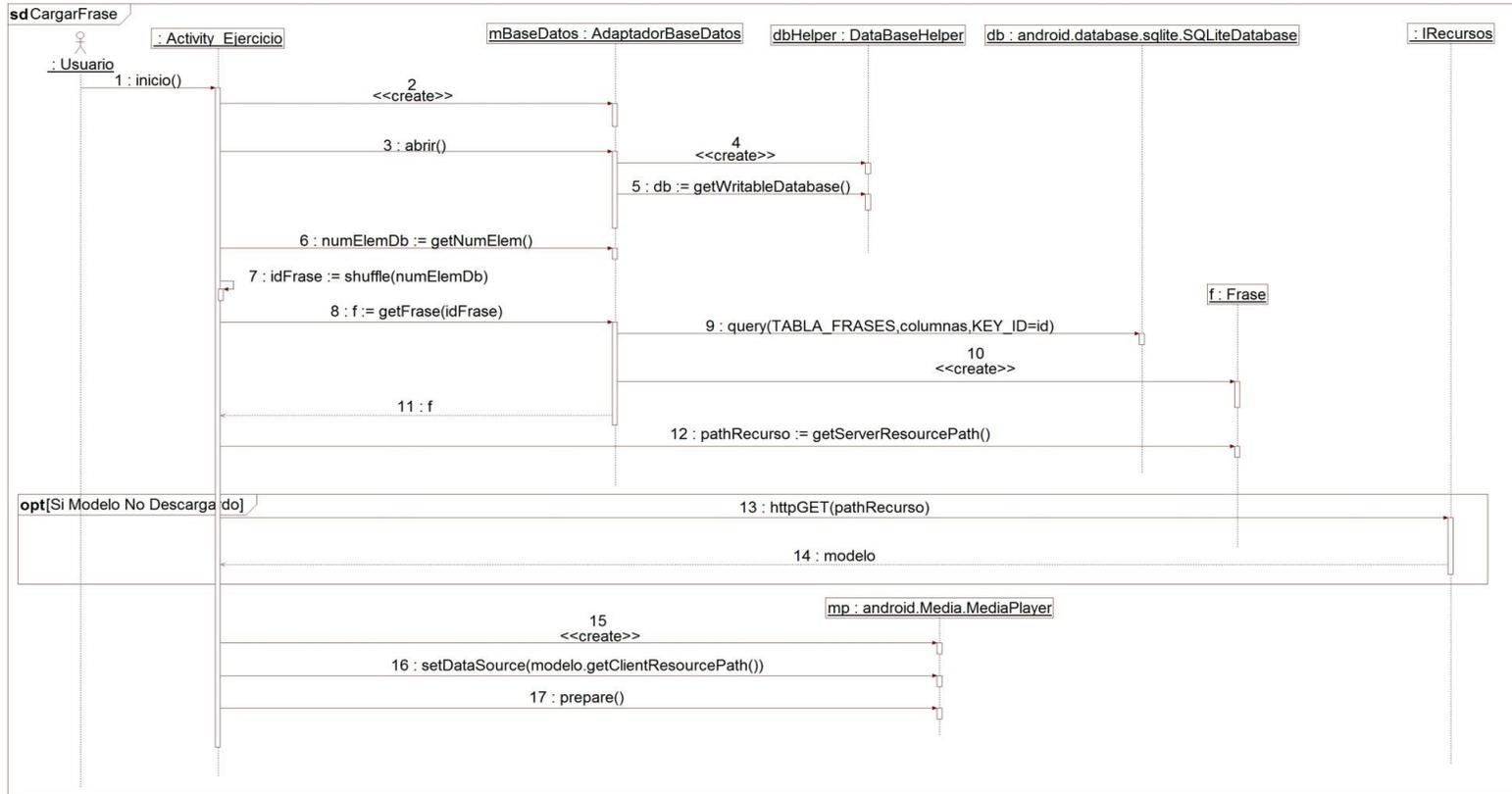


Ilustración 69: Diagrama de Secuencia Cargar Frase - Diseño

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

6.2.3.2.2. Diagrama de Secuencia de Diseño 2: Escuchar Frase

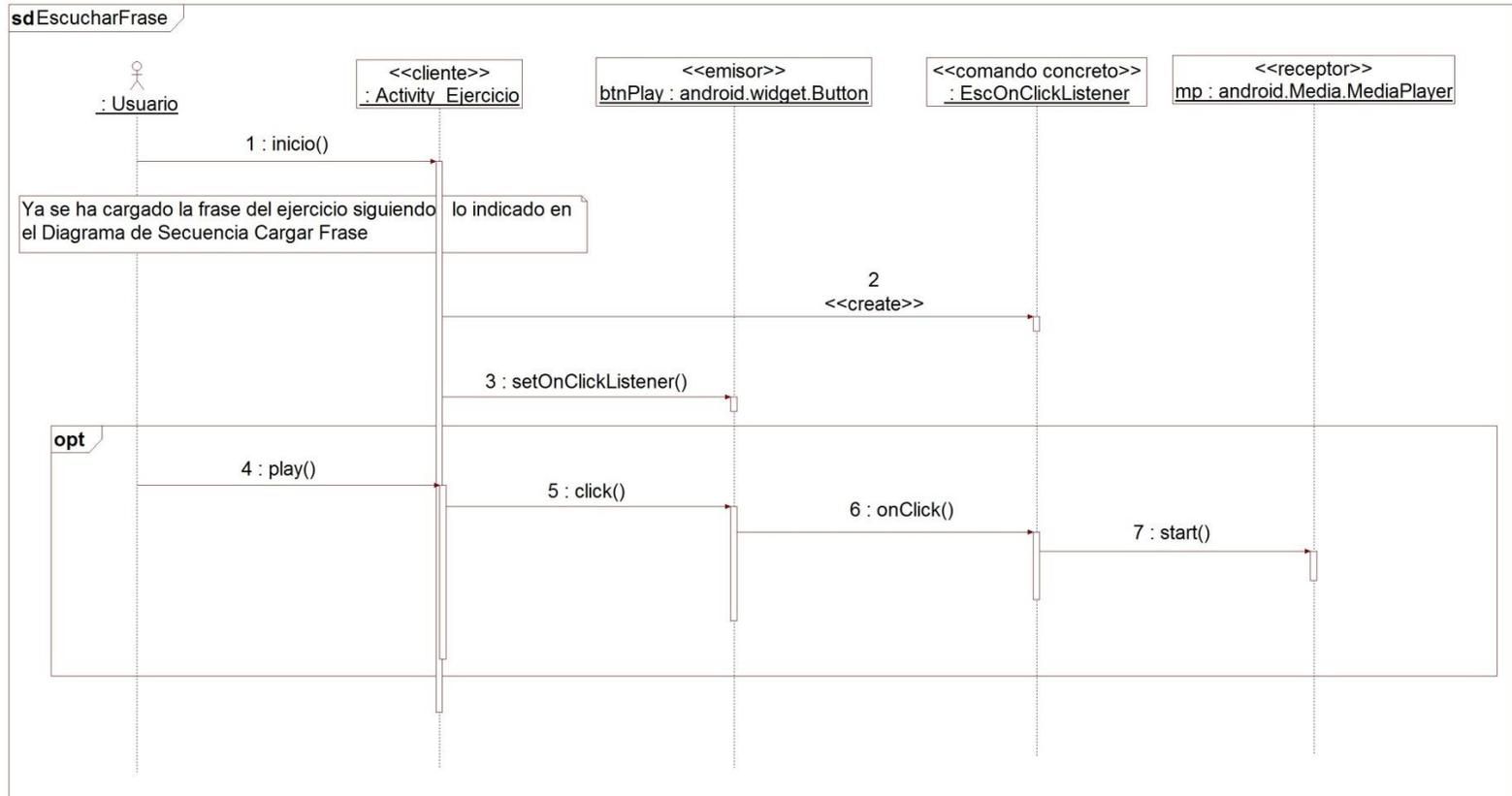


Ilustración 70: Diagrama de Secuencia Escuchar Frase - Diseño

6.2.3.2.3. Diagrama de Secuencia de Diseño 3: Grabar Ensayo

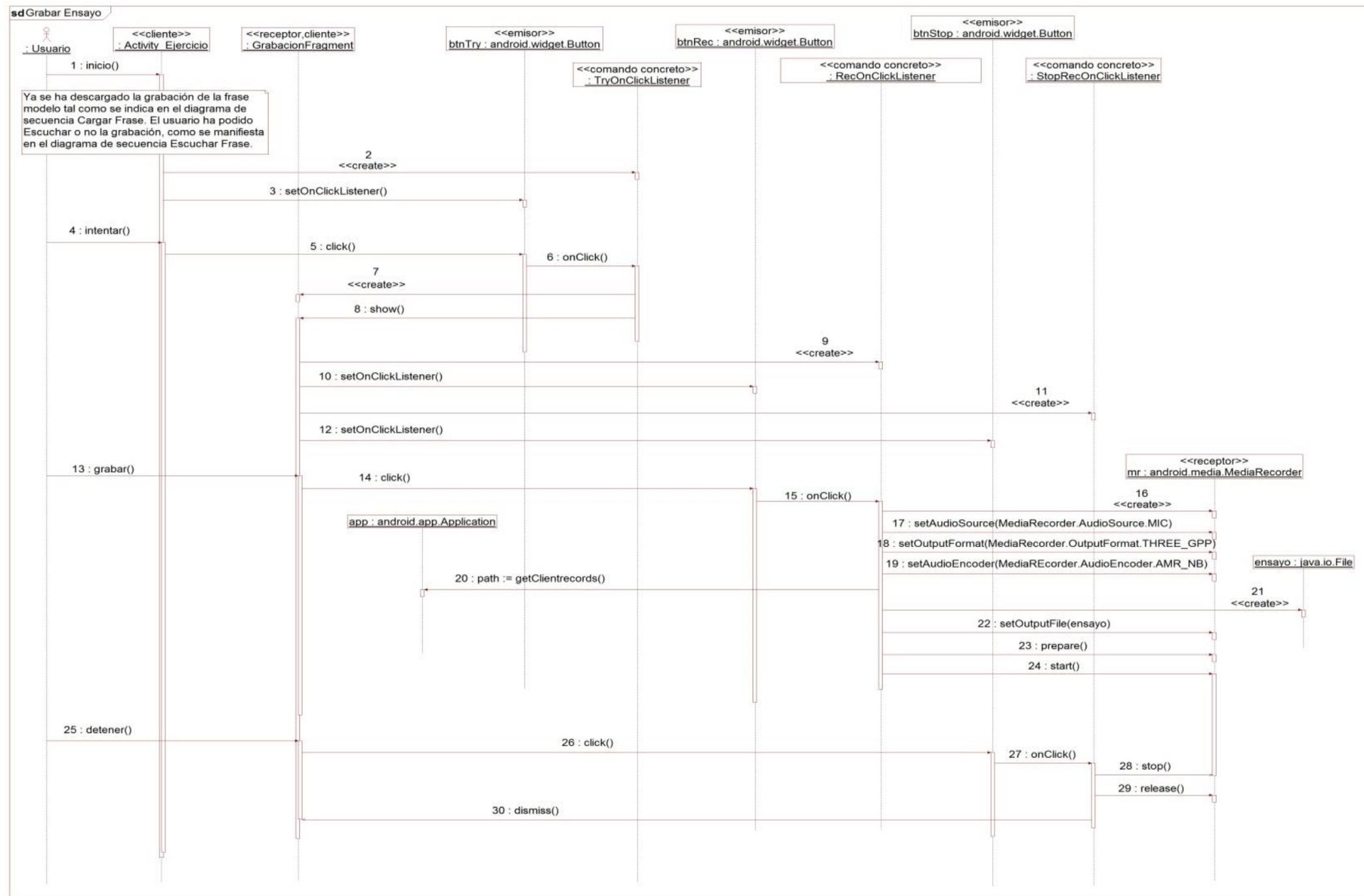


Ilustración 71: Diagrama de Secuencia Grabar Ensayo - Diseño

6.2.3.2.4. Diagrama de Secuencia de Diseño 4: Evaluar Ensayo

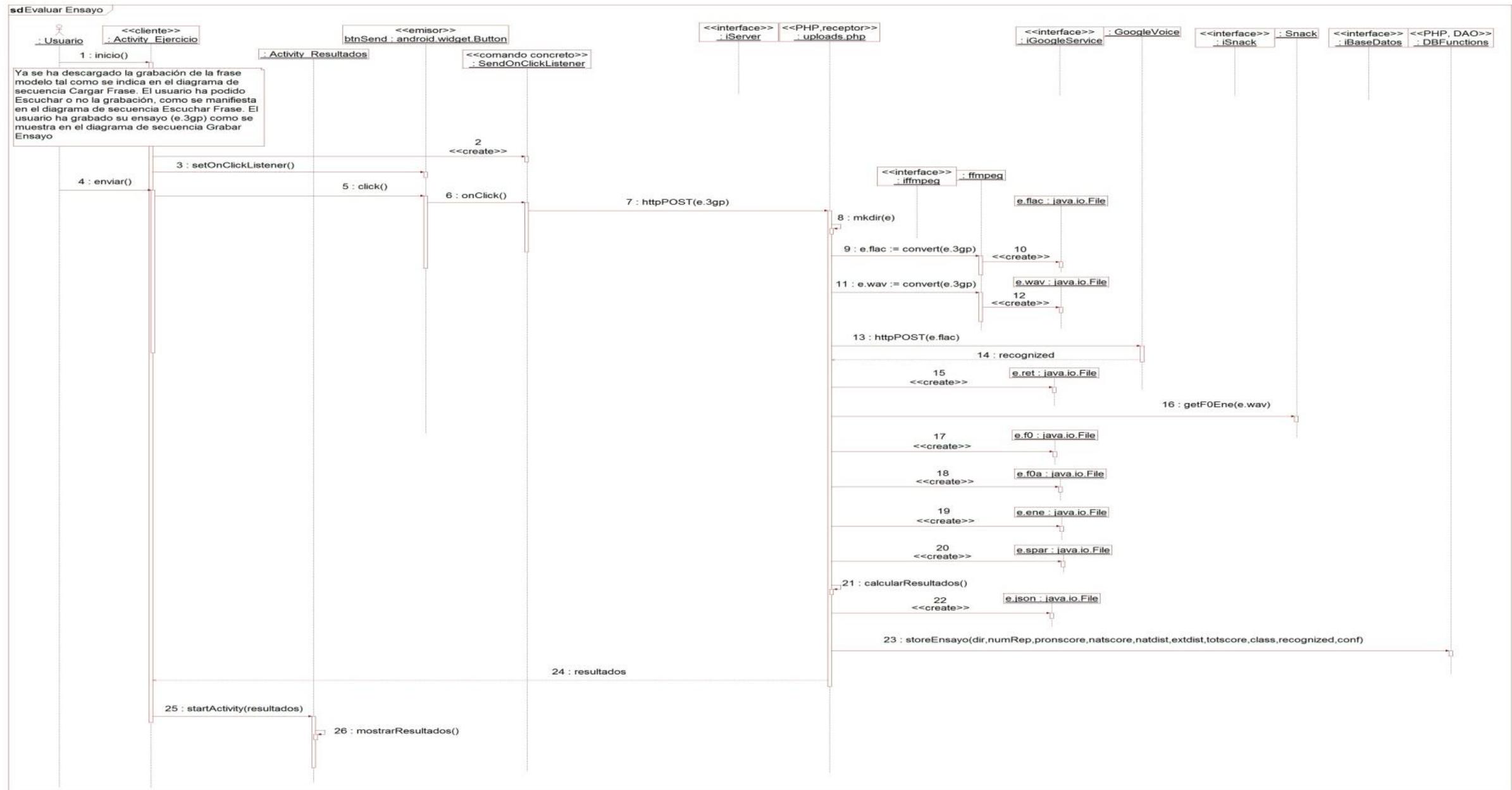


Ilustración 72: Diagrama de Secuencia Evaluar Ensayo - Diseño

6.3. Implementación

El proceso de codificación de la aplicación final está muy influenciado por el trabajo realizado a lo largo del trabajo. Se pretende conseguir un modelo de procesamiento similar al que hemos realizado en la fase de preparación del corpus (la generación de ficheros de características y su ubicación en las carpetas correspondientes), por lo que ha sido necesario adaptar la aplicación de partida para lograr este objetivo. El inconveniente es que tenemos que introducir demasiados cambios en el cliente móvil, pero el aspecto positivo es que en el servidor podemos aprovecharnos de los programas ya desarrollados para la preparación del corpus y utilizarlos de nuevo en el producto final, con lo cual tenemos trabajo ya avanzado.

Los grandes cambios en el cliente móvil son que ya no dispondremos de una síntesis de voz generada localmente en el dispositivo, sino que tendremos que descargar una grabación y reproducirla; y que el sistema de reconocimiento de voz debe preocuparse de mantener la grabación del ensayo. El enfoque es el mismo que el que seguimos en el proyecto del año pasado, pero sin duda, los desafíos que se plantean aquí son más complejos. Afortunadamente, gracias a aquel estudio tenemos experiencia con la grabación y reproducción de archivos de audio en Android, así que la mayor dificultad a salvar será la comunicación con el servidor y la programación de éste, de modo que podamos soportar el registro de usuarios, la subida y descarga de archivos, guardar la información de los ensayos y mantener un árbol de directorios con una estructura igual o similar a la que hemos venido utilizando en la etapa de preparación del corpus (locutores codificados por sexos, ensayos identificados por locutor, frase y repetición).

Por último, en esta ocasión se ha decidido "desnudar" a la aplicación de aquellas funcionalidades extra que decidimos incluir para que no resultara tan simple (escuchar las frases fuera del ejercicio, introducir nuevas frases...). Nos hemos centrado exclusivamente en la parte que nos interesa, la realización del ejercicio. Se ha considerado que el añadido de la gestión de inicio de sesión y registro de usuarios, junto con la evolución del prototipo dotándole de la capacidad de evaluar a un nivel más detallado las locuciones de los ensayos, ha supuesto una mejora cualitativa, por lo que podemos prescindir de aquellas funciones de relleno. Nuestro objetivo es ilustrar las posibilidades de la utilización de un sistema de este tipo, con lo que es suficiente para el alcance de este trabajo.

6.3.1. Lenguajes de programación empleados

Para el desarrollo de esta aplicación se han utilizado varios lenguajes de programación, cada uno para un cometido concreto.

En primer lugar, y de forma mayoritaria, el lenguaje que ha acaparado más líneas de código en la aplicación final es Java. Con él se han implementado la lógica que rige el comportamiento de UVaPronunciación, así como las clases auxiliares que encapsulan las características de objetos que se necesitan para el correcto funcionamiento del programa. Algunos de las clases y métodos que Android aporta en su API difieren de los originales en Java, pero en esencia son similares. Además, dentro del API de Android encontramos muchos paquetes Java que podemos importar en las clases que programemos para completar las carencias que puedan tener.

Otro de los lenguajes que han intervenido en el desarrollo de este trabajo ha sido XML. Con XML definimos la interfaz (los layouts) de las Activities de las que constan las aplicaciones. Por otra parte,

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

con XML definimos otros elementos que también aparecen en la interfaz, como pueden ser cadenas de texto o menús.

Finalmente, en menor medida, se ha utilizado SQL para realizar operaciones de inserción, acceso y actualización en una base de datos SQLite local del dispositivo. Con las sentencias SQL y otros métodos de clases relacionadas conseguimos gestionar adecuadamente la base de datos asociada a nuestra aplicación.

Para la implementación de los componentes del servidor también se han empleado diferentes lenguajes de programación. Como se ha venido adelantando, la lógica que se ocupa del análisis de los datos de la entonación está mayoritariamente codificada en C. Aun así, el programa en el que se computan las puntuaciones finales del ensayo está escrito en Java.

Por otra parte, la gestión de las funciones propias del servidor se ha realizado en PHP, tanto para las tareas de recepción de peticiones como para el enlace con la base de datos MySQL. En este caso, también nos hemos servido de las sentencias SQL para realizar búsquedas e inserciones en las tablas de la base de datos.

6.4. Prueba de la Aplicación

En este apartado se describe el proceso a seguir para la prueba de la aplicación desarrollada. Esencialmente, se identificarán los artefactos afectados por la especificación de requisitos, los cuales serán el objeto de nuestras pruebas; y se enumerarán los requisitos de alto nivel necesarios para realizar las pruebas. Además se definirán todas las condiciones a ser comprobadas en las pruebas. Estas comprobaciones serán obligatorias para asegurar una implementación exitosa y aceptable del producto final (siguiendo los casos de uso y las restricciones definidas). Finalmente, se resumirán los resultados de la actividad de realización de Pruebas en la aplicación desarrollada, proporcionando una evaluación basada en esos resultados.

6.4.1. Plan de Pruebas

El Plan de Pruebas se aplicará a las pruebas que se llevarán a cabo en nuestra aplicación UVaPronunciación. Se tratará de probar la calidad del sistema en cuanto al cumplimiento de las especificaciones definidas en el análisis y verificar el correcto funcionamiento de los aspectos funcionales.

6.4.1.1. Requisitos para las Pruebas

La siguiente lista identifica aquellos elementos (casos de uso, requisitos funcionales, y requisitos no funcionales) que han sido identificados como objetivos para las pruebas. Esta lista representa lo que será probado. Todas las acciones de la interfaz; los accesos a la base de datos y los cálculos internos también se muestran.

Funcionalidad	Interfaz	Base de Datos	Cálculo
Rq.1. La aplicación deberá reproducir en voz alta cadenas de texto.	<ul style="list-style-type: none">• La grabación de modelo que se descarga es la correcta.	Ninguna.	Ninguno.

<p>Rq-NF.10. El cliente móvil de la aplicación deberá descargar los audios correspondientes a las frases que ha leído en el fichero de texto del componente servidor de la aplicación.</p> <p>Rq-NF.11. El cliente móvil de la aplicación deberá reproducir los archivos de audio descargados desde el componente servidor.</p> <p>Rq-NF.27. La aplicación deberá permitir al usuario controlar el volumen de los contenidos auditivos que escuchará.</p>	<ul style="list-style-type: none"> • La grabación de modelo se descarga correctamente en el directorio adecuado. • Si la hubo un fallo en la descarga aparece un aviso para el usuario y se le permite reintentar. • La grabación de modelo se reproduce en el dispositivo. • Se puede subir el volumen de la grabación. 		
--	--	--	--

Tabla 37: Requisitos para las pruebas - Funcionalidad 1

Funcionalidad	Interfaz	Base de Datos	Cálculo
<p>Rq.2. La aplicación deberá reconocer las palabras habladas por un usuario.</p> <p>Rq-NF.13. El cliente móvil de la aplicación deberá subir al servidor las grabaciones correspondientes a las locuciones de los ensayos realizados por los usuarios.</p> <p>Rq-NF.17. La</p>	<ul style="list-style-type: none"> • Aparece un cuadro de diálogo para iniciar la grabación. • La grabación del ensayo se realiza correctamente. • La grabación del ensayo queda almacenada en el directorio correcto dentro del dispositivo del usuario para su posterior envío. 	<p>Ninguno.</p>	<p>Ninguno (Lo hace Google Voice).</p>

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

<p>aplicación deberá utilizar el servicio de reconocimiento de voz proporcionado por Google Voice.</p> <p>Rq-NF.26. La aplicación deberá permitir la introducción de datos a través de la entrada de voz para el reconocimiento de habla.</p>	<ul style="list-style-type: none"> • Si la grabación no se envió correctamente, aparecen un aviso para el usuario. 		
---	---	--	--

Tabla 38: Requisitos para las pruebas - Funcionalidad 2

Funcionalidad	Interfaz	Base de Datos	Cálculo
<p>Rq.3. La aplicación deberá permitir conservar la grabación del ensayo después del reconocimiento.</p> <p>Rq-NF.13. El cliente móvil de la aplicación deberá subir al servidor las grabaciones correspondientes a las locuciones de los ensayos realizados por los usuarios.</p> <p>Rq-NF.21. La aplicación creará un directorio para cada ensayo dentro del directorio del usuario que lo realizó.</p> <p>Rq-NF.22. La aplicación deberá almacenar la grabación del ensayo y los ficheros de datos que se generen a</p>	<ul style="list-style-type: none"> • Si la grabación no se envió correctamente, aparecen un aviso para el usuario. 	<p>Ninguno.</p>	<p>Ninguno.</p>

<p>partir de ella en el directorio que se creó a raíz de su realización en el componente servidor.</p> <p>Rq-NF.23. La aplicación deberá generar por cada ensayo ficheros con formato ".ene", ".f0", ".f0a", ".spar" con las características descritas en la sección "Ficheros de datos" del apartado "Contenidos" del capítulo anterior.</p> <p>Rq-NF.24. La aplicación deberá generar dos ficheros codificados en formato JSON con los datos del reconocimiento de voz y los de la entonación. Los ficheros tendrán formato ".ret" y ".json", respectivamente.</p>			
--	--	--	--

Tabla 39: Requisitos para las pruebas - Funcionalidad 3

Funcionalidad	Interfaz	Base de Datos	Cálculo
<p>Rq.4. La aplicación deberá calificar la pronunciación del usuario.</p> <p>Rq-NF.15. La aplicación puntuará la pronunciación de una frase por parte del usuario con una calificación entre 0 y 10.</p>	<p>Ninguno.</p>	<p>Ninguno.</p>	<ul style="list-style-type: none"> • Distancia de Levenshtein entre la cadena reconocida por Google Voice y la cadena de Origen. • Porcentaje de exactitud de la cadena reconocida en comparación con la original.

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

Rq-NF.16. La aplicación puntuará la pronunciación de una frase mediante un criterio basado en el cálculo de la distancia de Levenshtein.			
--	--	--	--

Tabla 40: Requisitos para las pruebas - Funcionalidad 4

Funcionalidad	Interfaz	Base de Datos	Cálculo
<p>Rq.5. La aplicación deberá extraer los parámetros de duración, frecuencia fundamental y energía de los ensayos y guardarlos: tiempo, número de muestras, máximos y mínimos de frecuencia y energía, tiempo en el que suceden los máximos y mínimos, tiempo normalizado en que se dan los máximos y los mínimos, rangos de frecuencia y energía, promedios de frecuencia y energía y desviación típica de frecuencia y energía; tanto para el global de la grabación, como para cada una de las series temporales (intervalos sonoros y sordos) identificadas dentro de la misma.</p> <p>Rq-NF.14. La</p>	Ninguno.	Ninguno.	<ul style="list-style-type: none"> • Máximos, tiempo en que aparece el máximo, tiempo normalizado en que aparece el máximo, mínimo, tiempo en que aparece el mínimo, tiempo normalizado en que aparece el mínimo, rango de valores, valor promedio y desviación típica de los valores de las frecuencia fundamental y energía del global de la grabación. • Máximos, tiempo en que aparece el máximo, tiempo normalizado en que aparece el máximo, mínimo, tiempo en que aparece el mínimo, tiempo normalizado en que aparece el mínimo, rango de valores, valor promedio y desviación típica de los valores de las frecuencia fundamental y

Capítulo 6: Aplicación Final

aplicación extraerá los valores que caractericen la locución del ensayo utilizando la herramienta Snack.			energía de los intervalos (sonoros y sordos) locales distinguibles en la grabación.
--	--	--	---

Tabla 41: Requisitos para las pruebas - Funcionalidad 5

Funcionalidad	Interfaz	Base de Datos	Cálculo
<p>Rq.6. La aplicación deberá calificar la entonación del usuario.</p> <p>Rq-NF.18. La aplicación puntuará la entonación de una frase por parte del usuario con una calificación entre 0 y 10.</p> <p>Rq-NF.19. La aplicación calificará un ensayo como nativo o extranjero.</p>	Ninguno.	Ninguno.	<ul style="list-style-type: none"> • Distancia con algoritmo DTW a nativos y extranjeros. • Calificación según la función sugerida en el apartado "Sistema de Puntuación" del capítulo anterior.

Tabla 42: Requisitos para las pruebas - Funcionalidad 6

Funcionalidad	Interfaz	Base de Datos	Cálculo
Rq.7. La aplicación deberá guardar la información referente a los ensayos: usuario que la realizó, frase que ensayó, fecha del ensayo, lugar en el que se almacena, puntuación de la pronunciación, palabras reconocidas para la evaluación de la pronunciación, puntuación de la entonación, distancia considerada para	Ninguno.	<ul style="list-style-type: none"> • Añadir en la tabla "Ensayos" (INSERT) los datos de cada ensayo (usuario que la realizó, frase que ensayó, fecha del ensayo, lugar en el que se almacena, puntuación de la pronunciación, palabras reconocidas para la evaluación de la pronunciación, puntuación de la entonación, distancia considerada para evaluar la 	Ninguno.

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

evaluar la entonación, clasificación como nativo o no nativo del ensayo y si el usuario realizó la audición previamente.		entonación, clasificación como nativo o no nativo del ensayo y si el usuario realizó la audición previamente).	
--	--	--	--

Tabla 43: Requisitos para las pruebas - Funcionalidad 7

Funcionalidad	Interfaz	Base de Datos	Cálculo
Rq.8. La aplicación deberá permitir saber al usuario la información de su ensayo: puntuación de la pronunciación, palabras reconocidas para la evaluación de la pronunciación, puntuación de la entonación, clasificación como nativo o no nativo del ensayo y nota global.	<ul style="list-style-type: none"> Se muestra por pantalla la información del intento: nota de pronunciación, palabras reconocidas, nota de entonación, pronóstico de si es nativo o extranjero, y puntuación total. 	Ninguno.	Ninguno.

Tabla 44: Requisitos para las pruebas - Funcionalidad 8

Funcionalidad	Interfaz	Base de Datos	Cálculo
<p>Rq.9. La aplicación deberá permitir el registro de varios usuarios.</p> <p>Rq-NF.20. La aplicación deberá crear un directorio para cada usuario que se registre en el componente servidor.</p>	<ul style="list-style-type: none"> Asegurar que el campo de contraseña está visualmente encriptado al introducir los datos. El sistema se ha cargado tras haber validado la identificación del usuario. Asegurar que todos los campos obligatorios para rellenar previenen 	<ul style="list-style-type: none"> Acceso a la tabla "Usuarios" (SELECT) y verificar los datos enviados (nick y contraseña) Añadir en la tabla "Usuarios" (INSERT) los datos de cada usuario (DNI, nombre, apellidos, nick, contraseña, email, sexo, fecha de nacimiento, nacionalidad y otros datos de 	Ninguno.

	al usuario de que no pueden estar vacíos.	interés)	
--	---	----------	--

Tabla 45: Requisitos para las pruebas - Funcionalidad 9

6.4.1.2. Estrategia para las Pruebas

La estrategia de pruebas define la forma de abordar la actividad de realización de pruebas y la definición de los objetivos de las pruebas. Las principales consideraciones que se realizan en esta sección son las técnicas a seguir y el criterio para determinar cuándo una prueba ha sido completada.

6.4.1.2.1. Tipos de Pruebas

Pruebas de Funcionalidad

Objetivo de la Prueba:	Garantizar la el correcto funcionamiento de la funcionalidad del sistema, incluyendo navegación, datos de entrada, proceso y recuperación.
Técnica:	Ejecutar cada caso de uso, cada flujo posible, utilizando datos válidos y erróneos, para verificar: <ul style="list-style-type: none"> • Los resultados esperados ocurren al introducir datos válidos • Se muestran los mensajes de error apropiados al introducir datos incorrectos • Las reglas del negocio se aplican de forma correcta
Criterio de completitud:	Todas las pruebas planificadas se han ejecutado. Todos los defectos identificados se han abordado.
Consideraciones especiales:	

Tabla 46: Estrategia para las Pruebas de Funcionalidad

Pruebas de Interfaz de Usuario

Objetivo de la Prueba:	Verificar: <ul style="list-style-type: none"> • La correcta navegación a través de las vistas de la aplicación • Se muestran los elementos y opciones correctas según el rol de cada usuario • Aparecen los datos correspondientes en el lugar adecuado
Técnica:	Crear/Modificar pruebas para verificar la adecuada navegación y el estado de los objetos para cada vista de la aplicación.
Criterio de completitud:	Cada vista verificada con éxito será consistente con otras pruebas similares.

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

Consideraciones especiales:	
------------------------------------	--

Tabla 47: Estrategia para las Pruebas de Interfaz de Usuario

Pruebas de Integridad de la Bases de Datos

Objetivo de la Prueba:	Asegurar que los métodos de acceso y manejo de la base de datos funcionan adecuadamente y sin ningún tipo de corrupción en los datos.
Técnica:	Crear varias pruebas de creación de objetos en el sistema y comprobar directamente en la base de datos si se han creado correctamente a través de consultas SQL.
Criterio de completitud:	Todos los métodos de creación funcionan adecuadamente y los elementos están almacenado con los datos correctos sin ningún tipo de corrupción.
Consideraciones especiales:	Requiere tener abierto un navegador con la página de phpmysql para consultar las tablas de la base de datos.

Tabla 48: Estrategia para las Pruebas de Interfaz de Integridad de la Base de Datos

6.4.1.3. Herramientas para las pruebas

Cometido	Herramienta
Documentación del proceso de prueba de la aplicación	Microsoft Office Word 2007
Comprobación de los datos introducidos en la base de datos	Navegador Web Google Chrome / phpMyAdmin
Prueba de la correcta navegación / Introducción de datos / Descarga de audio / Subida de audio	Dispositivo Android

Tabla 49: Herramientas para las pruebas

6.4.2. Casos de Prueba

La descripción de los casos de pruebas se realizará en forma de tablas, donde la primera fila contiene el identificador de la prueba, en la forma CP-XXX, donde XXX es un número que comienza en 001 y se incrementa en una unidad por cada nuevo caso de prueba. En el caso de que una prueba sea fallida, las siguientes pruebas, donde el error se haya resuelto, tendrán el mismo identificador seguido de la expresión “Revisión X”, donde X indica el número de revisión correspondiente, con inicio en uno. En caso de que existieran distintas posibilidades para una misma prueba, se indican con un sufijo al final del identificador con el siguiente formato: CP-XXXz donde z es un carácter en minúscula del alfabeto español (a-z incluida la ñ).

El campo “Situación a probar” indica el aspecto que se va a probar en forma de título o breve descripción del caso de prueba. A continuación viene el campo en el que se detallan las entradas a proporcionar al sistema. El campo “Descripción de la prueba” indica cual va ser la secuencia de

acciones a desarrollar por parte del sistema y del operador de las pruebas. “Salida esperada” y “Salida obtenida” permiten comparar las salidas para saber si la prueba es correcta o fallida, lo que reflejaremos en el campo “Resultado”.

Identificador de la Prueba CP-XXXz	
Elemento a probar	
Situación a probar	
Entrada	
Descripción de la prueba	
Salida esperada	
Salida obtenida	
Resultado	

Tabla 50: Plantilla Casos de Prueba

6.4.2.1. Pruebas de Funcionalidad

Identificación del Usuario en el Cliente Móvil CP-001a	
Elemento a probar	Activity_Login.java e index.php
Situación a probar	El usuario deja en blanco el campo "Nick" y/o el campo "Contraseña" y pulsa "Iniciar sesión".
Entrada	Ninguna, o su "Nick" o su "Contraseña".
Descripción de la prueba	Se abre la aplicación en el dispositivo, se pasa la Splash Activity y se llega a la pantalla de identificación. Se dejan los dos campos sin rellenar, luego sólo el primero y después sólo el segundo. Se pulsa "Iniciar sesión".
Salida esperada	No se inicia la sesión.
Salida obtenida	La esperada.
Resultado	OK

Tabla 51: Identificación del Usuario en el Cliente Móvil CP-001a

Identificación del Usuario en el Cliente Móvil CP-001b	
Elemento a probar	Activity_Login.java e index.php
Situación a probar	El usuario rellena los campos "Nick" y "Contraseña" pero la identificación es incorrecta.
Entrada	"Nick" y "Contraseña" del usuario.

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

Descripción de la prueba	Se abre la aplicación en el dispositivo, se pasa la Splash Activity y se llega a la pantalla de identificación. Se rellenan los campos "Nick" y "Contraseña" (no registrados). Se pulsa "Iniciar sesión".
Salida esperada	No se inicia la sesión.
Salida obtenida	La esperada.
Resultado	OK

Tabla 52: Identificación del Usuario en el Cliente Móvil CP-001b

Identificación del Usuario en el Cliente Móvil CP-001c	
Elemento a probar	Activity_Login.java e index.php
Situación a probar	El usuario rellena los campos "Nick" y "Contraseña" de forma correcta.
Entrada	"Nick" y "Contraseña" del usuario.
Descripción de la prueba	Se abre la aplicación en el dispositivo, se pasa la Splash Activity y se llega a la pantalla de identificación. Se rellenan los campos "Nick" y "Contraseña" (registrados). Se pulsa "Iniciar sesión".
Salida esperada	Se inicia la sesión del usuario.
Salida obtenida	La esperada.
Resultado	OK

Tabla 53: Identificación del Usuario en el Cliente Móvil CP-001c

Registro del Usuario en el Cliente Móvil CP-002a	
Elemento a probar	Activity_Registro.java e index.php
Situación a probar	El usuario deja en blanco cualquiera de los campos de la pantalla de registro y pulsa "Validar registro".
Entrada	Ningún campo, o el campo "Nombre", "Apellidos", "Sexo", "Fecha de Nacimiento", "Nick", "Contraseña", "Repetir Contraseña" o "email".
Descripción de la prueba	Se abre la aplicación en el dispositivo, se pasa la Splash Activity y se llega a la pantalla de identificación. Se pulsa en "Registrarse". En la pantalla de registro se dejan todos los campos en blanco y después se prueban las diferentes combinaciones de campos dejando uno o varios en blanco. Se pulsa en "Validar registro".
Salida esperada	El registro no se realiza.
Salida obtenida	La esperada.
Resultado	OK

Tabla 54: Registro del Usuario en el Cliente Móvil CP-002a

Registro del Usuario en el Cliente Móvil CP-002b	
Elemento a probar	Activity_Registro.java e index.php

Situación a probar	El usuario rellena todos los campos necesarios para el registro, pero un usuario con el mismo "Nick" ya existe.
Entrada	Todos los campos necesarios para el registro.
Descripción de la prueba	Se abre la aplicación en el dispositivo, se pasa la Splash Activity y se llega a la pantalla de identificación. Se pulsa en "Registrarse". En la pantalla de registro se rellenan todos los campos (se introduce un "Nick" ya existente). Se pulsa en "Validar registro"
Salida esperada	El registro no se realiza.
Salida obtenida	La esperada.
Resultado	OK

Tabla 55: Registro del Usuario en el Cliente Móvil CP-002b

Registro del Usuario en el Cliente Móvil CP-002c	
Elemento a probar	Activity_Registro.java e index.php
Situación a probar	El usuario rellena todos los campos necesarios para el registro de forma correcta.
Entrada	Todos los campos necesarios para el registro.
Descripción de la prueba	Se abre la aplicación en el dispositivo, se pasa la Splash Activity y se llega a la pantalla de identificación. Se pulsa en "Registrarse". En la pantalla de registro se rellenan todos los campos (se introduce un "Nick" no existente). Se pulsa en "Validar registro"
Salida esperada	El registro se realiza y se inicia la sesión del usuario.
Salida obtenida	La esperada.
Resultado	OK

Tabla 56: Registro del Usuario en el Cliente Móvil CP-002c

Descarga de la Grabación Modelo CP-003a	
Elemento a probar	Activity_Ejercicio.java y BajarAudio.java
Situación a probar	Al iniciar un ejercicio la grabación de modelo no se descarga correctamente.
Entrada	Ninguna.
Descripción de la prueba	Con la sesión de usuario ya iniciada, en el Menú Principal se pulsa sobre "Ir a ejercicio" .
Salida esperada	El sistema ofrece la posibilidad de reintentar la descarga.
Salida obtenida	La esperada.
Resultado	OK

Tabla 57: Descarga de la Grabación Modelo CP-003a

Descarga de la Grabación Modelo CP-003b

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

Elemento a probar	Activity_Ejercicio.java y BajarAudio.java
Situación a probar	Al iniciar un ejercicio la grabación de modelo se descarga correctamente.
Entrada	Ninguna.
Descripción de la prueba	Con la sesión de usuario ya iniciada, en el Menú Principal se pulsa sobre "Ir a ejercicio" .
Salida esperada	La grabación se almacena en el directorio. La grabación se carga en la Activity, permitiendo ser reproducida.
Salida obtenida	La esperada.
Resultado	OK

Tabla 58: Descarga de la Grabación Modelo CP-003b

Reproducción de la Grabación Modelo CP-004a	
Elemento a probar	Activity_Ejercicio.java
Situación a probar	La grabación modelo descarga se reproduce.
Entrada	Ninguna.
Descripción de la prueba	Con la sesión de usuario ya iniciada, en el Menú Principal se pulsa sobre "Ir a ejercicio" . En la Activity del ejercicio, se pulsa sobre "Play".
Salida esperada	Suena la grabación del ejercicio correspondiente.
Salida obtenida	La esperada.
Resultado	OK

Tabla 59: Reproducción de la Grabación Modelo CP-004a

Grabación del Ensayo CP-005a	
Elemento a probar	Activity_Ejercicio.java
Situación a probar	La grabación del ensayo se realiza correctamente.
Entrada	Voz del usuario intentando pronunciar y entonar correctamente la frase del ejercicio.
Descripción de la prueba	Con la sesión de usuario ya iniciada, en el Menú Principal se pulsa sobre "Ir a ejercicio" . En la Activity del ejercicio, independientemente de si se ha escuchado la grabación o no, se pulsa en "Intentar". En el diálogo de grabación que aparece, se pulsa en "Rec". Se habla a la entrada de voz del terminal y cuando se haya terminado se pulsa "Stop".
Salida esperada	La grabación queda almacenada en el directorio correspondiente del dispositivo para ser subida.
Salida obtenida	La esperada.
Resultado	OK

Tabla 60: Grabación del Ensayo CP-005a

Subida de la Grabación del Ensayo al Servidor CP-006a	
Elemento a probar	Activity_Ejercicio.java y uploads.php
Situación a probar	La grabación se envía de forma incorrecta al servidor.
Entrada	La grabación del ensayo.
Descripción de la prueba	Con la sesión de usuario ya iniciada, en el Menú Principal se pulsa sobre "Ir a ejercicio" . En la Activity del ejercicio, independientemente de si se ha escuchado la grabación o no, se pulsa en "Intentar". En el diálogo de grabación que aparece, se pulsa en "Rec". Se habla a la entrada de voz del terminal y cuando se haya terminado se pulsa "Stop". Independientemente de si se escucha el ensayo o no, se pulsa "Enviar".
Salida esperada	El servidor no guarda el ensayo.
Salida obtenida	La esperada.
Resultado	OK

Tabla 61: Subida de la Grabación del Ensayo al Servidor CP-006a

Subida de la Grabación del Ensayo al Servidor CP-006b	
Elemento a probar	Activity_Ejercicio.java y uploads.php
Situación a probar	La grabación se envía al servidor correctamente.
Entrada	La grabación del ensayo.
Descripción de la prueba	Con la sesión de usuario ya iniciada, en el Menú Principal se pulsa sobre "Ir a ejercicio" . En la Activity del ejercicio, independientemente de si se ha escuchado la grabación o no, se pulsa en "Intentar". En el diálogo de grabación que aparece, se pulsa en "Rec". Se habla a la entrada de voz del terminal y cuando se haya terminado se pulsa "Stop". Independientemente de si se escucha el ensayo o no, se pulsa "Enviar".
Salida esperada	Se crea una carpeta específica para la frase (si no estaba creada) y la repetición que toque en el servidor dentro del directorio del usuario que ha realizado el ensayo y se almacena el ensayo en ella.
Salida obtenida	La esperada.
Resultado	OK

Tabla 62: Subida de la Grabación del Ensayo al Servidor CP-006b

Reconocimiento de las palabras producidas en el Ensayo CP-007a	
Elemento a probar	gv.sh

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

Situación a probar	El reconocimiento de voz se realiza.
Entrada	La grabación del ensayo.
Descripción de la prueba	Se ha recibido la grabación del ensayo en el servidor. Se ejecuta el script "gv.sh".
Salida esperada	Se crean dos archivos de audio con extensión ".flac" y ".wav" con el mismo contenido que el recibido y un fichero de texto con los resultados del reconocimiento.
Salida obtenida	La esperada.
Resultado	OK

Tabla 63: Reconocimiento de las palabras producidas en el Ensayo CP-007a

Análisis de la entonación del Ensayo CP-008a	
Elemento a probar	doEntonacion.sh
Situación a probar	Se extraen los valores característicos y se generan los ficheros de datos del ensayo.
Entrada	La grabación del ensayo.
Descripción de la prueba	Se ha recibido la grabación del ensayo en el servidor. Se ejecuta el script "doEntonacion.sh".
Salida esperada	Se generan los ficheros ".ene", ".f0", ".f0a" y ".spar".
Salida obtenida	La esperada.
Resultado	OK

Tabla 64: Análisis de la entonación del Ensayo CP-008a

Puntuación Pronunciación CP-009a	
Elemento a probar	uploads.php
Situación a probar	Se calcula la nota de la pronunciación del ensayo.
Entrada	Fichero de resultados del reconocimiento de voz.
Descripción de la prueba	Después del análisis de la entonación, se abre el fichero con los datos del reconocimiento de voz y se extrae el campo "utterance". Con el algoritmo de Levenshtein se comparan la cadena original y la reconocida. Se calcula la nota como el tanto por 10 de acierto sobre el total de la cadena original.
Salida esperada	La puntuación de la pronunciación.
Salida obtenida	La esperada.
Resultado	OK

Tabla 65: Puntuación Pronunciación CP-009a

Puntuación Entonación CP-010a

Elemento a probar	doEntonacion.sh
Situación a probar	Se calcula la nota de la entonación del ensayo.
Entrada	Fichero con formato ".spar".
Descripción de la prueba	Después del análisis de la entonación, se calcula la distancia promedio al conjunto de los nativos y los extranjeros. Se compara con ambas y se clasifica el intento como nativo o extranjero. Se calcula la puntuación basándose en la función del apartado "Sistema de Puntuación" del capítulo anterior con respecto al conjunto de los nativos.
Salida esperada	La puntuación de la entonación y una clasificación como nativo o extranjero.
Salida obtenida	La esperada.
Resultado	OK

Tabla 66: Puntuación Entonación CP-010a

Recepción de Resultados en el Cliente Móvil CP-011a	
Elemento a probar	Activity_Resutlados.java
Situación a probar	Se reciben correctamente los resultados en el dispositivo.
Entrada	Fichero de resultados con formato ".json"
Descripción de la prueba	Después de los cálculos, se escribe la respuesta al cliente móvil, que la recibe y la procesa.
Salida esperada	El cliente móvil recibe los daos correctos y al completo.
Salida obtenida	La esperada.
Resultado	OK

Tabla 67: Recepción de Resultados en Cliente Móvil CP-011a

Cierre de Sesión del Usuario CP-012a	
Elemento a probar	Activity_MenuPrincipal.java
Situación a probar	El usuario pulsa en "Cerrar Sesión".
Entrada	Ninguna.
Descripción de la prueba	El usuario se ha identificado en el sistema y se encuentra en el Menú Principal de la aplicación. Pulsa el botón "Cerrar Sesión".
Salida esperada	El usuario ya no está identificado en sistema.
Salida obtenida	La esperada.
Resultado	OK

Tabla 68: Cierre de Sesión del Usuario CP-012

6.4.2.2. Pruebas de Interfaz de Usuario

Identificación del Usuario en el Cliente Móvil CP-001d	
Elemento a probar	Activity_Login.java
Situación a probar	El usuario deja en blanco el campo "Nick" y/o el campo "Contraseña" y pulsa "Iniciar sesión".
Entrada	Ninguna, o su "Nick" o su "Contraseña".
Descripción de la prueba	Se abre la aplicación en el dispositivo, se pasa la Splash Activity y se llega a la pantalla de identificación. Se dejan los dos campos sin rellenar, luego sólo el primero y después sólo el segundo. Se pulsa "Iniciar sesión".
Salida esperada	Un aviso al usuario con los campos que debe rellenar.
Salida obtenida	La esperada.
Resultado	OK

Tabla 69: Identificación del Usuario en el Cliente Móvil CP-001d

Identificación del Usuario en el Cliente Móvil CP-001e	
Elemento a probar	Activity_Login.java e index.php
Situación a probar	El usuario rellena los campos "Nick" y "Contraseña" pero la identificación es incorrecta.
Entrada	"Nick" y "Contraseña" del usuario.
Descripción de la prueba	Se abre la aplicación en el dispositivo, se pasa la Splash Activity y se llega a la pantalla de identificación. Se rellenan los campos "Nick" y "Contraseña" (no registrados). Se pulsa "Iniciar sesión".
Salida esperada	Un aviso al usuario indicándole que ha introducido mal o el "Nick" o la "Contraseña".
Salida obtenida	La esperada.
Resultado	OK

Tabla 70: Identificación del Usuario en el Cliente Móvil CP-001e

Identificación del Usuario en el Cliente Móvil CP-001f	
Elemento a probar	Activity_Login.java e index.php
Situación a probar	El usuario rellena los campos "Nick" y "Contraseña" de forma correcta.
Entrada	"Nick" y "Contraseña" del usuario.
Descripción de la prueba	Se abre la aplicación en el dispositivo, se pasa la Splash Activity y se llega a la pantalla de identificación. Se rellenan los campos "Nick" y "Contraseña" (registrados). Se pulsa "Iniciar sesión".
Salida esperada	Se muestra el Menú Principal de la aplicación.

Salida obtenida	La esperada.
Resultado	OK

Tabla 71: Identificación del Usuario en el Cliente Móvil CP-001f

Registro del Usuario en el Cliente Móvil CP-002d	
Elemento a probar	Activity_Registro.java e index.php
Situación a probar	El usuario deja en blanco cualquiera de los campos de la pantalla de registro y pulsa "Validar registro".
Entrada	Ningún campo, o el campo "Nombre", "Apellidos", "Sexo", "Fecha de Nacimiento", "Nick", "Contraseña""Repetir Contraseña" o "email".
Descripción de la prueba	Se abre la aplicación en el dispositivo, se pasa la Splash Activity y se llega a la pantalla de identificación. Se pulsa en "Registrarse". En la pantalla de registro se dejan todos los campos en blanco y después se prueban las diferentes combinaciones de campos dejando uno o varios en blanco. Se pulsa en "Validar registro"
Salida esperada	Un aviso al usuario con los campos que debe rellenar.
Salida obtenida	La esperada.
Resultado	OK

Tabla 72: Registro del Usuario en el Cliente Móvil CP-002d

Registro del Usuario en el Cliente Móvil CP-002e	
Elemento a probar	Activity_Registro.java e index.php
Situación a probar	El usuario rellena todos los campos necesarios para el registro, pero un usuario con el mismo "Nick" ya existe.
Entrada	Todos los campos necesarios para el registro.
Descripción de la prueba	Se abre la aplicación en el dispositivo, se pasa la Splash Activity y se llega a la pantalla de identificación. Se pulsa en "Registrarse". En la pantalla de registro se rellenan todos los campos (se introduce un "Nick" ya existente). Se pulsa en "Validar registro"
Salida esperada	Un aviso al usuario indicándole que ya existe un usuario con el mismo "Nick".
Salida obtenida	La esperada.
Resultado	OK

Tabla 73: Registro del Usuario en el Cliente Móvil CP-002e

Registro del Usuario en el Cliente Móvil CP-002f	
Elemento a probar	Activity_Registro.java e index.php
Situación a probar	El usuario rellena todos los campos necesarios para el registro de forma correcta.
Entrada	Todos los campos necesarios para el registro.

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

Descripción de la prueba	Se abre la aplicación en el dispositivo, se pasa la Splash Activity y se llega a la pantalla de identificación. Se pulsa en "Registrarse". En la pantalla de registro se rellenan todos los campos (se introduce un "Nick" no existente). Se pulsa en "Validar registro"
Salida esperada	Se muestra el Menú Principal de la aplicación.
Salida obtenida	La esperada.
Resultado	OK

Tabla 74: Registro del Usuario en el Cliente Móvil CP-002f

Descarga de la Grabación Modelo CP-003c	
Elemento a probar	Activity_Ejercicio.java y BajarAudio.java
Situación a probar	Al iniciar un ejercicio la grabación de modelo no se descarga correctamente.
Entrada	Ninguna.
Descripción de la prueba	Con la sesión de usuario ya iniciada, en el Menú Principal se pulsa sobre "Ir a ejercicio" .
Salida esperada	Se muestra un botón para reintentar la descarga.
Salida obtenida	La esperada.
Resultado	OK

Tabla 75: Descarga de la Grabación Modelo CP-003c

Descarga de la Grabación Modelo CP-003d	
Elemento a probar	Activity_Ejercicio.java y BajarAudio.java
Situación a probar	Al iniciar un ejercicio la grabación de modelo se descarga correctamente.
Entrada	Ninguna.
Descripción de la prueba	Con la sesión de usuario ya iniciada, en el Menú Principal se pulsa sobre "Ir a ejercicio" .
Salida esperada	Aparece el reproductor de audio en pantalla.
Salida obtenida	La esperada.
Resultado	OK

Tabla 76: Descarga de la Grabación Modelo CP-003d

Reproducción de la Grabación Modelo CP-004b	
Elemento a probar	Activity_Ejercicio.java
Situación a probar	La grabación modelo descarga se reproduce.
Entrada	Ninguna.

Descripción de la prueba	Con la sesión de usuario ya iniciada, en el Menú Principal se pulsa sobre "Ir a ejercicio" . En la Activity del ejercicio, se pulsa sobre "Play".
Salida esperada	Se puede ver en pantalla el estado de la reproducción. Se deshabilitan las opciones de "Intentar" y de pasar a otro ejercicio.
Salida obtenida	La esperada.
Resultado	OK

Tabla 77: Reproducción de la Grabación Modelo CP-004b

Grabación del Ensayo CP-005b	
Elemento a probar	Activity_Ejercicio.java
Situación a probar	La grabación del ensayo se realiza correctamente.
Entrada	Voz del usuario intentando pronunciar y entonar correctamente la frase del ejercicio.
Descripción de la prueba	Con la sesión de usuario ya iniciada, en el Menú Principal se pulsa sobre "Ir a ejercicio" . En la Activity del ejercicio, independientemente de si se ha escuchado la grabación o no, se pulsa en "Intentar". En el diálogo de grabación que aparece, se pulsa en "Rec". Se habla a la entrada de voz del terminal y cuando se haya terminado se pulsa "Stop".
Salida esperada	Aparece un cuadro de diálogo en pantalla con la imagen de un micrófono.
Salida obtenida	La esperada.
Resultado	OK

Tabla 78: Grabación del Ensayo CP-005b

Subida de la Grabación del Ensayo al Servidor CP-006c	
Elemento a probar	Activity_Ejercicio.java y uploads.php
Situación a probar	La grabación se envía al servidor de forma incorrecta.
Entrada	La grabación del ensayo.
Descripción de la prueba	Con la sesión de usuario ya iniciada, en el Menú Principal se pulsa sobre "Ir a ejercicio" . En la Activity del ejercicio, independientemente de si se ha escuchado la grabación o no, se pulsa en "Intentar". En el diálogo de grabación que aparece, se pulsa en "Rec". Se habla a la entrada de voz del terminal y cuando se haya terminado se pulsa "Stop". Independientemente de si se escucha el ensayo o no, se pulsa "Enviar".
Salida esperada	Se abre un diálogo de progreso indicando que la grabación se está subiendo. Cuando se produce el fallo, no se cierra el cuadro de diálogo y el botón de "Enviar" pone "Reintentar envío".
Salida obtenida	La esperada.
Resultado	OK

Tabla 79: Subida de la Grabación del Ensayo al Servidor CP-006c

Subida de la Grabación del Ensayo al Servidor CP-006d	
Elemento a probar	Activity_Ejercicio.java y uploads.php
Situación a probar	La grabación se envía al servidor correctamente.
Entrada	La grabación del ensayo.
Descripción de la prueba	Con la sesión de usuario ya iniciada, en el Menú Principal se pulsa sobre "Ir a ejercicio" . En la Activity del ejercicio, independientemente de si se ha escuchado la grabación o no, se pulsa en "Intentar". En el diálogo de grabación que aparece, se pulsa en "Rec". Se habla a la entrada de voz del terminal y cuando se haya terminado se pulsa "Stop". Independientemente de si se escucha el ensayo o no, se pulsa "Enviar".
Salida esperada	Se abre un diálogo de progreso indicando que la grabación se está subiendo. Al terminar de subir la grabación, se cierra el cuadro de diálogo. Se abre un nuevo diálogo de progreso a la espera de los resultados.
Salida obtenida	La esperada.
Resultado	OK

Tabla 80: Subida de la Grabación del Ensayo al Servidor CP-006d

Recepción de Resultados en el Cliente Móvil CP-011b	
Elemento a probar	Activity_Resutlados.java
Situación a probar	Se reciben correctamente los resultados en el dispositivo.
Entrada	Ninguna.
Descripción de la prueba	Se ha enviado una grabación de un ensayo al servidor. Se espera a la recepción de los resultados.
Salida esperada	Se cierra el diálogo de progreso y se cambia a la Activity de Resultados. Los resultados se muestran por pantalla en su lugar correspondiente.
Salida obtenida	La esperada.
Resultado	OK

Tabla 81: Recepción de Resultados en Cliente Móvil CP-011b

Cierre de Sesión del Usuario CP-012b	
Elemento a probar	Activity_MenuPrincipal.java
Situación a probar	El usuario pulsa en "Cerrar Sesión".

Entrada	Ninguna.
Descripción de la prueba	El usuario se ha identificado en el sistema y se encuentra en el Menú Principal de la aplicación. Pulsa el botón "Cerrar Sesión".
Salida esperada	Se pregunta al usuario si está seguro de que desea cerrar su sesión. Si pulsa "Cancelar" se mantiene en el Menú Principal, si pulsa "Aceptar" se cierra la aplicación.
Salida obtenida	La esperada.
Resultado	OK

Tabla 82: Cierre de Sesión del Usuario CP-012b

6.4.2.3. Pruebas de Integridad de la Bases de Datos

Identificación del Usuario en el Cliente Móvil CP-001g	
Elemento a probar	Activity_Login.java e index.php
Situación a probar	El usuario rellena los campos "Nick" y "Contraseña" pero la identificación es incorrecta.
Entrada	"Nick" y "Contraseña" del usuario.
Descripción de la prueba	Se abre la aplicación en el dispositivo, se pasa la Splash Activity y se llega a la pantalla de identificación. Se rellenan los campos "Nick" y "Contraseña" (no registrados). Se pulsa "Iniciar sesión".
Salida esperada	La base de datos no devuelve ningún usuario.
Salida obtenida	La esperada.
Resultado	OK

Tabla 83: Identificación del Usuario en el Cliente Móvil CP-001g

Identificación del Usuario en el Cliente Móvil CP-001h	
Elemento a probar	Activity_Login.java e index.php
Situación a probar	El usuario rellena los campos "Nick" y "Contraseña" de forma correcta.
Entrada	"Nick" y "Contraseña" del usuario.
Descripción de la prueba	Se abre la aplicación en el dispositivo, se pasa la Splash Activity y se llega a la pantalla de identificación. Se rellenan los campos "Nick" y "Contraseña" (registrados). Se pulsa "Iniciar sesión".
Salida esperada	La base de datos devuelve el usuario de la tabla "Usuarios" cuyo "Nick" y "Contraseña" coinciden con los del usuario que se está identificando.
Salida obtenida	La esperada.
Resultado	OK

Tabla 84: Identificación del Usuario en el Cliente Móvil CP-001h

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

Registro del Usuario en el Cliente Móvil CP-002g	
Elemento a probar	Activity_Registro.java e index.php
Situación a probar	El usuario rellena todos los campos necesarios para el registro, pero un usuario con el mismo "Nick" ya existe.
Entrada	Todos los campos necesarios para el registro.
Descripción de la prueba	Se abre la aplicación en el dispositivo, se pasa la Splash Activity y se llega a la pantalla de identificación. Se pulsa en "Registrarse". En la pantalla de registro se rellenan todos los campos (se introduce un "Nick" ya existente). Se pulsa en "Validar registro"
Salida esperada	La base de datos devuelve un usuario de la tabla "Usuarios" cuyo "Nick" coincide con el del usuario que se está intentando registrar.
Salida obtenida	La esperada.
Resultado	OK

Tabla 85: Registro del Usuario en el Cliente Móvil CP-002g

Registro del Usuario en el Cliente Móvil CP-002h	
Elemento a probar	Activity_Registro.java e index.php
Situación a probar	El usuario rellena todos los campos necesarios para el registro de forma correcta.
Entrada	Todos los campos necesarios para el registro.
Descripción de la prueba	Se abre la aplicación en el dispositivo, se pasa la Splash Activity y se llega a la pantalla de identificación. Se pulsa en "Registrarse". En la pantalla de registro se rellenan todos los campos (se introduce un "Nick" no existente). Se pulsa en "Validar registro"
Salida esperada	Se crea una nueva entrada en la tabla "Usuarios" de la base de datos "UVaPronunciacion" con los datos del nuevo usuario registrado.
Salida obtenida	La esperada.
Resultado	OK

Tabla 86: Registro del Usuario en el Cliente Móvil CP-002h

Subida de la Grabación del Ensayo al Servidor CP-006e	
Elemento a probar	Activity_Ejercicio.java y uploads.php
Situación a probar	La grabación se envía al servidor correctamente.
Entrada	La grabación del ensayo.
Descripción de la prueba	Con la sesión de usuario ya iniciada, en el Menú Principal se pulsa sobre "Ir a ejercicio". En la Activity del ejercicio, independientemente de si se ha escuchado la grabación o no, se pulsa en "Intentar". En el diálogo de grabación que aparece, se pulsa en "Rec". Se habla a la entrada de voz del terminal y cuando

	se haya terminado se pulsa "Stop". Independientemente de si se escucha el ensayo o no, se pulsa "Enviar".
Salida esperada	Se crea una nueva entrada en la base de datos "UvaPronunciacion" con los datos del ensayo realizado por el usuario.
Salida obtenida	La esperada.
Resultado	OK

Tabla 87: Subida de la Grabación del Ensayo al Servidor CP-006e

6.4.3. Resultado de las Pruebas

En esta sección trataremos de resumir los resultados de la actividad de realización de Pruebas en la aplicación UVaPronunciación desarrollada y proporcionar una evaluación de la misma basada en esos resultados. Intentaremos identificar también posibles mejoras que puedan mejorar el sistema.

6.4.3.1. Variaciones

Todas las pruebas realizadas respetan el Plan de Pruebas, y han seguido las directrices marcadas en el apartado de Plan de Pruebas. Se han diseñado 37 casos de prueba, de los cuales se han comprobado la totalidad, tal como se había previsto.

6.4.3.2. Resumen de Resultados

Los casos de prueba definidos previamente para nuestra aplicación UVaPronunciación fueron efectuados siguiendo la estrategia definida en el documento del Plan de Pruebas.

La cobertura de pruebas se refiere a que se abordan todos los casos de prueba tal como se definen en el apartado de Plan de Pruebas. Como se indica en la sección de Variaciones, se han realizado un total de 37 pruebas. La cobertura de las pruebas ha sido completa.

	Pruebas realizadas	Éxito	Fallo
Total	37	37	0

Tabla 88: Resumen de los Resultados de las Pruebas

Las pruebas de prestaciones no han sido consideradas debido a la escasez de tiempo para conseguir un rendimiento óptimo del sistema. Se ha primado la funcionalidad frente a aspectos como la eficiencia.

6.4.3.3. Cobertura de las Pruebas basada en Requisitos

Las pruebas que se han realizado con el prototipo están definidas en el apartado de Plan de Pruebas junto con sus criterios de completitud. La cobertura de resultados ha sido la siguiente:

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

	Casos de Prueba realizados	Casos de Prueba con éxito	Casos de Prueba con fallo
Funcionalidad	18	18	0
Interfaz de Usuario	14	14	0
Integridad de Base de Datos	5	5	0
Total	37	37	0

Tabla 89: Cobertura de las Pruebas basada en Requisitos

6.4.3.4. Tipos de Errores e Impacto en el Sistema

Ante estas pruebas, pese a los buenos resultados obtenidos, nos hemos percatado de que existe un pequeño problema con el tiempo de procesamiento de la grabación: en grabaciones de locuciones de frases largas, es posible que se tarde más en procesar la señal y que la aplicación móvil se mantenga demasiado tiempo, bloqueando el hilo principal, a la espera de una respuesta por parte del servidor. Esto puede llevar a que el sistema operativo detenga la aplicación de forma indeseada. Podríamos catalogar este fallo como error en la interfaz e incluso de funcionalidad, puesto que la respuesta no llegaría a recibirse. El impacto en el sistema podría ser fatal, por lo que es conveniente pensar en una alternativa que nos permita recibir la respuesta sin problemas de que el cliente móvil pueda cerrarse inesperadamente durante el procesamiento de la grabación.

6.4.3.5. Acciones Sugeridas

Las acciones que se recomiendan tras los resultados obtenidos y las observaciones recogidas del comportamiento de la aplicación son las siguientes:

- Repetir las pruebas con locuciones mucho más lentas, prolongando las grabaciones durante más tiempo del esperado, para comprobar si el cliente móvil aguanta la espera sin cerrarse.
- Establecer un valor más alto para el timeout de la conexión entre el cliente móvil y el componente servidor de la aplicación y comprobar si el comportamiento anómalo existe.
- Diseñar un sistema por el que el cliente móvil no bloquee el hilo principal de la aplicación y reciba sus calificaciones en una especie de "buzón de entrada", recibiendo el aviso de la calificación a modo de notificación.

Manual de Usuario

Este capítulo contiene el manual de usuario de la aplicación UvaPronunciacion. En él se comentarán los requisitos y pasos a seguir para la instalación y puesta en marcha tanto del servidor como del cliente móvil Android. Además, se describirán las diferentes pantallas (Activities) que la componen tratando de explicar a modo de instrucciones cómo se debe utilizar para su correcto funcionamiento.

7.1. Servidor

En este apartado se describirán las instrucciones de uso para manejar UVaPronunciacion, tanto su componente servidor como la aplicación móvil cliente. Se detallarán los requisitos y los pasos necesarios para su instalación, así como la utilización de las funciones de la aplicación en sí.

7.1.1. Requisitos para la Instalación

Algunos de los requisitos expuestos a continuación pueden ser modificados, sin embargo no se garantiza la plena funcionalidad del servidor con otras configuraciones.

Se requiere un servidor web para poder realizar la comunicación entre el cliente y el servidor mediante peticiones GET/POST.

El servidor también almacenará información relativa a los usuarios, un registro de sus ensayos y la localización de las grabaciones utilizadas como referencia. Las consultas están escritas en SQL y son ejecutadas por PHP. Éstas han sido adaptadas para un gestor de base de datos MySQL por lo que se recomienda su elección. Sin embargo, no es necesario siempre y cuando el administrador adecue las consultas para otro gestor de su preferencia.

Requisitos mínimos	Requisitos recomendados
GNU/Linux	GNU/Linux Debian
Gestor de base de datos	MySQL
Servidor Web	Apache HTTP
Wget	

Tabla 90: Requisitos mínimos para la instalación del servidor

7.1.2. Instalación

Para explicar cómo se instala el servidor de UVaPronunciacion se describirá cómo se ha procedido para la realización de este Proyecto. De debe apuntar que la distribución de las máquinas que se han utilizado es Ubuntu 12.04.

Antes de comenzar, es conveniente tener actualizado el equipo, para ello, en el terminal tecleamos:

```
sudo apt-get update
sudo apt-get upgrade
```

A continuación se nos sugerirá que reiniciemos el sistema para que se complete la actualización y se puedan aplicar los cambios.

Ahora sí, ya podemos iniciar la instalación. Se deben instalar los componentes típicos necesarios para el soporte de un servidor web: Apache, MySQL y phpMyAdmin. Podemos hacerlo todo de una vez escribiendo la siguiente entrada por línea de comandos:

```
sudo apt-get install apache2 php5 libapache2-mod-php5 mysql-server
mysql-client phpmyadmin
```

Durante el transcurso de la instalación se nos solicitarán algunos datos como opciones de Apache o las contraseñas de la base de datos.

Los archivos de configuración de apache2 se encuentran en el directorio /etc/apache2. El archivo principal de configuración es /etc/apache2/apache2.conf. Necesitaremos editarlo para poder terminar la instalación y asegurar el funcionamiento del servidor, debemos hacerlo con privilegios de administrador:

```
sudo gedit /etc/apache2/apache2.conf
```

Y añadimos al final la siguiente línea:

```
Include /etc/phpmyadmin/apache.conf
```

Por defecto, el directorio raíz del servidor web es la carpeta /var/www. Todos los documentos que se encuentren dentro de esta carpeta serán accesibles vía web. Dentro del directorio raíz de documentos es donde debemos copiar la carpeta /uvapronunciacion incluida en el directorio /Código/Servidor del CD que se adjunta con esta memoria (ver Apéndice G).

Para acceder a la web principal del servidor (en nuestro caso a sus servicios) se conecta con la dirección IP del servidor (<http://ip-del-servidor>) o utilizando du nombre si disponemos de un DNS funcionando (<http://nombre-del-servidor>).

Una vez tenemos el servidor preparado, debemos instalar los complementos que utilizará para sus funciones.

7.1.2.1. Instalación de la base de datos

Para cargar la base de datos diseñada (tiene tres sencillas tablas, Usuarios, Ensayos y Frases; hemos optado por una solución rápida y fácil), abrimos phpMyAdmin en el navegador, y después de habernos identificado, importamos el fichero ".sql" incluido en la carpeta /Código/Servidor del CD que se adjunta con esta memoria (ver Apéndice G). Acto seguido se cargará la base de datos UvaPronunciación y podremos ver las tablas en nuestro panel.

7.1.2.2. Instalación de ffmpeg

El software ffmpeg es con el que el servidor llevará a cabo las conversiones entre formatos de audio. Las grabaciones que recibe tienen el formato ".3gp", por tanto, necesitamos convertir los archivos de audio a formato ".wav" para el análisis de la entonación del ensayo; y a formato ".flac" para el reconocimiento de voz con Google Voice. Para su instalación, introducimos la siguiente orden:

```
sudo apt-get install ffmpeg
```

7.1.2.3. Instalación de Snack

Como hemos comentado en el capítulo de "Preparación del Corpus" (capítulo 5), el software Snack es el que hemos elegido para la extracción de los parámetros que caracterizan las locuciones realizadas por los usuarios del sistema. Lo instalaremos escribiendo los siguientes comandos:

```
sudo apt-get install libsnack2
```

```
sudo apt-get install libsnack2-dev
```

```
sudo apt-get install libsnack2-alsa
```

```
sudo apt-get install python-tksnack
```

7.1.3. Puesta en Marcha

El servidor web apache2 y la base de datos MySQL, al igual que todos los servicios en Debian, disponen de un script de arranque y parada en la carpeta /etc/init.d. Para poner en marcha debemos utilizar los privilegios de administrador, así que escribimos por línea de comandos:

```
sudo /etc/init.d/apache2 start
```

```
sudo /etc/init.d/mysql start
```

7.1.4. Desconexión

Como indicamos en la sección anterior, en el mismo directorio en el que se encuentra un script de arranque de Apache2 y MySQL también hay un script de parada para cada uno. Al igual que hicimos antes, utilizaremos los privilegios de administrador para detenerlos. En este caso introduciremos:

```
sudo /etc/init.d/apache2 stop
```

```
sudo /etc/init.d/mysql stop
```

7.2. Cliente Móvil Android

Tal como se indicó previamente, el cliente es una aplicación móvil que se ejecuta en terminales con plataforma Android. Interactúa con el servidor para descargar los audios de referencia y para subir las grabaciones de los ensayos que serán evaluadas y calificadas. Al igual que se hizo con el servidor, a continuación se explicarán los requisitos y pasos a seguir para su instalación y correcto funcionamiento.

7.2.1. Requisitos para la Instalación

La aplicación está desarrollada para dispositivos Android con una versión 2.2 o superior, si no posee un terminal con estas características no podrá ejecutar UVaPronunciación.

Para poder descargar los datos y contenidos del servidor, se requiere una conexión de datos 3G a internet o tener acceso a una red Wi-Fi.

El almacenamiento de los archivos y componentes multimedia necesarios para el correcto funcionamiento de la aplicación y las Actividades se realizará en la tarjeta microSD del dispositivo, por tanto es necesario que el terminal este provisto de una.

7.2.2. Instalación

Para instalar UVaPronunciación en su dispositivo móvil (smartphone o tablet), se debe guardar el ejecutable ".apk" que se encuentra en el directorio /bin de la carpeta del Proyecto de Eclipse en una carpeta del dispositivo a la elección del usuario (se recomienda que sea de fácil acceso). Dirigirse a la carpeta donde esté ubicado el ejecutable ".apk" y pulsar sobre él.

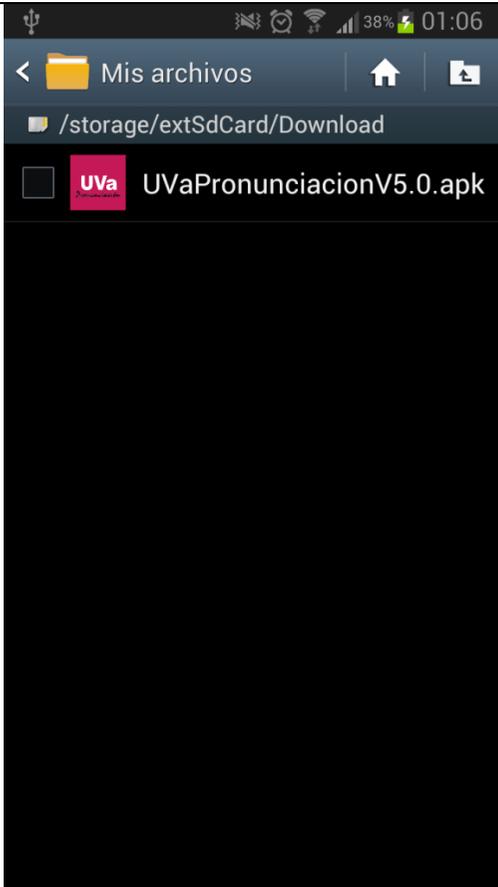


Ilustración 73: Ejecutable ".apk" de UVaPronunciación

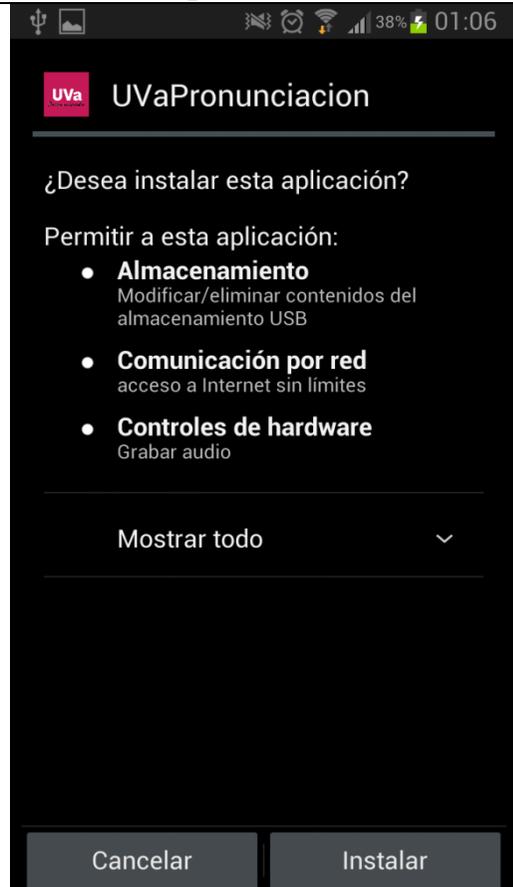


Ilustración 74: Confirmación para la instalación de UVaPronunciación

A continuación se mostrará una pantalla solicitando el permiso del usuario para instalar la aplicación e informando de los servicios que la aplicación utiliza (almacenamiento, comunicación por red y controles de hardware), pulsar "Instalar".

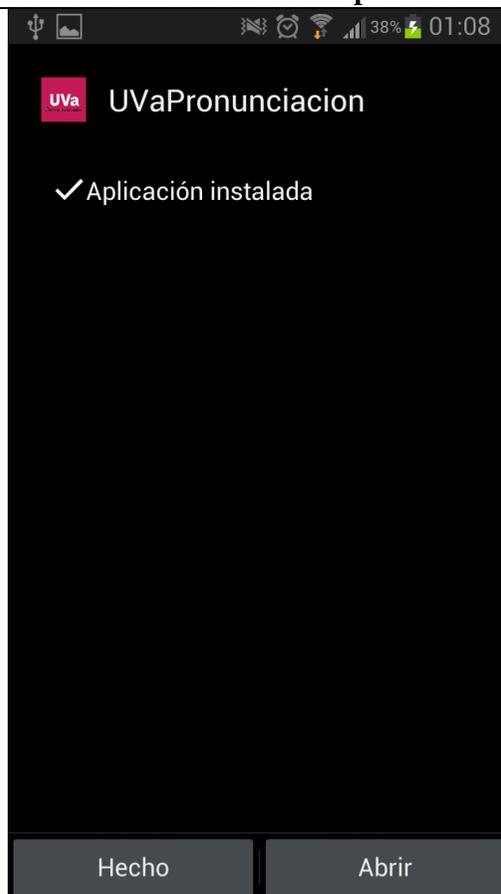


Ilustración 75: UVaPronunciación instalada

Si la instalación se ha realizado con éxito, se informa al usuario de ello, y ya estará lista para su uso.

7.2.3. Puesta en Marcha

Una vez instalada la aplicación, para acceder a ella el usuario tan sólo debe pulsar el icono en el menú de aplicaciones. Una vez haya pulsado, se lanzará la aplicación UVaPronunciación en el dispositivo del usuario.



Ilustración 76: Menú de aplicaciones con icono de UVaPronunciación



Ilustración 77: Splash Activity de UVaPronunciación

7.2.4. Instrucciones

A continuación se detallan las funciones de la aplicación UVaPronunciación y las instrucciones para que el usuario pueda manejarla correctamente.

7.2.4.1. Configuración del Host

La configuración del Host sirve para especificar a la aplicación la IP de donde debe obtener los contenidos e información. Es una opción que se ha incluido debido a que la aplicación ha sido probada con el servidor instalado en varias máquinas. Se trata de hacer más cómoda la prueba.

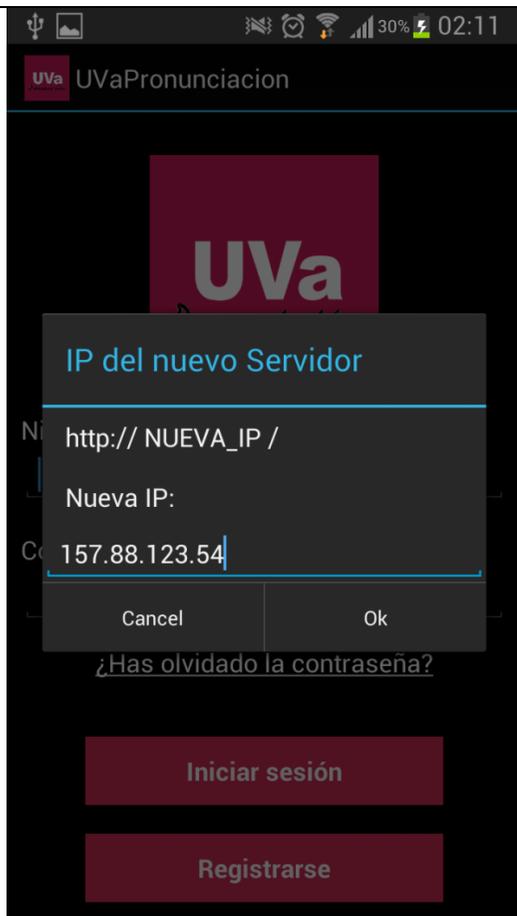


Ilustración 78: Cambiar dirección IP del servidor

7.2.4.2. Registro

Para poder utilizar la aplicación UVaPronunciación, el usuario debe estar registrado en el sistema.

Para registrarse, el usuario debe pulsar en la pantalla de “Inicio de Sesión” (Login) el botón “Registrarse”. A continuación se muestra al usuario una pantalla con un formulario que debe rellenar al completo para poder finalizar el registro. Los campos a rellenar por el usuario son los siguientes:

UVa UVaPronunciacion

Nombre:
Diego

Apellidos:
Vallejo Alonso

Sexo:
 Hombre
 Mujer

Fecha de nacimiento:
1-12-1987

Nacionalidad:
Spain

Email:
diego.vallejo@alumnos.uva.es

Nick:
dvall

Contraseña:
.....

Validar registro

- Nombre: Nombre del usuario.
- Apellidos: Apellidos de usuario.
- Sexo: Hombre o mujer.
- Fecha de nacimiento: Día mes y año en que nació el usuario. Al pulsar el botón "Elegir Fecha" se abre un cuadro de diálogo para seleccionar la fecha.
- Nacionalidad: País de nacimiento del usuario. Sólo se puede especificar una única nacionalidad de la lista desplegable.
- Email: Dirección de correo electrónico de contacto. En ella podrá recuperar la nueva contraseña cuando la solicite (función sin implementar en este proyecto).
- Nick: Nombre de usuario con el que se identificará en el sistema. Debe ser único.
- Contraseña: Password con el que verificará su identificación al iniciar sesión.

Ilustración 79: Activity de Registro (desplegada) de UVaPronunciación

Para finalizar el registro, pulsar "Validar Registro". Si alguno de los campos no ha sido rellenado o ha sido rellenado de forma incorrecta, la aplicación informará al usuario a través de un texto emergente. En caso de que el registro se haya efectuado apropiadamente, se iniciará automáticamente la sesión del usuario en el sistema y se abrirá la pantalla del Menú Principal de la aplicación en su terminal.

7.2.4.3. Identificación / Inicio de sesión en el sistema (Login)

Antes de poder utilizar la aplicación UVaPronunciación, el usuario debe identificarse en el sistema (como se ha señalado es indispensable haberse registrado previamente).

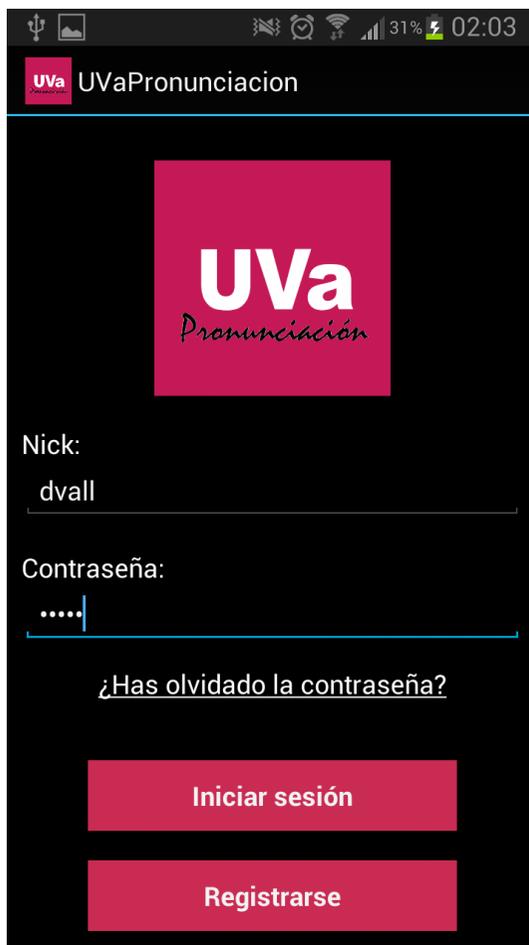


Ilustración 80: Activity de Inicio de Sesión de UVaPronunciación

Para identificarse, el usuario debe introducir en los cuadros de texto habilitados, su nick de usuario (no confundir con su nombre) y su contraseña, los cuales especificó en el momento de su registro, y pulsar el botón "Iniciar Sesión".

7.2.4.4. Realizar un ejercicio

Para comenzar a realizar una ejercicio, el usuario debe encontrarse en el Menú Principal y seleccionar la opción "Ir a Ejercicios".

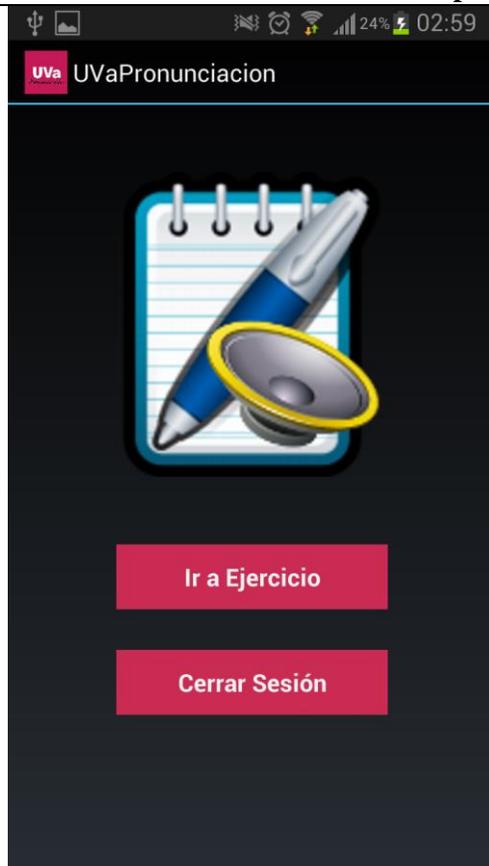


Ilustración 81: Activity Menú Principal de UVaPronunciación

Aparecerá en pantalla un nuevo ejercicio, seguramente precedida de un mensaje de carga de contenidos (en ese momento se estará descargando la grabación de audio que acompaña al ejercicio que toque, si es que no la tiene ya descargada en su dispositivo). En esta pantalla se muestran el número de la frase, la frase y los controles que el usuario necesita para escuchar la grabación de modelo y realizar un ensayo. Los ejercicios se componen de dos fases: reproducción y grabación.

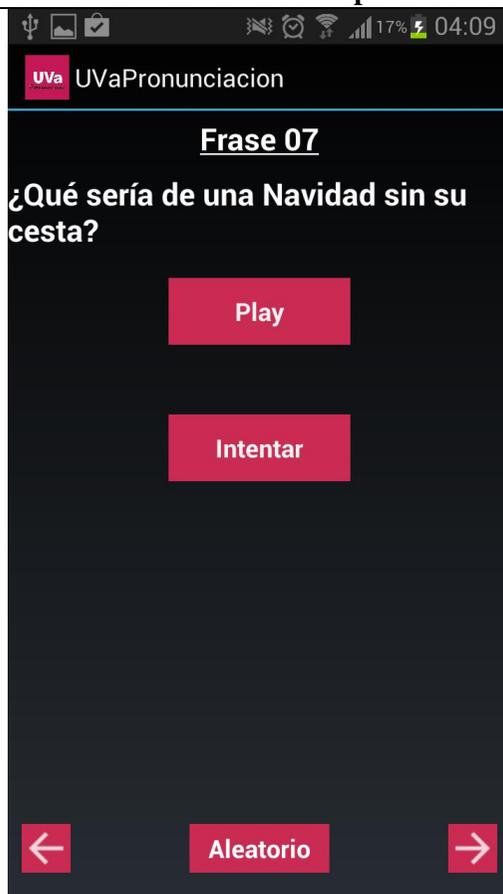


Ilustración 82: Activity Ejercicio de UVaPronunciación

7.2.4.4.1. Escuchar la Grabación de Referencia

En la etapa de reproducción el usuario decide cuándo se reproducen los recursos y cuántas veces. Para iniciar o repetir la reproducción del contenido multimedia tan solo hay que pulsar sobre el botón "Play". En ese momento, podrá escuchar la grabación modelo en voz alta. En el momento que lo desee, puede detener la reproducción pulsando el botón "Stop". en lo que suena la grabación queda deshabilitada la opción de realizar el ensayo.



Ilustración 83: Activity Ejercicio de UVaPronunciación - Escuchar la grabación de modelo

7.2.4.4.2. Grabación y Envío del Ensayo

La etapa de grabación se inicia cuando el usuario pulsa sobre el botón "Intentar". En ese momento, se lanzará un dialogo emergente en el que podrá iniciar la grabación del ensayo pulsando el botón "Rec". En este momento, el usuario deberá hablar claramente a la entrada de voz del dispositivo leyendo la frase que aparece en pantalla, y procurando entonar como el modelo. Cuando haya terminado el ensayo, deberá pulsar el botón "Stop" para detener la grabación.

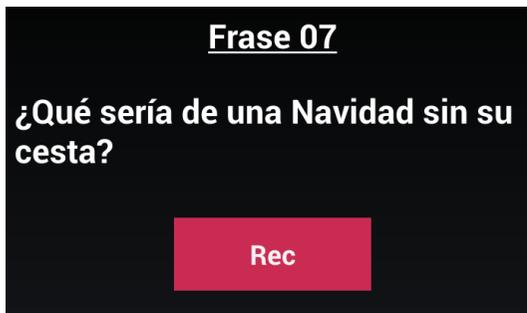


Ilustración 84: Cuadro de Diálogo de Grabación

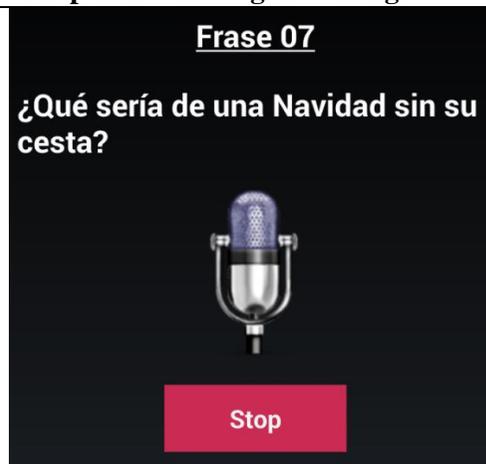


Ilustración 85: Cuadro de Diálogo de Grabación - Grabando

A continuación, se le ofrecerá escuchar su ensayo, por si deseara comprobar que la grabación no presenta ruidos, o simplemente para que se asegure de si está satisfecho con su ensayo.

Independientemente de si escucha su ensayo o no, se le mostrará la opción de "Enviar" para poder enviar la grabación y ser evaluado.

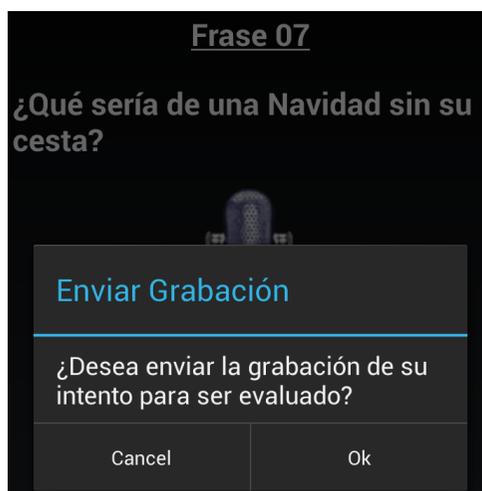


Ilustración 86: Confirmación de Envío de la Grabación del Ensayo al Servidor

7.2.4.4.3. Recepción de los Resultados

Al enviar la grabación se procederá a su calificación en el servidor y posterior lanzamiento de la pantalla de resultados en la pantalla del usuario. En esta pantalla se muestra de nuevo el número de frase y su contenido, y además la puntuación obtenida en el intento (de 0 a 10). La calificación viene detallada por los diferente aspectos que hemos comentado a lo largo del ejercicio: la nota de su

pronunciación (de 0 a 10), qué palabras han sido reconocidas en su ensayo, la nota de entonación (de 0 a 10) y una observación indicando su pronóstico, si es nativo o extranjero.

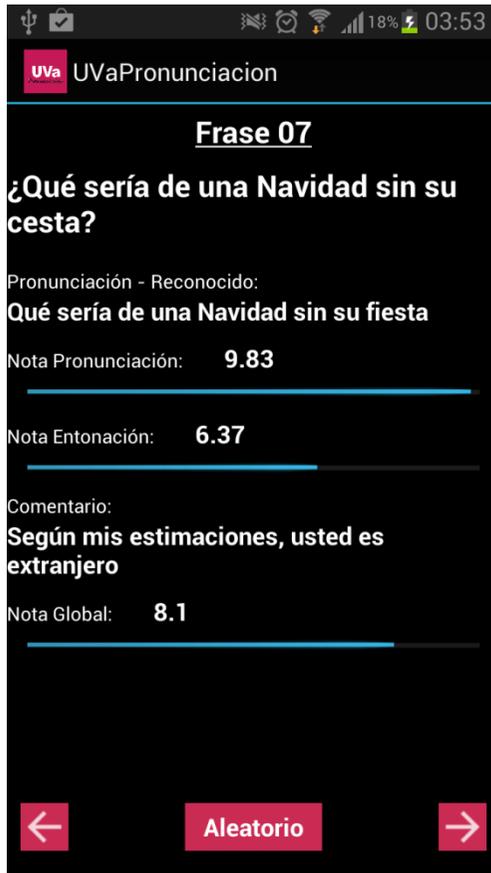


Ilustración 87:Activity de Resultados en UVaPronunciación

Después de haber visto su puntuación el usuario podrá volver a la Activity del mismo ejercicio (por si deseara repetir el intento), pasar al siguiente, volver al anterior, realizar uno aleatorio, o salir al menú principal.

7.2.4.5. Cerrar Sesión (Logout)

Cuando el usuario lo desee, puede cerrar su sesión pulsando la opción "Cerrar sesión" pulsando el botón de menú cuando se encuentre en la pantalla del Menú Principal de la aplicación.

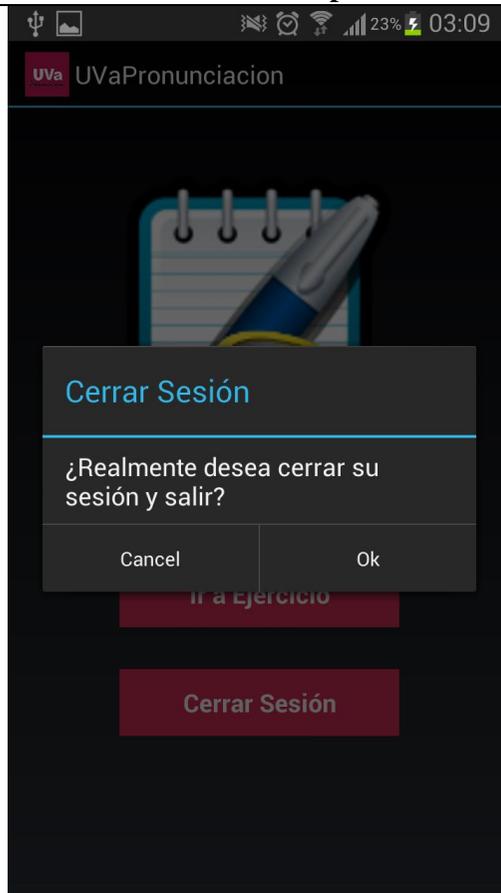


Ilustración 88: Cerrar Sesión de Usuario

Conclusiones y Trabajo Futuro

En este capítulo se presentarán los resultados obtenidos en este proyecto, realizaremos una reflexión global de todo el trabajo comparando los resultados que hemos cosechado durante el transcurso de este estudio con los objetivos que nos marcamos al principio. A continuación, expondremos las conclusiones a las que hemos llegado una vez que ya hemos tenido la oportunidad de probar nuestro módulo de valuación de la entonación integrado en el prototipo. Para finalizar, propondremos una serie de ideas que el proyecto nos sugiere que podrían ser buenas opciones para ampliar el trabajo desarrollado hasta este punto.

8.1. Resultados Finales

Para interpretar los resultados finales de este proyecto, debemos recordar cuáles fueron las metas que nos marcamos cuando decidimos afrontar este desafío, lo cual nos ayudará a tomar cierta perspectiva para comprobar si hemos cumplido con los objetivos establecidos. En aquel caso identificamos un objetivo principal que después subdivimos, a su vez, en otros menores:

Desarrollo de un módulo de comparación de entonación entre locuciones de una misma frase que permita obtener una medida cuantitativa de similitud entre la entonación esperada de esa frase y la producida por el locutor.

Para poder enfrentarnos a la cuestión de si hemos abordado el objetivo central del proyecto, primero comprobaremos que uno a uno hemos superado los objetivos en los que desglosamos éste y que nos impusimos en el inicio de este trabajo:

- *Recopilación de un corpus vocal en el que se disponga de varias pronunciaciones de un conjunto de frases controlado (siguiendo las directrices marcadas para la obtención de los DELE) por varios locutores y que pueda emplearse para elaborar el modelo de comparación. El corpus debe incluir grabaciones de los locutores repitiendo los ejercicios en diferentes situaciones (repetición del mismo ejercicio 5 veces seguidas, repetición del ejercicio a las dos horas, a la semana...), para poder comprobar las variaciones intra-locutor.*

En este caso, la recopilación del corpus corrió a cargo del Grupo de Investigación ECA-SIMM (Entornos de Computación Avanzada y Sistema de Interacción Multimodal) de la Universidad de Valladolid. Las características de este corpus cumplen exactamente con lo establecido en los objetivos (véase capítulo 5, Preparación del Corpus, apartado de Protocolo de Grabación), con lo cual, podemos asegurar que este objetivo se ha cumplido en su totalidad. Además, como apoyo, se ha recopilado un conjunto de ejercicios tipo que suelen aparecer en los exámenes para la obtención de los DELE, agrupados por niveles, y se ha investigado cuáles son los contenidos exactos, en relación con la entonación, que un candidato debe conocer para poder optar al diploma de un nivel concreto. De este

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

modo, tendríamos una idea de qué ejercicios podríamos incluir en una hipotética nueva versión de la aplicación, en caso de que deseáramos adaptarla para un nivel o un aspecto específico.

- *Preparación del corpus anterior para extraer de cada locución el conjunto de parámetros prosódicos que se vayan a emplear para caracterizar dicha locución. Para ello se analizarán diversas herramientas (SPTK, Praat, Snack...) existentes que facilitan la extracción de la secuencia de valores de la frecuencia fundamental (F0) de una señal hablada y a partir de dichas secuencias se elaborarán los parámetros adecuados según el tipo de unidad de habla elegida (sílabas, palabras, grupo acentual, grupo de entonación). Básicamente, se usarán en principio valores máximos, mínimos, medios y evolución temporal de las duraciones, pausas y niveles de F0. Se desea que las herramientas que extraigan las secuencias de valores de la frecuencia fundamental puedan ser ejecutadas en un servidor, y que el tiempo que tarden en procesar una locución sea razonable.*

Esta ha sido la tarea que más se ha prolongado durante la realización del proyecto. En el capítulo 5 de este documento hemos podido ver todo el trabajo realizado en relación con este objetivo y el estudio llevado a cabo para la toma de decisiones derivadas de los resultados obtenidos. Dentro del abanico de posibilidades de las que disponíamos para la elección de las herramientas que nos permitieran extraer parámetros que caracterizaran locuciones de habla humana, se optó por Praat y Snack, lo cual nos ha permitido comparar dos "puntos de vista" diferentes sobre el mismo conjunto de estudio (nuestro corpus). Ambas herramientas pueden ser ejecutadas como componentes en un sistema que corra en un servidor, por lo que desde esta perspectiva, este objetivo también se ha cubierto. La única incógnita es su tiempo de respuesta ante el análisis de locuciones muy largas, aunque en las pruebas realizadas con la aplicación parece que en la mayoría de los casos se comportan de la forma deseada.

- *Evaluación experimental del método de comparación de entonación. Para ello se analizarán diversos métodos de clasificación que permitan determinar cuándo dos locuciones de una misma frase son iguales o distintas, dentro de unos márgenes de error suficientemente pequeños. En este sentido, las pronunciaciones perceptualmente equivalentes de una misma frase, realizadas por el mismo locutor o por locutores diferentes, deberán ser clasificadas como iguales de forma que las que no resulten equivalentes a éstas estén lo más separadas que sea posible de ellas.*

Esta parte del estudio es la que pudimos llevar a cabo gracias a la ayuda de WEKA. Como se ha expuesto anteriormente, se consideraron diferentes alternativas para la clasificación de los ensayos, en busca de la que mayor porcentaje de acierto tuviera. Estas alternativas se tradujeron en experimentos sobre el conjunto de los datos globales y locales de las locuciones extraídos con Praat y Snack, atendiendo a diversos criterios para la elección de los pesos (sonoridad, sexo, sonoridad y sexo). En cuanto obtuvimos los pesos procedentes de cada experimento, los probamos para comprobar cuál aseguraba más aciertos para nuestro corpus. A raíz de este evaluación, se decidió utilizar la herramienta Snack y la ponderación deducida del análisis de los valores locales distinguiendo sexo y sonoridad con WEKA. En este caso, WEKA nos daba una estimación de clasificaciones correctas que oscilaba entre el 83% y el 90% con su método de correlación cruzada de 10 hojas. Con el método que hemos utilizado en este estudio, a nosotros nos resultó un valor cercano al 77%, lo cual, para nuestros intereses es ligeramente menor que a lo que aspirábamos. Sin embargo, esta baja tasa de acierto no quiere decir que hayamos obrado mal en nuestros experimentos, sino que ha podido verse perjudicada por la escasez de locuciones de hablantes nativos (tampoco disponíamos de repeticiones de las frases realizadas por los locutores de referencia) en comparación con la gran cantidad de locuciones de extranjeros. WEKA podría no tener la cantidad de información necesaria como para hacer una estimación más ajustada de los pesos. Otro factor que ha podido influir en este sentido es el propio método de prueba, si en lugar de considerar las distancias entre las primeras repeticiones de las frases pronunciadas por los extranjeros hubiéramos propuesto las segundas, o incluso terceras repeticiones, estos resultados podrían haber

cambiado. Por último, también podríamos haber realizado un tratamiento diferente sobre los elementos espúrios de las grabaciones. A priori, una opción sería computar los datos de manera que cada vez que se encontrar un elemento espúrio sus valor de F0 fuera sustituido por el resultante de utilizar interpolación con la regla del punto medio con los datos inmediato anterior y posterior. Esto quizás habría ayudado a mejorar la calidad de las distancias utilizadas como entrada para el algoritmo DTW. Aún así, puesto que nuestro objetivo apuntaba a realizar la evaluación empírica y no nos marcamos una cota del porcentaje de acierto deseado, podemos pensar en que una tasa de acierto de casi 4 de cada 5 usuarios es más que suficiente para una primera aproximación.

Como se ha expuesto, con mayor o menor éxito del esperado, los objetivos de este proyecto se han superado. Nos queda la duda de saber si los resultados hubieran sido los mismos si hubiéramos dispuesto de una gama más amplia de locuciones de los locutores de referencia, pero ya podemos extraer algunas conclusiones del trabajo realizado.

8.2. Conclusiones

La evaluación de la entonación en sistemas de enseñanza de un segundo idioma es una rama de estudio relativamente reciente. Podemos encontrar sistemas de reciente aparición del tipo del que se ha pretendido diseñar en este proyecto para otros idiomas. El tratar de hacer un trabajo parecido para nuestro idioma, debido a la importancia que éste tiene en el mundo, parece pues, una buena idea.

Sin embargo, la tarea del análisis de la voz con este propósito es una labor costosa y complicada. Se requieren unos conocimientos previos y un dominio de la materia, además de un sólida base lingüística, que pueden hacer que el trabajo sea muy difícil para una persona no relacionada con ello. Por tanto, se destaca que la etapa de documentación inicial es esencial para conocer cuáles son los aspectos fundamentales a tratar para poder comenzar la elaboración de un sistema con las características del que aquí se ha descrito.

En cuanto al proceso del trabajo realizado en sí, existen etapas en las que quizás hubiera sido ventajosa la presencia de un lingüista que aportara su opinión especializada en relación con aspectos puramente lingüísticos como el etiquetado de las frases (pudiendo así realizar un análisis detallado a nivel de palabra y de sílaba), la elección de un locutor de referencia ("Golden Speaker") para la aplicación, o para la evaluación del sistema, comparando sus resultados y puntuaciones con las proporcionadas por el especialista.

En comparación con el trabajo realizado en el proyecto del año pasado, "Aplicación Móvil de Ayuda a la Mejora de la Pronunciación del Español", en este proyecto se ha profundizado más en los aspectos relacionados con la caracterización del habla. Mientras que en aquél nos preocupábamos por el estudio de las tecnologías del habla, y en concreto por los resultados que nos pudiera ofrecer un reconocimiento de voz para contrastar con la frase original, en éste hemos descendido a un nivel más cercano a la voz, estudiando y analizando los conceptos que influyen en la descripción de una locución.

El sistema diseñado se trata de un módulo de evaluación de la entonación que pretende complementar la formación recibida en las aulas por los estudiantes de español. Los resultados de clasificación que hemos obtenido en una primera prueba se aproximan al 80% de acierto para los hablantes extranjeros, probablemente un poco alejados de nuestras expectativas iniciales, pero un buen resultado teniendo en cuenta que hemos entrenado al sistema con un 87.5% de muestras extranjeras y tan sólo un 12.5% de muestras nativas. Esto nos hace ser optimistas si pensamos en una evolución del sistema en que equiparemos las muestras nativas con las extranjeras. Seguramente la tasa de acierto será mayor y la clasificación más precisa y fiable.

8.3. Trabajo Futuro

Después de los resultados obtenidos y las conclusiones extraídas del trabajo realizado, se desprenden algunas propuestas de trabajo pendiente y posibles vías de ampliación del proyecto.

En primer lugar, quizás el punto al que se le debería dar más prioridad es a la adquisición de locuciones de hablantes nativos bajo las mismas condiciones, o similares, que las que se han utilizado para los locutores extranjeros que han participado aportando su voz para este corpus. El requisito mínimo indispensable es que al menos se disponga de varias repeticiones de las frases de los mismos locutores para poder considerar las distancias intra-persona de sus pronunciaciones. A partir de aquí, sería interesante repetir el estudio de los parámetros más relevantes con WEKA y observar las posibles diferencias. Es posible que de este modo la clasificación pronosticada por el módulo se ajustara más a la realidad. También sería interesante realizar un tratamiento distinto a los elementos espurios al que aquí se ha llevado cabo para comparar si existe una mejora en los resultados.

Además de esto, podríamos incluir al estudio otras herramientas como SPTK o HTK. Por una parte podría dificultar más las evaluaciones, pero sería enriquecedor de cara a la posible mejora del módulo y podríamos realizar un análisis más detallado comprobando diferencias y coincidencias entre los resultados obtenidos de los valores extraídos con las distintas herramientas, cada una con su (o sus) algoritmo (s) de extracción de la frecuencia fundamental.

Se recomienda además seguir trabajando en un método de evaluación eficaz que refleje de la forma más fiel posible el verdadero nivel de una entonación. En este trabajo se ha propuesto una que atiende a los criterios que nosotros hemos considerado, pero quizás existan mejores alternativas.

Otra posible vía de ampliación sería estudiar la posibilidad de que el módulo de evaluación "aprendiera", quizás mediante retroalimentación procedente de los propios usuarios, o mediante otras técnicas. El sistema podría recalcular y volver a estimar si sus parámetros son adecuados en cada uso, lo que en una situación ideal llevaría a que sus clasificaciones y puntuaciones fueran más reales. También es cierto que si facilitáramos ese "poder" a los usuarios, el sistema podría ser sensible frente a un uso malintencionado, si bien habría que protegerlo ante estas situaciones.

En cuanto al producto final, queda pendiente su optimización. Ahora mismo disponemos de una aplicación que recibe y envía grabaciones desde y hacia un servidor, respectivamente. Quizás esta propuesta esté más relacionado con la mejora de las tecnologías empleadas (conectividad, capacidad de cómputo del servidor, número de operaciones de los algoritmos de extracción...), pero se hace necesario analizar opciones que permitan aumentar y acelerar su rendimiento. Existe el riesgo de que ante locuciones de larga duración el hilo principal de la aplicación móvil quede bloqueado a la espera de la respuesta del servidor, y eso puede resultar desagradable al usuario. Esta situación se daría en el mejor de los casos, en el peor el cliente ni siquiera recibiría una respuesta o la aplicación se cerraría. Otra opción sería que el usuario pudiera enviar su ensayo y el sistema responderle más tarde de forma asíncrona. El usuario podría continuar utilizando su dispositivo sin tener bloqueada la UI a la espera, y recibiría sus puntuaciones en una especie de "buzón de entrada", siendo avisado mediante una notificación, al estilo de otras muchas famosas aplicaciones (sistemas de mensajería, por ejemplo).

Por otra parte, queda también pendiente el estudio y mejora de su interfaz. Se ha procurado diseñarla de un modo que fuera más atractiva para los usuarios, pero aún así exige un esfuerzo para incrementar la calidad de la experiencia del usuario.

Capítulo 8: Conclusiones y Trabajo Futuro

Por último, se propone la ampliación de la escala de la aplicación final. Actualmente se trata de un sistema por el cual un usuario puede hacer frente a diferentes desafíos propuestos por el sistema (supuestamente propuestos por especialistas). Incluir competencia podría suponer un aliciente para el usuario para esforzarse más en mejorar su dominio del español, de modo que tratara de superarse ya no sólo a sí mismo, sino también a otros usuarios. es por esto que podríamos pensar en una aplicación de tipo social en la cual los usuarios pudieran desafiarse unos a otros en busca de obtener la mayor puntuación. Si el sistema avanzara lo suficiente, incluso podríamos dotar a los usuarios de la capacidad de subir sus propios desafíos para otros usuarios y que el sistema los evaluara. Sería una buena forma de aumentar la base de datos del sistema para su estudio.

A - Competencias Requeridas y Ejercicios Tipo de Exámenes para la obtención del DELE

En la siguiente sección revisaremos los aspectos relacionados con la entonación que se encuentran en el inventario del apartado de "Pronunciación" del Plan Curricular del Instituto Cervantes, clasificando los contenidos por los niveles especificados, para poder elaborar un conjunto de ejercicios tipo que sirvan de ayuda para la mejora de la prosodia de los estudiantes del español como segunda lengua. Además, se han recopilado una serie de ejercicios procedentes de los diferentes modelos de examen según el nivel que también nos podrían resultar de interés.

A.1. Contenidos para la obtención del DELE por niveles en relación con la entonación

En este apartado veremos, clasificados por niveles, los contenidos que, atendiendo al Plan Curricular, deben conocer los candidatos que aspiren a obtener un diploma de un determinado nivel. Como podremos apreciar, dentro de cada nivel, a su vez, los contenidos se dividen en cuatro bloques: "Segmentación del discurso en unidades melódicas", "Identificación y producción de los patrones melódicos correspondientes a la entonación enunciativa", "Identificación y producción de los patrones melódicos correspondientes a la entonación interrogativa" e "Identificación y producción de los patrones melódicos correspondientes a los distintos actos de habla". Dentro de estos bloques es donde podremos comprobar las diferencias entre las competencias requeridas para cada nivel.

A.1.1. A1-A2

A.1.1.1. Segmentación del discurso en unidades melódicas

- **Correspondencia entre las unidades melódicas y el sistema de puntuación**
 - El punto
 - La coma
 - Signos de interrogación y exclamación

- **Extensión de la unidad melódica en la secuencia hablada: irregularidad relativa**
 - Unidad mínima: palabra monosilábica
¿Quién ha venido?
Yo.

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

- Segmentación en unidades más extensas [rasgos característicos del español]:
 - Unidades de entre cinco y diez sílabas: predominio de unidades de siete y ocho sílabas
Trabaja en un banco.
 - Escasez de unidades inferiores a cinco sílabas y de unidades superiores a diez o quince
¿Qué? / No creo que tengas razón en lo que dices.

- **Correspondencias sintácticas de la división de unidades melódicas: identificación y producción de patrones melódicos correspondientes a las estructuras sintácticas básicas**
 - Entonación del vocativo
Sara, ¿me dejas tus ejercicios?
 - Oraciones simples de no más de siete u ocho sílabas: una sola unidad
Me llamo Pedro.
 - Oraciones simples por encima de las quince sílabas: mínimo dos unidades
Vivo en Madrid. vs. Vivo en Madrid porque mi novio es español y porque me gusta España.
 - Enumeraciones: cada miembro, una unidad
Tengo un lápiz, un cuaderno, un libro de español.
 - Oraciones complejas coordinadas: unidades distintas
Quiero ir // pero no puedo.

A.1.1.2. Identificación y producción de los patrones melódicos correspondientes a la entonación enunciativa

- **Al principio de la unidad enunciativa**
 - Inflexión de la voz: tono grave de las sílabas o palabras átonas iniciales (en comparación con la primera sílaba fuerte del mismo grupo)
Escribo pocas palabras.
 - Tono medio desde el principio si el inicio es sílaba tónica.
Tengo dos hermanos.

- **En el cuerpo de la unidad enunciativa**
 - Entonación uniforme: carácter sobrio de la entonación española

- **Al final de la unidad enunciativa**
 - Localización de la zona de mayor relevancia significativa (siempre al final)
No te gusta el cine. / ¿No te gusta el cine?
 - Sílabas finales a partir de la sílaba que lleva el último acento:
 - Unidad finalizada en palabra aguda: ámbito del final melódico en una sílaba
Es malo para mi corazón.
 - Unidad finalizada en palabra llana: ámbito del final melódico en dos sílabas
Hace mucho que llegué a casa.
 - Unidad finalizada en palabra esdrújula: ámbito del final melódico en tres sílabas
Ese es su comportamiento típico.
 - Producción del tonema de cadencia: terminación absoluta de la enunciativa; tono grave
 - Tonema de anticadencia (ascenso 4 ó 5 semitonos)

- Tonema de suspensión (sin ascenso ni descenso)
- Importancia del valor expresivo en la inflexión final: reconocimiento de los tonemas
No es muy bueno (cadencia muy marcada: aseveración categórica) / *No es muy bueno* (semicadencia o cadencia menos marcada: aseveración dubitativa)

A.1.1.3. Identificación y producción de los patrones melódicos correspondientes a la entonación interrogativa

- Sensibilización hacia la gran variedad de entonaciones de las interrogativas españolas
- Entonación de la interrogativa absoluta frente a la entonación de la interrogativa relativa
- Entonación de la interrogativa introducida por pronombre interrogativo
¿Quién es tu profesor?
- Semejanzas y diferencias entre la entonación de una interrogativa simple y la primera parte de una enunciativa
¿Están estudiando? vs. *Están estudiando // desde las tres de la tarde.*

A.1.1.4. Identificación y producción de los patrones melódicos correspondientes a los distintos actos de habla

- La entonación como exponente de los distintos actos de habla: los orientados hacia el hablante (actos de habla asertivos, exclamativos y expresivos) y los orientados hacia el oyente (actos de habla interrogativos y actos de habla imperativos):
 - Actos de habla exclamativos y expresivos en sus formas básicas
 - La entonación en los saludos y expresiones de cortesía básicas:
Muchas gracias
Buenas tardes
Adiós
¿Qué tal?

A.1.2. B1-B2

A.1.2.1. Segmentación del discurso en unidades melódicas

- **Correspondencia entre las unidades melódicas y el sistema de puntuación**
 - Los dos puntos
 - El punto y coma
 - Los puntos suspensivos
 - El paréntesis y el corchete
- **Extensión de la unidad melódica en el diálogo: mayor grado de irregularidad**
- **Correspondencias sintácticas de la división de unidades melódicas: identificación y producción de patrones melódicos correspondientes a ciertas estructuras sintácticas**
 - Anteposición de los complementos: posibilidad de establecer más de una unidad
En el río // se veían barcos. vs. *En el río se veían barcos.*
 - Anteposición del verbo al sujeto: una unidad
Vino Isabel.

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

- Subordinadas (independientemente de su extensión): exigencia de división de unidades entre subordinante y subordinado
Si quieres // puedes venir.
- Entonación de cláusulas reducidas y construcciones en posiciones parentéticas
*La chica, cansada de llevar tacones, se descalzó en la fiesta.
Por lo que me dices, sin duda, todo salió muy bien.*

A.1.2.2. Identificación y producción de los patrones melódicos correspondientes a la entonación enunciativa

- **Al principio de la unidad enunciativa**
 - Inflexión de la voz: tono bajo o tono alto dependiendo, no ya de que la sílaba inicial sea una sílaba tónica o átona, porque puede ser tanto lo uno como lo otro, sino del valor gramatical de la palabra en cuestión.
Donde está escrito. / Dónde está escrito.
- **En el cuerpo de la unidad enunciativa**
 - Perfil básicamente horizontal del contorno tonal
- **Al final de la unidad enunciativa**
 - Producción de los tonemas con mayores matizaciones:
 - Cadencia (descenso 8 semitonos)
 - Semicadencia (descenso 3 ó 4 semitonos)
 - Anticadencia (ascenso 4 ó 5 semitonos)
 - Semianticadencia (ascenso 2 ó 3 semitonos)
 - Suspensión (sin ascenso ni descenso)
- **Identificación y producción de los patrones melódicos de la entonación enunciativa correspondiente a los diversos tipos de aseveración: modificaciones en la cadencia**
 - Aseveración ordinaria
Me levanto a las ocho.
 - Aseveración categórica
La pronunciación // es lo más difícil del español.
 - Aseveración insinuativa: relación con la pragmática (intencionalidad, contenidos implícitos)
Es una opinión muy lógica.
 - Aseveración con tono de incertidumbre
Como no sabía qué ibas a hacer tú.
 - Enumeraciones:
 - Enumeración completa final de frase
Sé francés, inglés, alemán y español.
 - Enumeración incompleta final de frase
Me pareció que era lenta, aburrida, pesada.
 - Enumeración descriptiva
Me levanto, me ducho, tomo un café y me voy a trabajar.
 - Enumeración distributiva
Corta por aquí este lado del papel, por este otro dóblalo, por el otro lado corta otra vez.

A.1.2.3. Identificación y producción de los patrones melódicos correspondientes a la entonación interrogativa

- Entonación de las interrogativas aseverativas
¿A que es muy guapo?
- Entonación de las interrogativas aseverativas con el rasgo de cortesía
¿Querriás hacerme el favor de venir?
- Entonación de las interrogativas alternativas o disyuntivas
¿Fue sin querer o fue aposta?

A.1.2.4. Identificación y producción de los patrones melódicos correspondientes a los distintos actos de habla

- La entonación como exponente de los distintos actos de habla: los orientados hacia el hablante (actos de habla asertivos, exclamativos y expresivos) y los orientados hacia el oyente (actos de habla interrogativos y actos de habla imperativos):
 - Entonación volitiva
 - El tonema desiderativo
 - Saludos y expresiones de cortesía: preguntas
 - Actos de habla interrogativos (diversidad de procedimientos entonativos para marcarlos)
 - Actos de habla imperativos
 - Patrones entonativos básicos de mandato
Ven.
¡Silencio!
Más despacio, por favor
 - Exclamaciones de mandato
¡Que te calles de una vez, hombre!
 - Entonación exclamativa de urgencia
¡Pasa! ¡Vamos!
 - Recomendaciones o mandatos atenuados
No fumes tanto.
 - Exclamaciones de mandato irónico
¡Tú hazle caso!
 - Saludos y expresiones de cortesía más complejas
¿Cómo está usted? ¿Qué tal te fue?

A.1.3. C1-C2

A.1.3.1. Segmentación del discurso en unidades melódicas

- **La entonación en la lectura**
 - Influencia del estilo y del registro del texto
 - Interpretación o captación de la intención implícita del autor
 - Principales diferencias entre la entonación de la lectura y la del habla espontánea
 - División de las unidades melódicas en la lectura sin signos de puntuación: subjetividad en la interpretación del texto

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

- **Efectos sintácticos de la división en unidades melódicas: identificación y producción de patrones melódicos correspondientes a determinadas estructuras sintácticas**
 - Casos de ambigüedad: motivados por la entonación o resueltos por la entonación
Es todo un hombre // de las cavernas vs. Los chicos (//) que estudian (//) aprobarán.
 - Producción de patrones melódicos de oraciones con orden marcado:
A María la atendieron muy bien [topicalización]
Mucho pides tú [focalización]

A.1.3.2. Identificación y producción de los patrones melódicos correspondientes a la entonación enunciativa

- **Al principio de la unidad enunciativa**
 - Inflexión inicial, tono enfático:
 - *Pero* inicial
PERO, si está nevando. ¡¿PERO qué dices?!
 - Reforzamiento de artículos, adverbios o preposiciones
¡¡POR amor de Dios!!
- **En el cuerpo de la unidad enunciativa**
 - Movimientos más amplios cuanto mayor grado de intensificación semántica se desee alcanzar
- **Al final de la unidad enunciativa**
 - Grados sutiles de diferenciación y oposición entre los tonemas:
 - Enumeración interior
En una situación así, las risas, la diversión, la animación terminan en tristeza.
 - Enumeración acumulativa
Su mirada, su sonrisa, su forma de gesticular, todo es maravilloso en ella.
 - Enumeración intensificativa
No encontré ni un apoyo, ni una ayuda, ni nada.
 - Enumeración calificativa
Volverás a los reproches, los insultos, las susceptibilidades, como siempre.
 - Enumeración valorativa
Y van tan orgullosas a fiestas, certámenes, ruedas de prensa y desfiles.
 - Enumeración enfática
Y me dan unos sudores y unos nervios y un estado de descontrol tremendos.

A.1.3.3. Identificación y producción de los patrones melódicos correspondientes a la entonación interrogativa

- Entonación de la interrogativa restrictiva: se pregunta sobre un elemento del enunciado
¿Eres tú el que me ha llamado?
- Entonación de la interrogativa reiterativa
¿Que si sé lo que digo?

- ¿Cuánto dices que te ha costado?
 - ¿Si lo supiera mi padre? Pues no quiero pensarlo.
- Entonación de la interrogativa exclamativa
 - ¿Qué te he dicho yo?
 - ¿A eso le llaman ellos trabajar?
- Entonación de la interrogativa hipotética
 - ¿Supongo que me llamarás?
- Entonación de la interrogativa compuesta por varios grupos melódicos
 - ¿No has visto nunca al mirar hacia el horizonte en un cielo despejado la línea que separa el cielo de la tierra?
- Entonación del paréntesis interrogativo
 - Me lo prometiste (¿lo has olvidado ya?) el mismo día de conocernos.
- Entonación del vocativo en posición parentética dentro de una interrogativa
 - ¿No te alegras, mujer, de esta sorpresa?

A.1.3.4. Identificación y producción de los patrones melódicos correspondientes a los distintos actos de habla

- La entonación como exponente de los distintos actos de habla: los orientados hacia el hablante (actos de habla asertivos, exclamativos y expresivos) y los orientados hacia el oyente (actos de habla interrogativos y actos de habla imperativos):
 - Actos de habla expresivos. La entonación afectiva y emocional
 - Entonación volitiva que incluya mandato, etc.
 - Exclamaciones desiderativas
 - ¿Lo que daría yo por que vinieras un día a verme!
 - Expresión exclamativa de la decepción
 - ¿Yo que creía que lo hacía por mí!
 - Exclamaciones de protesta
 - ¿Si yo siempre he sido correcta con él!
 - Exclamaciones de resignación
 - ¿Qué le vamos a hacer!
 - Exclamaciones de reproche
 - ¿Buena la hemos hecho con invitarle!
 - Exclamaciones de extrañeza y sorpresa
 - ¿Que no haya venido a verme!
 - Exclamación ondulada (estados de ánimo muy marcados)
 - ¿Mira que te conozco, Carlos, que te conozco!
 - Exhortación persuasiva para dar ánimo
 - Verás como no es nada.
 - Exhortación con ímpetu
 - ¿Hay que trabajar más!
 - Expresión exhortativa de lamentación hipotética y no hipotética
 - ¿Si no fuera por tu ayuda!
 - ¿Vaya día me espera!
 - Patrones entonativos básicos de ruego o súplica
 - Por favor [indicando ruego]
 - En locuciones gramaticales
 - Desearía, quisiera, le agradecería.
 - Entonación en la expresión de ruego en una pregunta
 - ¿Por qué no me puedes tratar como a una persona?

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

- El énfasis articulatorio
 - Énfasis de intensidad
¿Dices que has visto A QUIÉN?
 - Énfasis de entonación (contrastivo)
Ya te he dicho que eres TÚ quien debe disculparse.
 - Énfasis interrogativo
¿QUE ESE TIPO te ha pedido el teléfono?
 - Énfasis de gravedad
Tú -CRÉEME- no tienes la culpa.
 - Falta de correspondencia entre la entonación y la palabra: entonación irónica
Me gusta mucho el bolso que te has comprado [sugiriendo que no le gusta nada, por lo tanto suprimiendo todo movimiento tonal intensificador]
 - Inflexiones circunflejas propias de los enunciados con gran carga afectiva o emocional
Con deciros / que cada vez que llamo / me salta el contestador [Tres unidades melódicas con perfiles entonativos no «rectos», sino ascendentes y descendentes, esto es, circunflejos]
- Saludos y expresiones de cortesía más elaborados
Que sea enhorabuena.
Que lo paséis muy bien.

A.2. Ejercicios tipo de los exámenes para la obtención del DELE

Como deseamos diseñar un módulo de evaluación de la entonación, necesitamos “entrenar” el sistema con un corpus que represente lo suficientemente bien los aspectos que deseamos que el usuario practique. Una buena fuente para extraer ejemplos de ejercicios que pudiéramos plantear a los estudiantes, son los exámenes para la obtención del DELE. Por ello, se ha procurado recoger algunos ejercicios, agrupados por niveles, que suelen aparecer en estas pruebas.

A.2.1. A1

- Tarea 1: (1-2 minutos)
 - INSTRUCCIONES:
En esta tarea el candidato debe proporcionar información muy básica sobre sí mismo y sobre su entorno inmediato: identidad, personas conocidas, pertenencias, lugares, actividad académica o profesional, etc., mediante enunciados breves y sencillos.

La tarea consiste en un monólogo sostenido breve, previamente preparado, sobre datos de identidad personal.

El candidato dispone de una lámina con estímulos verbales, centrados en los datos de identidad que debe proporcionar el candidato.
- Tarea 2: (2-3 minutos)
 - INSTRUCCIONES:
En esta tarea el candidato debe proporcionar información muy básica sobre sí mismo y sobre su entorno inmediato: identidad, personas conocidas, pertenencias,

lugares, actividad académica o profesional, etc., mediante enunciados breves y sencillos.

La tarea consiste en un monólogo sostenido breve, previamente preparado, sobre un tema del entorno personal inmediato.

El candidato dispone de una lámina con estímulos verbales, centrados en diferentes aspectos del tema y dispuestos gráficamente en forma de menú de opciones.

▪ Tarea 3: (3-4 minutos)

• INSTRUCCIONES:

En esta tarea el candidato debe participar en conversaciones breves, con un interlocutor operativo, en relación con la información previamente proporcionada.

La tarea consiste en conversar con el examinador, sobre la información previamente proporcionada.

▪ Tarea 4: (2-3 minutos)

• INSTRUCCIONES:

En esta tarea el candidato debe satisfacer necesidades inmediatas, en el transcurso de transacciones o contactos sociales, en que el candidato saluda o se despide, realiza o responde a peticiones, plantea o responde a preguntas...si se le habla despacio y con claridad.

La tarea consiste en una serie de intercambios conversacionales muy breves (dos o tres turnos) entre candidato y entrevistador, que ambos inician o continúan de forma alterna.

El candidato dispone de cuatro láminas que aportan estímulos visuales y verbales muy sencillos, a partir de los cuales, candidato y entrevistador inician el turno de palabra o reaccionan.

A.2.2. A2

▪ Tarea 1: (3-5 minutos)

• INSTRUCCIONES:

En esta tarea el candidato debe describir, en una presentación oral breve, ensayada y con términos sencillos, asuntos y experiencias de un aspecto concreto de su vida cotidiana.

La tarea consiste en un monólogo sostenido breve, previamente preparado, a partir de un tema que el candidato elige entre dos opciones.

El candidato dispone de una lámina con estímulos verbales que desglosan diversos aspectos del tema sobre los que se puede desarrollar su monólogo.

▪ Tarea 2: (2-3 minutos)

• INSTRUCCIONES:

En esta tarea el candidato debe describir de manera breve y sencilla los elementos de una escena de la vida cotidiana en la que se reflejan asuntos de tipo práctico, como pueden ser las compras, el uso de medios de transporte, etc.

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

La tarea consiste en un monólogo sostenido breve, previamente ensayado, a partir de una fotografía.

El candidato dispone de una fotografía que refleja una situación de la vida cotidiana y de indicaciones sobre el contenido de la descripción que debe realizar.

▪ Tarea 3: (2-3 minutos)

• INSTRUCCIONES:

En esta tarea el candidato debe participar en conversaciones breves, de tipo transaccional, con el fin de satisfacer necesidades de supervivencia básica.

La tarea consiste en conversar con el examinador, tras una planificación previa, simulando una situación a partir de la fotografía descrita en la tarea 2.

El candidato dispone de la fotografía de la tarea 2 y de indicaciones sobre el contenido y el desarrollo de la conversación.

▪ Tarea 4: (2-4 minutos)

• INSTRUCCIONES:

En esta tarea el candidato debe participar en conversaciones informales breves y sencillas, que tienen lugar en intercambios sociales de tipo cotidiano y predecible, en que se tratan temas de interés para el candidato o asuntos de la vida diaria: intercambios sencillos y directos de información, opiniones y puntos de vista sencillos, gustos, deseos y preferencias, planes, experiencias, sugerencias y propuestas, etc.

La tarea consiste en conversar con el examinador simulando una situación determinada a partir de una tarjeta de rol.

El candidato dispone de una tarjeta de rol con estímulos verbales o visuales muy sencillos, como iconos o fotografías, en relación con el tema de la conversación y el rol que debe adoptar.

A.2.3. B1

▪ Tarea 1: (2-3 minutos + 15 minutos de preparación)

• INSTRUCCIONES:

En esta tarea el candidato debe describa sus experiencias o hablar de sus deseos respecto a un tema o un titular que se le proporciona previamente.

La tarea consiste en un monólogo sostenido breve, previamente preparado, a partir de un tema y unas preguntas. Dos opciones, de las que debe elegir una para preparar.

El candidato dispone de una lámina con un tema o un titular (de entre 80 y 100 palabras) y preguntas para pautar su respuesta.

▪ Tarea 2: (3-4 minutos)

• INSTRUCCIONES:

En esta tarea el candidato debe conversar con el entrevistador sobre el tema de la Tarea 1. El entrevistador preguntará al candidato por su opinión o su experiencia personal.

La tarea consiste en una conversación entre el candidato y el entrevistador a partir de la Tarea 1.

El candidato dispone de unas indicaciones sobre el desarrollo de la conversación.

▪ Tarea 3: (2-3 minutos)

• INSTRUCCIONES:

En esta tarea el candidato debe describir una fotografía, siguiendo las pautas establecidas, y responder a las preguntas del entrevistador, que relacionarán la imagen con el entorno del candidato.

La tarea consiste en una breve presentación de una imagen elegida entre dos opciones y posterior conversación con el entrevistador, sin una planificación previa.

El candidato dispone de una lámina con una fotografía y unas pautas de intervención.

▪ Tarea 4: (2-3 minutos)

• INSTRUCCIONES:

En esta tarea el candidato debe participar en conversaciones para satisfacer necesidades cotidianas o intereses personales: hacer cambios o devoluciones, solicitar un servicio, hacer una queja o una consulta, confirmar o concertar una cita, solicitar una información, quedar con amigos, etc.

La tarea consiste en conversar con el examinador simulando una situación cotidiana, a partir de la fotografía de la Tarea 3.

El candidato dispone de tarjetas de rol que proporcionan la información que debe conocer para contextualizar la situación.

A.2.4. B2

▪ Tarea 1: (4-5 minutos + 15 minutos de preparación)

• INSTRUCCIONES:

En esta tarea el candidato debe valorar las ventajas y desventajas de una serie de propuestas dadas para resolver una situación problemática, e intercambiar opiniones acerca de ese tema.

La tarea consiste en un monólogo sostenido breve, previamente preparado, y una conversación sobre una serie de propuestas para resolver un problema. Dos opciones, de las que debe elegir una para preparar.

El candidato dispone de una lámina con una situación problemática y de cinco a siete propuestas para solucionarla o mejorarla; con un breve ejemplo de intervención.

▪ Tarea 2: (3-4 minutos)

• INSTRUCCIONES:

En esta tarea el candidato debe describir una situación a partir de un enunciado, una foto y unas pautas dadas, y conversar brevemente sobre experiencias y opiniones respecto a ese tema.

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

La tarea consiste en un breve monólogo sostenido sobre una situación, previamente preparado, y una conversación sobre la misma. Dos opciones, de las que debe elegir una para preparar.

El candidato dispone de una lámina con una foto, un enunciado que describe la situación y unas pautas para su intervención.

- Tarea 3: (3-4 minutos)
 - INSTRUCCIONES:
En esta tarea el candidato debe conversar con el entrevistador a partir de un estímulo escrito o gráfico.

La tarea consiste en una breve conversación informal, sin preparación previa, a partir de un estímulo escrito o gráfico. Dos opciones, de las que hay que elegir una.

El candidato dispone de una o dos láminas con las instrucciones de la tarea y el estímulo para conversar (resultados de encuestas, noticias sobre tendencias contemporáneas, etc.).

A.2.5. C1

- Tarea 1: (3-5 minutos + 20 minutos de preparación)
 - INSTRUCCIONES:
En esta tarea el candidato debe comprender la información fundamental y más relevante de textos escritos extensos y complejos, y resumirlos y valorarlos de forma oral.

La tarea consiste en hacer un monólogo sostenido breve, previamente preparado, a partir de un texto escrito.

El candidato dispone de un texto escrito de aproximadamente 800 palabras perteneciente al ámbito académico, profesional o público, y de indicaciones sobre el contenido o el tipo de monólogo que tiene que enunciar.

- Tarea 2: (4-6 minutos)
 - INSTRUCCIONES:
En esta tarea el candidato debe intervenir en una conversación argumentando su postura y respondiendo de manera fluida a preguntas, comentarios y argumentaciones contrarias de carácter complejo.

La tarea consiste en participar en un debate formal, sobre el monólogo que el candidato ha llevado a cabo en la Tarea 1, a partir de una serie de preguntas e intervenciones del entrevistador.

El candidato dará respuesta a las preguntas y los comentarios del examinador.

- Tarea 3: SIMULACIÓN (4-6 minutos)
 - INSTRUCCIONES:
En esta tarea el candidato debe intercambiar ideas, expresar y justificar opiniones o hacer valoraciones con el fin de negociar y llegar a un acuerdo con su interlocutor.

A partir de un material de entrada (fotografías, anuncios, breves currículums, etc.), el candidato debe mantener una conversación de ámbito académico, profesional o público con el entrevistador, con el fin de llegar a un acuerdo entre los dos en tareas como la selección de una fotografía para un concurso, candidatos para un trabajo o artículos de cualquier tipo para algún fin concreto.

El candidato dispone de una lámina con una contextualización de la tarea, material que refleja diferentes aspectos de un mismo tema u opciones para un determinado objetivo, indicaciones sobre el asunto en el que deben ponerse de acuerdo y criterios que debe tener en cuenta en su elección.

A.2.6. C2

▪ Tarea 1: (6-8 minutos + 30 minutos de preparación)

• INSTRUCCIONES:

En esta tarea el candidato debe comprender y transferir información de dos o tres textos (entre 700 y 800 palabras en total) y una o dos imágenes (gráficos, esquemas, humor gráfico, etc.), que expresan diferentes aspectos, circunstancias u opiniones sobre un mismo tema (pueden ser contrarios o no, pueden tener informaciones diferentes y registros diferentes).

La tarea consiste en hacer un monólogo sostenido en el que el candidato utilizará los materiales que se le han proporcionado como base para elaborar un discurso adaptado al nivel de formalidad que requiere una presentación preparada en una situación de examen.

El candidato dispone de textos escritos y estímulos gráficos pertenecientes al ámbito académico, profesional o público, y de indicaciones sobre el contenido o el tipo de monólogo que tiene que enunciar.

▪ Tarea 2: (5-6 minutos)

• INSTRUCCIONES:

En esta tarea el candidato debe intervenir en una conversación con el entrevistador, respondiendo de manera fluida a preguntas, argumentos y peticiones de aclaración o detalles por parte de su interlocutor.

La tarea consiste en participar en un debate formal, sobre el monólogo que el candidato ha llevado a cabo en la Tarea 1, puntualizando y desarrollando la información de la presentación que ha llevado a cabo en la tarea anterior.

El candidato dará respuesta a las preguntas y los comentarios del examinador.

▪ Tarea 3: SIMULACIÓN (4-6 minutos)

• INSTRUCCIONES:

En esta tarea el candidato debe participar en una conversación informal con el entrevistador.

A partir de una selección de titulares de periódicos (entre tres y cuatro) que comentan aspectos de un mismo tema, el candidato debe improvisar una

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

conversación informal con el entrevistador con el fin de intercambiar opiniones personales en torno a un tema.

El candidato dispone de una lámina con unos titulares de periódicos.

B - Reconocimiento de Voz con Google Voice

Durante el m transcurso de la memoria, nos hemos preocupado por el análisis de la entonación puesto que es el tema principal del trabajo. Sin embargo, puesto que en el prototipo final hemos modificado un aspecto tan importante para la evaluación de la pronunciación como es el sistema de reconocimiento de voz, debemos realizar un breve apunte aunque sea en este apéndice.

Cabe destacar que seguimos utilizando el mismo motor para el reconocimiento, Google Voice, pero la diferencia es que en este caso no lo hacemos lanzando directamente en streaming la grabación de nuestra voz al reconocedor, quedando así inhabilitados para mantenerla y poder hacer un estudio de sus características prosódicas, sino que hemos tenido que buscar una alternativa para poder optar a trabajar ambos aspectos.

El script con las órdenes que hemos utilizado en el servidor para conseguir mantener la grabación de la locución de un ensayo y además obtener los resultados de reconocimiento automático del habla de Google Voice (incluidos sus parámetros de estimación de confianza) es gv.sh, el cual se ejecuta en el código de uploads.php.

Al llegar la grabación al servidor, lo hace en formato ".3gp" desde el cliente móvil Android. Debemos convertirla a un formato que reconozca Google Voice., ".flac". Con tal motivo se ha utilizado el software ffmpeg, capaz de convertir entre formatos de audio. Además, hemos aprovechado para convertir también a formato ".wav" para el análisis de la entonación que también se realiza posteriormente. Las líneas concretas con las que convertimos la grabación enviada por el usuario a formato ".flac" son las siguientes:

```
ruta=`echo $1 | cut -d"." -f2`
```

Quitamos al nombre del archivo recibido en \$1 la extensión.

```
rgoogle=`echo $ruta.flac`  
rproceso=`echo $ruta.wav`
```

Creamos el nombre del archivo con la extensión ".flac", que pasaremos como parámetro en la siguiente línea (\$rgoogle). Hacemos lo mismo para el archivo con extensión ".wav".

```
ffmpeg -y -i $rproceso -ar 16000 -f flac $rgoogle
```

Finalmente realizamos la conversión: la opción "-y" es para forzar al programa a que realice la conversión sin preguntar; -i indica el fichero que recibe como entrada (input file); -ar especifica la frecuencia de muestreo del audio (fundamental para el correcto funcionamiento del script); y con -f señalamos qué codificación deseamos como formato de salida, para nuestro cometido, ".flac".

Una vez hemos convertido el archivo, debemos enviarlo al servicio de Google Voice para su reconocimiento. se trata de una petición HTTP POST que llevaremos a cabo con el siguiente comando:

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

```
wget -q -U "Mozilla/5.0" --post-file $rgoogle --header="Content-Type: audio/x-flac; rate=16000" -O - "http://www.google.com/speech-api/v1/recognize?lang=es-es&client=chromium" > $rgRetorno
```

Es importante especificar el idioma para el reconocimiento en la etiqueta "lang=" que aparece en la dirección url a la que enviamos la grabación. Los resultados los obtendremos codificados en formato json en \$rgRetorno, que como hemos comentado en el cuerpo de la memoria, será un archivo con formato ".ret".

```
{"status": 0,"id": "b3447b5d98c5653e0067f35b32c0a8ca-1","hypotheses": [{"utterance": "i like pickles","confidence": 0.9012539},{utterance": "qué sería de una navidad sin su cesta"}]}
```

Esencialmente nos interesan dos campos de la respuesta del servicio Google Voice: "utterance", que son las palabras que ha reconocido en el archivo de audio; y "confidence", un valor normalizado entre 0 y 1 que indica la estimación de confianza del reconocimiento.

C - Soporte y Desarrollo

Para la elaboración de este trabajo se han requerido diferentes componentes hardware y software que se detallan a continuación.

C.1. Hardware

Las máquinas que se han utilizado para las tareas de investigación, elaboración de documentos intermedios y desarrollo de las aplicaciones que ilustren los conceptos estudiados han sido las siguientes:

Ordenador de sobremesa (Ordenador del laboratorio 2L018)	
SO	Windows 7 Professional (64 bit)
CPU	Intel Core 2 Quad Q6600 @ 2.40 GHz
VRAM	NVIDIA GeForce 7100 1 GB
Memoria	3GB DDRII
Disco duro	500 GB

Tabla 91: Características Equipo 1

Notebook Samsung RC530	
SO	Windows 7 Home Premium (64 bit) / Ubuntu 12.04 (64 bit)
CPU	Intel Core i7 2670QM
VRAM	NVIDIA GT540 gDDR3 2 GB Hybrid
Memoria	6 GB
Disco duro	750 GB

Tabla 92: Características Equipo 2

Ordenador de sobremesa Asus R5QL PRO	
SO	Windows 7 Professional (64 bit) / Ubuntu 12.04 (64 bit)
CPU	Intel Core Quad CPU Q8300
VRAM	NVIDIA GeForce GT 220 1 GB
Memoria	4 GB
Disco duro	500 GB

Tabla 93: Características Equipo 3

Los dispositivos sobre los que se ha probado la aplicación creada se corresponden con un teléfono móvil personal y una tablet proporcionada por el tutor del proyecto:

Samsung Galaxy SIII I-9300	
SO	Android 4.1.2

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

CPU	1.4 GHz Quad Core
Memoria	1 GB
Almacenamiento Externo	16 GB

Tabla 94: Características Equipo 4

Motorola XOOM (MZ604)	
SO	Android 3.2 (Actualizado desde la versión 3.2)
CPU	1 GHz Dual Core
Memoria	1 GB
Almacenamiento Externo	32 GB

Tabla 95: Características Equipo 5

C.2. Software

La principal fuente de documentación ha sido la web. La búsqueda de la información pertinente se llevó a cabo mediante el navegador de internet Google Chrome. La visualización de los documentos con formato ".pdf" se efectuaba con Adobe Reader XI.

Tal como se ha indicado en el capítulo Preparación de Corpus (capítulo 5), para el procesamiento de las grabaciones de audio y la extracción de los valores característicos de las locuciones contenidas en las mismas, nos hemos servido de Praat 5.3.45 y Snack 2.2.10. El análisis de los datos extraídos fue resuelto con el uso del software WEKA, versión 3.6.9.

Para la conversión entre formatos de audio en el componente servidor de la aplicación final se ha empleado ffmpeg. Recientemente se ha descubierto que es posible que se deba substituir por otro software que realiza la misma función llamado avconv.

Por otra parte, para la edición de los programas que se han desarrollado durante la preparación del corpus, se han utilizado los clientes ssh XShell 4 y Xming 7.5.0.47. El paso de archivos entre la máquina en la que se trabajaba localmente y los servidores remotos tauro y aries los llevamos a cabo con un cliente FileZilla con versión 3.7.1.1. Además, con el fin de lograr una visualización más cómoda de los ficheros de datos que se iban generando en la preparación del corpus, optamos por el software notepad++ versión 5.9.

El software empleado para la implementación de la aplicación que posteriormente fue testada en los terminales señalados previamente fue el entorno de desarrollo Eclipse Indigo con el plugin Android Development Tools (ADT) versión 16.0.1. La versión de la SDK de Android es la r18.

Para compartir los archivos (tanto documentos como proyectos de Eclipse) con el tutor se ha utilizado una carpeta compartida mediante el software Dropbox versión 1.2.52. Además, para la redacción de esta memoria nos hemos servido de Microsoft Word 2007 y 2010. Las capturas de pantalla fueron recortadas con Microsoft Paint, incluido en cualquier versión del sistema operativo Microsoft Windows (en nuestro caso, Windows 7). Por último, los diagramas de clases, de casos de uso y de secuencia que aparecen en los capítulos de contenido más técnico han sido realizados mediante el programa StarUML versión 5.0.2.

D - Glosario

A

Acento (stress): se refiere a la prominencia (elevación de algo sobre lo que está alrededor o cerca de ello) perceptual relativa de las unidades del habla (discurso) mayores que segmentos fonémicos.

Acento contrastivo: se refiere a la posición de las sílabas prominentes para aclaraciones sintácticas.

Acento enfático: se refiere a las sílabas que se hacen prominentes conscientemente para poner énfasis en una palabra partícula.

Acento gramatical: se refiere a los patrones de acento léxico que aparecen en frases o en habla conectada, que se rigen por la gramática del idioma. El uso impropio de las fuerzas relativas de este tipo de acento puede resultar en ambigüedades sintácticas.

Acento léxico: se refiere a la posición de las sílabas relativamente prominentes en palabras habladas de forma aislada.

Activity: representa el componente principal de la interfaz gráfica de una aplicación Android. Se puede pensar en una Activity como el elemento análogo a una ventana en cualquier otro lenguaje visual.

Alófonos: cada una de las variantes que se dan en la pronunciación de un mismo fonema según la posición de éste en la palabra o la sílaba, según el carácter de los fonemas vecino, por ejemplo la b oclusiva de tubo y la b fricativa de tubo

Android: sistema operativo móvil basado en Linux, que junto con aplicaciones middleware, está enfocado para ser utilizado en dispositivos móviles como smartphones, tablets, Google TV y otros dispositivos.

Android Development Tools (ADT): plugin para Eclipse IDE diseñado para proporcionar al desarrollador un entorno en el que construir sus aplicaciones Android.

API: conjunto de métodos que ofrece una biblioteca para ser utilizada por otro software como capa de abstracción.

Aplicaciones: programa informático diseñado como herramienta para permitir a un usuario realizar uno o diversos tipo de trabajo. En el marco de nuestro trabajo se referirá a los programas que ejecutamos en los dispositivos móviles Android.

C

CALL: Computer-Assisted Language Learning. Rama de la Ingeniería Lingüística que consiste en la aplicación de distintos enfoques de enseñanza de lenguas utilizando como soporte una aplicación informática.

Corpus: conjunto estructurado de textos que constituyen una muestra lo más realista posible del uso de la lengua.

D

DELE: Diplomas del Español como Lengua Extranjera.

Difonemas: segmento acústico que incluye la transición entre dos fonemas consecutivos, formado por la parte estacionaria del primero, la transición del primero al segundo y la parte estacionaria del tercero.

Distancia: separación existente entre las ondas de las señales acústicas de un ensayo y una locución de referencia.

DTW (Dynamic Time Warping, algoritmo): algoritmo de alineamiento que nos ayuda a relacionar los segmentos similares de dos locuciones, aunque estos se produzcan en instantes temporales diferentes, es decir, a encontrar la mínima distancia (o el camino más corto) entre ellos.

Duración: longitud de un componente particular del discurso.

E

Eclipse: entorno de desarrollo integrado de código abierto multiplataforma para desarrollar aplicaciones.

Energía: en la curva de la señal acústica, podemos ver la energía como la amplitud, esto es, la distancia entre la posición de reposo y el punto de máximo desplazamiento.

Ensayo: intento de pronunciación de una frase.

Entonación: modulación de la voz en la secuencia de sonidos del habla que puede reflejar diferencias de sentido, de intención, de emoción y de origen del hablante, y que, en algunas lenguas, puede ser significativa

F

Fonación: proceso mediante el cual se produce en la voz, resultado de la acción de la laringe

Fonética: todo aquello que es perteneciente o relativo a la voz humana. Dicho de un alfabeto, de una ortografía o de un sistema de transcripción se refiere a la representación de los sonidos con mayor exactitud que la ortografía convencional. Conjunto de los sonidos de un idioma.

Formantes: rasgos identificables de un sonido o fonema.

Frame: ventana de análisis utilizada durante el proceso de extracción de la frecuencia fundamental del habla.

Frecuencia Fundamental del Habla (F0): proporción de pulsos glotales generados por la vibración de las cuerdas vocales durante la reproducción en voz de los segmentos de una unidad entonativa.

G

Glottis: espacio limitado por las cuerdas vocales dentro de la laringe.

Google, Inc: empresa propietaria de la marca Google, cuyo principal producto es el motor de búsqueda de contenido en internet del mismo nombre. Compró Android Inc. en 2005.

Grabación: proceso de capturar datos o convertir la información a un formato almacenado en un medio de almacenamiento. El resultado del proceso se denomina con el mismo nombre.

Grafemas: unidad mínima e indivisible de una lengua, pueden ser letras, números u otros signos lingüísticos.

I

Ingeniería lingüística: aplicación de las técnicas informáticas al desarrollo de aplicaciones que incluyen componentes relacionados con el tratamiento del lenguaje y del habla.

Inteligible: que puede ser entendido, que se oye clara y distintamente.

L

Levenshtein (algoritmo): algoritmo que calcula un valor entero que sugiere una distancia entre dos cadenas. La distancia es el mínimo número de operaciones que hay que realizar para transformar una cadena en la otra, y se calcula contando las transformaciones que es necesario hacer sobre una de estas cadenas para obtener la otra. Las posibles transformaciones son borrado e inserción de un carácter y substitución de un carácter por otro.

Locución: acto de hablar.

M

Modelo de la fuente y filtro: el aparato fonador puede considerarse en términos de una fuente, el lugar donde se produce la corriente de aire indispensable para la producción de sonido; y un filtro, el conjunto de cavidades que, por el fenómeno de la resonancia, modifican las características de la fuente.

N

Normalizado (tiempo): aplicado a los puntos temporales en los que se produce máximo de frecuencia o energía en un intervalo, nos referimos al valor del instante temporal en el que sucede el punto singular correspondiente, dividido por el tiempo total del intervalo considerado.

O

Onda sonora: producida en la fonación, es el resultado del paso de aire por la glotis en la emisión de una serie de sucesivas bocanadas de aire al ritmo de abertura y cierre de las cuerdas vocales. Las moléculas de aire vibran al pasar por los pliegues vocálicos y se produce el sonido. Puede definirse en función de la duración del movimiento de vibración, su frecuencia y su energía.

P

Plan Curricular: elaborado por el Instituto Cervantes, es un documento de referencia que desarrolla y fija los niveles de referencia para el español según las recomendaciones que, en 2001, publicó el Consejo de Europa en su Marco Común Europeo de Referencia para las lenguas en cuanto a aprendizaje, enseñanza, evaluación.

Ponderación: asignación de pesos a los parámetros del vector que caracteriza una locución.

Praat: software para el análisis científico del habla utilizado en lingüística desarrollado por Paul Boersma y David Weenink en el Instituto de Ciencias Fonéticas de la Universidad de Amsterdam.

Pronunciación: emitir y articular sonidos para hablar.

Prosodia: parte de la gramática que enseña la recta pronunciación y acentuación.

R

Reconocimiento de voz: parte de la Inteligencia Artificial que tiene como objetivo permitir la comunicación hablada entre seres humanos y computadoras electrónicas. Lo que se pretende con esta tecnología es transformar una señal sonora - el habla - en su correspondiente representación simbólica que, en general, será un texto escrito.

S

Segmental (plano): hace referencia a cualquier unidad discreta que se puede identificar en la cadena hablada, ya sea por medios físicos o auditivos.

Segunda lengua: cualquier idioma aprendido por un individuo, después de haber adquirido su lengua materna (o primera).

Sílaba: sonido o sonidos articulados que constituyen un solo núcleo fónico entre dos depresiones sucesivas de la emisión de voz

Síntesis de la voz: generación automática de mensajes orales, partiendo de un texto escrito, en la denominada conversión de texto en habla, o de otros tipos de representación simbólica

Smartphones: teléfono móvil construido sobre una plataforma informática móvil, con una mayor capacidad de computación y conectividad que un teléfono móvil convencional.

Snack: software para el análisis científico del habla utilizado en lingüística desarrollado por el KTH Royal Institute of Technology (Suecia).

Suprasegmental (plano): rama de estudio dedicada a los fenómenos fonéticos que implican más de un fonema o segmento.

T

Tecnologías de lenguaje escrito: tecnologías que se centran en la lengua escrita y que se engloban en el campo conocido como procesamiento del lenguaje natural.

Tecnologías del lenguaje: también conocidas como tecnologías lingüísticas o tecnologías para el lenguaje humano (TLH o HLT, Human Language Technologies), son todas aquellas que se integran en aplicaciones informáticas para permitir el tratamiento de textos escritos o el procesamiento del habla.

Tecnologías vocales: tecnologías cuyo objeto es el habla.

U

Unvoiced: muestras o intervalos sordos dentro de una señal acústica.

V

Voiced: muestras o intervalos sonoros dentro de una señal acústica.

Voz: sonido generado por el aparato fonador humano.

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

Voz sintética: voz artificial (no pregrabada) generada mediante un proceso de sintetización del habla.

W

WEKA (Waikato Environment for Knowledge Analysis): plataforma software para el aprendizaje automático y minería de datos escrito en Java y desarrollado en la Universidad de Waikato, Nueva Zelanda.

E - Índice de Tablas

TABLA 1: FASES DEL PROYECTO.....	21
TABLA 2: LOCUTORES DE REFERENCIA DEL CORPUS Y SUS CORRESPONDIENTES SEXOS, IDENTIFICADORES Y DIRECTORIOS.....	44
TABLA 3: LOCUTORES EXTRANJEROS DEL CORPUS Y SUS CORRESPONDIENTES SEXOS, NACIONALIDADES, IDENTIFICADORES, NICK NAMES, Y DIRECTORIOS.....	45
TABLA 4: CONJUNTO DE FRASES QUE COMPONEN EL CORPUS	46
TABLA 5: MATRIZ DE CONFUSIÓN MODALIDAD 1	95
TABLA 6: MATRIZ DE CONFUSIÓN MODALIDAD 2 – LOCUTORES FEMENINOS.....	96
TABLA 7: MATRIZ DE CONFUSIÓN MODALIDAD 2 – LOCUTORES MASCULINOS.....	97
TABLA 8: MATRIZ DE CONFUSIÓN MODALIDAD 3	98
TABLA 9: MATRIZ DE CONFUSIÓN MODALIDAD 4 – LOCUTORES FEMENINOS.....	99
TABLA 10: MATRIZ DE CONFUSIÓN MODALIDAD 4 – LOCUTORES MASCULINOS.....	100
TABLA 11: MATRIZ DE CONFUSIÓN MODALIDAD 5 – INTERVALOS VOICED	101
TABLA 12: MATRIZ DE CONFUSIÓN MODALIDAD 5 – INTERVALOS UNVOICED.....	102
TABLA 13: MATRIZ DE CONFUSIÓN MODALIDAD 6 – LOCUTORES FEMENINOS E INTERVALOS VOICED	104
TABLA 14: MATRIZ DE CONFUSIÓN MODALIDAD 6 – LOCUTORES FEMENINOS E INTERVALOS UNVOICED	105
TABLA 15: MATRIZ DE CONFUSIÓN MODALIDAD 6 – LOCUTORES MASCULINOS E INTERVALOS VOICED	106
TABLA 16: MATRIZ DE CONFUSIÓN MODALIDAD 6 – LOCUTORES MASCULINOS E INTERVALOS UNVOICED	107
TABLA 17: MATRIZ DE CONFUSIÓN MODALIDAD 7	108
TABLA 18: MATRIZ DE CONFUSIÓN MODALIDAD 8 – LOCUTORES FEMENINOS.....	109
TABLA 19: MATRIZ DE CONFUSIÓN MODALIDAD 8 – LOCUTORES MASCULINOS.....	110
TABLA 20: MATRIZ DE CONFUSIÓN MODALIDAD 9	111
TABLA 21: MATRIZ DE CONFUSIÓN MODALIDAD 10 – LOCUTORES FEMENINOS.....	113
TABLA 22: MATRIZ DE CONFUSIÓN MODALIDAD 10 – LOCUTORES MASCULINOS	114
TABLA 23: MATRIZ DE CONFUSIÓN MODALIDAD 11 – INTERVALOS VOICED	115
TABLA 24: MATRIZ DE CONFUSIÓN MODALIDAD 11 – INTERVALOS UNVOICED.....	116
TABLA 25: MATRIZ DE CONFUSIÓN MODALIDAD 12 – LOCUTORES FEMENINOS E INTERVALOS VOICED	117
TABLA 26: MATRIZ DE CONFUSIÓN MODALIDAD 12 – LOCUTORES FEMENINOS E INTERVALOS UNVOICED	118
TABLA 27: MATRIZ DE CONFUSIÓN MODALIDAD 12 – LOCUTORES MASCULINOS E INTERVALOS VOICED	120
TABLA 28: MATRIZ DE CONFUSIÓN MODALIDAD 12 – LOCUTORES MASCULINOS E INTERVALOS UNVOICED	121
TABLA 29: RESULTADOS DE LAS PRUEBAS DEL SISTEMA: PORCENTAJES DE ACIERTO EN LA CLASIFICACIÓN DE LAS LOCUCIONES COMO NATIVO O EXTRANJERO	126
TABLA 30: ACTORES QUE INTERVIENEN EN LOS CASOS DE USO	131
TABLA 31: CASO DE USO - HACER EJERCICIO (ANÁLISIS)	133

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

TABLA 32: DESCRIPCIÓN DE LA CLASE USUARIO	135
TABLA 33: DESCRIPCIÓN DE LA CLASE FRASE	135
TABLA 34: DESCRIPCIÓN DE LA CLASE ENSAYO.....	135
TABLA 35: DESCRIPCIÓN DE LA CLASE PUNTUACIÓN.....	135
TABLA 36: CASO DE USO - HACER EJERCICIO (DISEÑO).....	142
TABLA 37: REQUISITOS PARA LAS PRUEBAS - FUNCIONALIDAD 1	153
TABLA 38: REQUISITOS PARA LAS PRUEBAS - FUNCIONALIDAD 2	154
TABLA 39: REQUISITOS PARA LAS PRUEBAS - FUNCIONALIDAD 3	155
TABLA 40: REQUISITOS PARA LAS PRUEBAS - FUNCIONALIDAD 4	156
TABLA 41: REQUISITOS PARA LAS PRUEBAS - FUNCIONALIDAD 5	157
TABLA 42: REQUISITOS PARA LAS PRUEBAS - FUNCIONALIDAD 6	157
TABLA 43: REQUISITOS PARA LAS PRUEBAS - FUNCIONALIDAD 7	158
TABLA 44: REQUISITOS PARA LAS PRUEBAS - FUNCIONALIDAD 8	158
TABLA 45: REQUISITOS PARA LAS PRUEBAS - FUNCIONALIDAD 9	159
TABLA 46: ESTRATEGIA PARA LAS PRUEBAS DE FUNCIONALIDAD	159
TABLA 47: ESTRATEGIA PARA LAS PRUEBAS DE INTERFAZ DE USUARIO	160
TABLA 48: ESTRATEGIA PARA LAS PRUEBAS DE INTERFAZ DE INTEGRIDAD DE LA BASE DE DATOS .	160
TABLA 49: HERRAMIENTAS PARA LAS PRUEBAS	160
TABLA 50: PLANTILLA CASOS DE PRUEBA.....	161
TABLA 51: IDENTIFICACIÓN DEL USUARIO EN EL CLIENTE MÓVIL CP-001A.....	161
TABLA 52: IDENTIFICACIÓN DEL USUARIO EN EL CLIENTE MÓVIL CP-001B.....	162
TABLA 53: IDENTIFICACIÓN DEL USUARIO EN EL CLIENTE MÓVIL CP-001C	162
TABLA 54: REGISTRO DEL USUARIO EN EL CLIENTE MÓVIL CP-002A	162
TABLA 55: REGISTRO DEL USUARIO EN EL CLIENTE MÓVIL CP-002B	163
TABLA 56: REGISTRO DEL USUARIO EN EL CLIENTE MÓVIL CP-002C	163
TABLA 57: DESCARGA DE LA GRABACIÓN MODELO CP-003A	163
TABLA 58: DESCARGA DE LA GRABACIÓN MODELO CP-003B	164
TABLA 59: REPRODUCCIÓN DE LA GRABACIÓN MODELO CP-004A	164
TABLA 60: GRABACIÓN DEL ENSAYO CP-005A	165
TABLA 61: SUBIDA DE LA GRABACIÓN DEL ENSAYO AL SERVIDOR CP-006A	165
TABLA 62: SUBIDA DE LA GRABACIÓN DEL ENSAYO AL SERVIDOR CP-006B	165
TABLA 63: RECONOCIMIENTO DE LAS PALABRAS PRODUCIDAS EN EL ENSAYO CP-007A	166
TABLA 64: ANÁLISIS DE LA ENTONACIÓN DEL ENSAYO CP-008A	166
TABLA 65: PUNTUACIÓN PRONUNCIACIÓN CP-009A	166
TABLA 66: PUNTUACIÓN ENTONACIÓN CP-010A.....	167
TABLA 67: RECEPCIÓN DE RESULTADOS EN CLIENTE MÓVIL CP-011A	167
TABLA 68: CIERRE DE SESIÓN DEL USUARIO CP-012.....	167
TABLA 69: IDENTIFICACIÓN DEL USUARIO EN EL CLIENTE MÓVIL CP-001D	168
TABLA 70: IDENTIFICACIÓN DEL USUARIO EN EL CLIENTE MÓVIL CP-001E.....	168
TABLA 71: IDENTIFICACIÓN DEL USUARIO EN EL CLIENTE MÓVIL CP-001F.....	169
TABLA 72: REGISTRO DEL USUARIO EN EL CLIENTE MÓVIL CP-002D	169
TABLA 73: REGISTRO DEL USUARIO EN EL CLIENTE MÓVIL CP-002E	169
TABLA 74: REGISTRO DEL USUARIO EN EL CLIENTE MÓVIL CP-002F	170
TABLA 75: DESCARGA DE LA GRABACIÓN MODELO CP-003C	170
TABLA 76: DESCARGA DE LA GRABACIÓN MODELO CP-003D	170
TABLA 77: REPRODUCCIÓN DE LA GRABACIÓN MODELO CP-004B	171
TABLA 78: GRABACIÓN DEL ENSAYO CP-005B	171
TABLA 79: SUBIDA DE LA GRABACIÓN DEL ENSAYO AL SERVIDOR CP-006C	172
TABLA 80: SUBIDA DE LA GRABACIÓN DEL ENSAYO AL SERVIDOR CP-006D	172

TABLA 81: RECEPCIÓN DE RESULTADOS EN CLIENTE MÓVIL CP-011B	172
TABLA 82: CIERRE DE SESIÓN DEL USUARIO CP-012B	173
TABLA 83: IDENTIFICACIÓN DEL USUARIO EN EL CLIENTE MÓVIL CP-001G.....	173
TABLA 84: IDENTIFICACIÓN DEL USUARIO EN EL CLIENTE MÓVIL CP-001H	173
TABLA 85: REGISTRO DEL USUARIO EN EL CLIENTE MÓVIL CP-002G.....	174
TABLA 86: REGISTRO DEL USUARIO EN EL CLIENTE MÓVIL CP-002H.....	174
TABLA 87: SUBIDA DE LA GRABACIÓN DEL ENSAYO AL SERVIDOR CP-006E	175
TABLA 88: RESUMEN DE LOS RESULTADOS DE LAS PRUEBAS	175
TABLA 89: COBERTURA DE LAS PRUEBAS BASADA EN REQUISITOS	176
TABLA 90: REQUISITOS MÍNIMOS PARA LA INSTALACIÓN DEL SERVIDOR.....	177
TABLA 91: CARACTERÍSTICAS EQUIPO 1.....	215
TABLA 92: CARACTERÍSTICAS EQUIPO 2.....	215
TABLA 93: CARACTERÍSTICAS EQUIPO 3.....	215
TABLA 94: CARACTERÍSTICAS EQUIPO 4.....	216
TABLA 95: CARACTERÍSTICAS EQUIPO 5.....	216

F - Índice de Ilustraciones

ILUSTRACIÓN 1: PROTOTIPO DE PARTIDA.....	15
ILUSTRACIÓN 2: FORMAS ENUNCIATIVAS (FUENTE: LLISTERRI, JOAQUIM. "LA DESCRIPCIÓN FONÉTICA Y FONOLÓGICA DEL ESPAÑOL: LOS ELEMENTOS SUPRASEGMENTALES")	30
ILUSTRACIÓN 3: ILUSTRACIÓN 2: FORMAS INTERROGATIVAS (FUENTE: LLISTERRI, JOAQUIM. "LA DESCRIPCIÓN FONÉTICA Y FONOLÓGICA DEL ESPAÑOL: LOS ELEMENTOS SUPRASEGMENTALES")	31
ILUSTRACIÓN 4: ILUSTRACIÓN 3: FORMAS VOLITIVAS (FUENTE: LLISTERRI, JOAQUIM. "LA DESCRIPCIÓN FONÉTICA Y FONOLÓGICA DEL ESPAÑOL: LOS ELEMENTOS SUPRASEGMENTALES")	32
ILUSTRACIÓN 5: ILUSTRACIÓN 4: FORMAS AFECTIVAS (FUENTE: LLISTERRI, JOAQUIM. "LA DESCRIPCIÓN FONÉTICA Y FONOLÓGICA DEL ESPAÑOL: LOS ELEMENTOS SUPRASEGMENTALES")	33
ILUSTRACIÓN 6: DIAGRAMA DE ESTADOS FINITO DE POSIBLES SECUENCIAS DE TONOS H/L (PIERREHUMBER, 1980)	39
ILUSTRACIÓN 7: INTERIOR DE LA CAVIDAD LARÍNGEA (FUENTE: LA FONACIÓN - PÁGINA WEB PERSONAL DE JOAQUIM LLISTERRI)	40
ILUSTRACIÓN 8: SEÑAL DE LA ONDA SONORA (PARTE SUPERIOR) Y FRECUENCIA FUNDAMENTAL (PARTE CENTRAL).....	41
ILUSTRACIÓN 9: CARGAR UN ARCHIVO EN PRAAT.....	50
ILUSTRACIÓN 10: EXTRAER UN CANAL DE UN OBJETO DE SONIDO (SOUND).....	51
ILUSTRACIÓN 11: EXTRAER EL CANAL IZQUIERDO DEL OBJETO DE SONIDO	51
ILUSTRACIÓN 12: ETIQUETAR UN ARCHIVO DE SONIDO.....	52
ILUSTRACIÓN 13: NOMBRE DE LAS CAPAS QUE SE UTILIZARÁN EN EL ETIQUETADO DEL OBJETO DE SONIDO	52
ILUSTRACIÓN 14: OBJETO DE SONIDO ETIQUETADO	53
ILUSTRACIÓN 15: GUARDAR EL OBJETO TEXTGRID.....	54
ILUSTRACIÓN 16: NUEVO SCRIPT	55
ILUSTRACIÓN 17: PEGAR HISTORIAL DE ACCIONES REALIZADAS	55
ILUSTRACIÓN 18: FUNCIONES RELACIONADAS CON OBJETOS .TEXTGRID	56
ILUSTRACIÓN 19: EXTRAER UNA PARTE DE UN OBJETO DE SONIDO	57
ILUSTRACIÓN 20: GUARDAR OBJETO DE SONIDO CON FORMATO WAV	57
ILUSTRACIÓN 21: ESTABLECER LA ETIQUETA DE UN INTERVALO	58
ILUSTRACIÓN 22: OBTENCIÓN DE UN OBJETO PITCH A PARTIR DE UN OBJETO SOUND	59
ILUSTRACIÓN 23: PARÁMETROS PARA LA EXTRACCIÓN DEL PITCH	60
ILUSTRACIÓN 24: CREACIÓN DEL OBJETO PITCHTIER A PARTIR DE UN OBJETO PITCH	60
ILUSTRACIÓN 25: CREACIÓN DE UN OBJETO INTENSITY A PARTIR DE UN OBJETO SOUND.....	61
ILUSTRACIÓN 26: PARÁMETROS DEL OBJETO INTENSITY	61
ILUSTRACIÓN 27: XLAUNCH EN MENÚ DE INICIO DE MICROSOFT WINDOWS.....	65
ILUSTRACIÓN 28: XLAUNCH - CONFIGURACIÓN DEL TERMINAL.....	66
ILUSTRACIÓN 29: XLAUNCH - CONFIGURACIÓN DE SESIÓN	66
ILUSTRACIÓN 30: XLAUNCH - ESTABLECER LOGIN DEL CLIENTE Y SERVIDOR.....	67
ILUSTRACIÓN 31: XLAUNCH - PARÁMETROS ADICIONALES.....	67
ILUSTRACIÓN 32: XLAUNCH – FINALIZAR	68

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

ILUSTRACIÓN 33: XLAUNCH - TERMINAL.....	68
ILUSTRACIÓN 34: DISTANCIA EUCLÍDEA PUNTO A PUNTO VS. DTW (FUENTE: "BASES DE DATOS DE SERIES TEMPORALES: REPRESENTACIÓN Y CONSULTAS", ANTONIO MORENO GARCÍA).....	73
ILUSTRACIÓN 35: ALINEAMIENTO TEMPORAL DE DOS SEÑALES CORRESPONDIENTES A DOS LOCUCIONES (FUENTE: UNIVERSIDAD AUTÓNOMA DE MADRID).....	74
ILUSTRACIÓN 36: FUNCIÓN DE ALINEAMIENTO (FUENTE: UNIVERSIDAD AUTÓNOMA DE MADRID).....	74
ILUSTRACIÓN 37: EFECTO DEL ALINEAMIENTO SOBRE DOS SEÑALES SIN ALINEAR (FUENTE: UNIVERSIDAD AUTÓNOMA DE MADRID).....	75
ILUSTRACIÓN 38: COMPARACIÓN DE VECTORES EN LOS PUNTOS DE LA TRAYECTORIA DE LA FUNCIÓN DE ALINEAMIENTO (FUENTE: UNIVERSIDAD AUTÓNOMA DE MADRID).....	75
ILUSTRACIÓN 39: ALGORITMO DTW - RESTRICCIÓN DE VENTANA (FUENTE: UNIVERSIDAD AUTÓNOMA DE MADRID).....	77
ILUSTRACIÓN 40: ALGORITMO DTW - RESTRICCIÓN DE PENDIENTE (FUENTE: UNIVERSIDAD AUTÓNOMA DE MADRID).....	77
ILUSTRACIÓN 41: ALGORITMO DTW - TIPOS DE RESTRICCIÓN DE PENDIENTE (FUENTE: UNIVERSIDAD AUTÓNOMA DE MADRID).....	78
ILUSTRACIÓN 42: PANTALLA DE INICIO DE WEKA.....	82
ILUSTRACIÓN 43: ABRIR FICHERO EN WEKA.....	82
ILUSTRACIÓN 44: PREPROCESS CON FICHERO CARGADO EN WEKA.....	83
ILUSTRACIÓN 45: ESCOGER FILTRO EN PREPROCESS.....	84
ILUSTRACIÓN 46: FILTRO DE ATRIBUTOS NO SUPERVISADO "REMOVE".....	85
ILUSTRACIÓN 47: PARÁMETROS DEL FILTRO DE ATRIBUTOS NO SUPERVISADO "REMOVE".....	86
ILUSTRACIÓN 48: FILTRO NO SUPERVISADO DE ATRIBUTOS "NOMINALTOBINARY".....	86
ILUSTRACIÓN 49: PARÁMETROS DEL FILTRO DE ATRIBUTOS NO SUPERVISADO "NOMINALTOBINARY".....	87
ILUSTRACIÓN 50:EFECTO DE APLICAR EL FILTRO NO SUPERVISADO DE ATRIBUTOS "NOMINALTOBINARY" AL ATRIBUTO NOMINAL "SENT", CON 15 VALORES POSIBLES.....	87
ILUSTRACIÓN 51: FILTRO NO SUPERVISADO DE INSTANCIAS "REMOVewithVALUES".....	88
ILUSTRACIÓN 52: PARÁMETROS DEL FILTRO NO SUPERVISADO DE INSTANCIAS "REMOVewithVALUES".....	89
ILUSTRACIÓN 53: SELECT ATTRIBUTES EN WEKA.....	90
ILUSTRACIÓN 54: RESULTADOS DE LA SELECCIÓN DE ATRIBUTOS CON WEKA.....	91
ILUSTRACIÓN 55: CLASSIFY EN WEKA.....	92
ILUSTRACIÓN 56: RESULTADOS DE CLASIFICACIÓN EN WEKA.....	93
ILUSTRACIÓN 57: SITUACIÓN IDEAL - DOS CONJUNTOS BIEN DIFERENCIADOS.....	122
ILUSTRACIÓN 58: LOCUCIÓN PERTENECE AL CONJUNTO DE LOS NATIVOS.....	123
ILUSTRACIÓN 59: LOCUCIÓN PERTENECE AL CONJUNTO DE LOS EXTRANJEROS.....	124
ILUSTRACIÓN 60: LOCUCIÓN PERTENECIENTE A AMBOS CONJUNTOS.....	124
ILUSTRACIÓN 61: LOCUCIÓN NO PERTENECE A NINGUNO DE LOS DOS CONJUNTOS.....	125
ILUSTRACIÓN 62: MODELO DE CASOS DE USO.....	132
ILUSTRACIÓN 63: DIAGRAMA DE CLASES DE ANÁLISIS.....	134
ILUSTRACIÓN 64: DIAGRAMA DE SECUENCIA HACER EJERCICIO - ANÁLISIS.....	136
ILUSTRACIÓN 65: DIAGRAMA DE DESPLIEGUE DEL SISTEMA.....	137
ILUSTRACIÓN 66: ARQUITECTURA DEL SISTEMA (3 CAPAS + BIBLIOTECAS).....	137
ILUSTRACIÓN 67: ARQUITECTURA DE CAPAS DETALLADO (3 CAPAS + BIBLIOTECAS).....	140
ILUSTRACIÓN 68: DIAGRAMA DE CLASES DE DISEÑO.....	143
ILUSTRACIÓN 69: DIAGRAMA DE SECUENCIA CARGAR FRASE - DISEÑO.....	145
ILUSTRACIÓN 70: DIAGRAMA DE SECUENCIA ESCUCHAR FRASE - DISEÑO.....	146
ILUSTRACIÓN 71: DIAGRAMA DE SECUENCIA GRABAR ENSAYO - DISEÑO.....	147
ILUSTRACIÓN 72: DIAGRAMA DE SECUENCIA EVALUAR ENSAYO - DISEÑO.....	149
ILUSTRACIÓN 73: EJECUTABLE ".APK" DE UVAPRONUNCIACIÓN.....	181

ILUSTRACIÓN 74: CONFIRMACIÓN PARA LA INSTALACIÓN DE UVAPRONUNCIACIÓN.....	181
ILUSTRACIÓN 75: UVAPRONUNCIACIÓN INSTALADA	182
ILUSTRACIÓN 76: MENÚ DE APLICACIONES CON ICONO DE UVAPRONUNCIACIÓN.....	183
ILUSTRACIÓN 77: SPLASH ACTIVITY DE UVAPRONUNCIACIÓN.....	183
ILUSTRACIÓN 78: CAMBIAR DIRECCIÓN IP DEL SERVIDOR.....	184
ILUSTRACIÓN 79: ACTIVITY DE REGISTRO (DESPLEGADA) DE UVAPRONUNCIACIÓN	185
ILUSTRACIÓN 80: ACTIVITY DE INICIO DE SESIÓN DE UVAPRONUNCIACIÓN.....	186
ILUSTRACIÓN 81: ACTIVITY MENÚ PRINCIPAL DE UVAPRONUNCIACIÓN.....	187
ILUSTRACIÓN 82: ACTIVITY EJERCICIO DE UVAPRONUNCIACIÓN	188
ILUSTRACIÓN 83: ACTIVITY EJERCICIO DE UVAPRONUNCIACIÓN - ESCUCHAR LA GRABACIÓN DE MODELO.....	189
ILUSTRACIÓN 84: CUADRO DE DIÁLOGO DE GRABACIÓN	190
ILUSTRACIÓN 85: CUADRO DE DIÁLOGO DE GRABACIÓN - GRABANDO	190
ILUSTRACIÓN 86: CONFIRMACIÓN DE ENVÍO DE LA GRABACIÓN DEL ENSAYO AL SERVIDOR.....	190
ILUSTRACIÓN 87:ACTIVITY DE RESULTADOS EN UVAPRONUNCIACIÓN	191
ILUSTRACIÓN 88: CERRAR SESIÓN DE USUARIO	192

G - Contenido del CD

El CD contiene el código de la aplicación desarrolladas, así como toda la documentación no añadida en esta memoria impresa. Está estructurado de la siguiente forma:

- **Carpeta Código:** Contiene el código fuente del cliente móvil android de la aplicación UVaPronunciación y el del servidor, ubicados en sus carpetas correspondientes:
 - **Carpeta Cliente:** Contiene el proyecto UVaPronunciación para la plataforma Android desarrollado en Eclipse. El código de la aplicación se encuentran en la subcarpeta /src del proyecto. El archivo para instalar la aplicación en un dispositivo Android es el archivo con extensión ".apk" de la subcarpeta /bin del proyecto.
 - **Carpeta Servidor:** Contiene el código necesario para el correcto funcionamiento del componente servidor de la aplicación UVaPronunciación, el archivo ".sql" que se debe importar para la creación de la base de datos, y los directorios en los que deben almacenarse las grabaciones de los ensayos y los recursos a descargar por los usuarios.
- **Carpeta Documentación:** Contiene el documento en formato ".pdf" del Proyecto de Fin de Carrera "Aplicación Móvil de Ayuda a la Mejora de la Pronunciación del Español" y un archivo html con links a documentación consultada, disponible en web, y que también pueden ser de interes.
- **TFG-Memoria.pdf:** memoria del proyecto.

H - Bibliografía

- [1] ARLOW, Jim. "UML 2.0 (Programación)". Madrid. Anaya Multimedia, 2006 ISBN: 84-415-2033-X.
- [2] BAGSHAW, Paul Cristopher. "Automatic Prosodic Analysis For Computer Aided Pronunciation Teaching" [en línea]. Fecha: 1994. Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.55.3401>
- [3] BELLARD, Fabrice. "FFmpeg" [Computer program]. Fecha: 2013. Disponible en: <http://www.ffmpeg.org/legal.html>
- [4] BOERSMA, Paul. "Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-to-Noise Ratio of a Sampled Sound"[en línea]. Fecha: 1993. Disponible en: http://xeds.eu/other/P_Boersma_Accurate_short-term_analysis_of_the_fundamental_freq.pdf
- [5] BOERSMA, Paul & WEENINK, David. "Praat: doing phonetics by computer" [Computer program]. Version 5.3.55. Fecha: Septiembre 2013. Disponible en: <http://www.praat.org/>
- [6] CABRÉ LUNAS, Laura – Universitat de Barcelona. "Prosodia" [en línea]. Disponible en: <http://www.ub.edu/diccionarilinguistica/content/prosodia-0>
- [7] CAMPILLOS LLANOS, Leonardo. "Tecnologías del habla y análisis de la voz. Aplicaciones en la enseñanza de la lengua", en *Diálogo de la Lengua* [en línea]. Fecha: 2010. Disponible en: http://www.dialogodelalengua.com/articulo/pdf/2/1_campillos_DL_2010.pdf
- [8] GAMMA, E. et. al. "Patrones de diseño: Elementos de software orientado a objeto reutilizable." Madrid: Addison Wesley, 2003. ISBN: 84-7829-059-1
- [9] GARCÍA MORATE, Diego. "Manual de WEKA" [en línea]. Disponible en: <http://www.metaemotion.com/diego.garcia.morate/download/weka.pdf>
- [10] KTH ROYAL INSTITUTE OF TECHNOLOGY. "The Snack Sound Toolkit" [Computer program]. Fecha: 2005. Disponible en: <http://www.speech.kth.se/snack/>
- [11] INTERVAL RESEARCH CORPORATION. "dtw.c" [Computer program]. Fecha: 1995-1997. Disponible en: http://read.pudn.com/downloads16/sourcecode/multimedia/audio/60305/dtw.c_.htm
- [12] LI, Kun et al. "Detention of Intonation in L2 English Speech of Native Mandarin Learners" [en línea]. Fecha: 2010
- [13] LLISTERRI, Joaquim. "La descripción fonética y fonológica del español: los elementos suprasegmentales" [en línea]. Disponible en: http://liceu.uab.es/~joaquim/phonetics/fon_esp/fonetica_espanol_suprasegmental.html
- [14] LLISTERRI, Joaquim. "La fonación" [en línea]. Disponible en: http://liceu.uab.es/~joaquim/phonetics/fon_produccio/fonacion.html

Evaluación de la Entonación en la Enseñanza del Español como Segunda Lengua

- [15] LLISTERRI, Joaquim. “*Las características acústicas de los sonidos del habla*” [en línea]. Disponible en: http://liceu.uab.es/~joaquim/phonetics/fon_anal_acus/fon_acust.html#amplitud
- [16] LLISTERRI, J. (2003) “*La evaluación de la pronunciación en la enseñanza del español como segunda lengua*”, en REYZÁBAL, M.V. (Dir.) *Perspectivas teóricas y metodológicas: Lengua de acogida, educación intercultural y contextos inclusivos* [en línea]. Disponible en: http://liceu.uab.es/~joaquim/publicacions/Eval_Pron_EL2.pdf
- [17] LLISTERRI, Joaquim. “*Lingüística y tecnologías del lenguaje*” [en línea]. Fecha: 2003. Disponible en: http://liceu.uab.es/~joaquim/publicacions/Llisterri_03_Linguistica_Tecnologias_Lenguaje.pdf
- [18] LORD, Gillian – College of Liberal Arts & Science, University of Florida. “*Fonética Española, capítulo 19*” [en línea]. Disponible en: www.clas.ufl.edu/users/glord/fonetica/pres/cap19.ppt
- [19] MARQUÉS CORRAL, José Manuel. “*Ingeniera del software I*”. 2º curso de Ingeniería Técnica en Informática de Sistemas. Universidad de Valladolid, 2010.
- [20] MARQUÉS CORRAL, José Manuel. “*Ingeniera del software II*”. 3º curso de Ingeniería Técnica en Informática de Sistemas. Universidad de Valladolid, 2011.
- [21] MORENO ALONSO, Antonio. “*Bases de Datos de Series Temporales: Representación y Consultas*” [en línea]. Fecha: 2012. Disponible en: <https://ruidera.uclm.es/xmlui/bitstream/handle/10578/2440/TFM-AntonioMorenoGarcia.pdf?sequence=1>
- [22] MOUSTROFAS, N., DIGALAKIS, V. “*Automatic pronunciation evaluation of foreign speakers using unknown text*” [en línea]. Fecha: 2006.
- [23] ORTEGA, J. *DTW_AlinTempDinam_ASAL* [en línea]. Disponible en Web: http://arantxa.ii.uam.es/~jortega/DTW_AlinTempDinam_ASAL.pdf
- [24] SCHALKWYK, Johan et al. “*Google Search by Voice: A case study*” [en línea]. Fecha: 2010. Disponible en: http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/es//pubs/archive/36340.pdf
- [25] UNIVERSITY OF WAIKATO. “*Weka*” [Computer program] Fecha: 2006. Version 3.6.9. Disponible en: <http://www.cs.waikato.ac.nz/ml/weka/downloading.html>
- [26] XU, Y. “*ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. In Proceedings of Tools and Resources for the Analysis of Speech Prosody*” [en línea]. Fecha: 2013. Disponible en: <http://www.phon.ucl.ac.uk/home/yi/ProsodyPro/>
- [27] ZECHNER, Klaus, XI, Xiaoming, CHEN, Lei. “*Evaluating Prosodic Features for Automated Scoring of Non-Native Read Speech*” [En línea]. Fecha: 2011
- [28] Android Developers. *Dev Guide*. [en línea]. Disponible en: <http://developer.android.com/index.html>
-

- [29] Central Intelligence Agency. “*The world fact book*” [en línea]. Disponible en: <https://www.cia.gov/library/publications/the-world-factbook/fields/2098.html>
- [30] Centro Virtual Cervantes. “*Diccionario de términos clave de ELE*” [en línea]. Disponible en: http://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/indice.htm
- [31] Centro Virtual Cervantes. “*El español: una lengua viva. Informe 2012*” [en línea]. Disponible en: http://cvc.cervantes.es/lengua/anuario/anuario_12/default.htm
- [32] Centro Virtual Cervantes. “*Plan Curricular del Instituto Cervantes*” [en línea]. Disponible en: http://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/default.htm
- [33] Diccionario de la Real Academia Española de la Lengua. “*Prosodia*” [en línea]. Disponible en: <http://lema.rae.es/drae/>
- [34] Fono.Ele. “*La competencia fónica*” [en línea]. Disponible en: <http://www2.uah.es/fonoele/proyecto-competencia-fonica.php>
- [35] Instituto Cervantes – Diplomas de Español como Lengua Extranjera. “*Descripción DELE*” [en línea]. Disponible en: http://diplomas.cervantes.es/informacion/descripcion_dele.html
- [36] Instituto Cervantes – Diplomas de Español como Lengua Extranjera. “*Niveles DELE*” [en línea]. Disponible en: <http://diplomas.cervantes.es/informacion/niveles/tipos.html>
- [37] Instituto Cervantes – Diplomas de Español como Lengua Extranjera. “*Niveles y certificados*” [en línea]. Disponible en: http://www.cervantes.es/lengua_y_ensenanza/portales_espanol/diplomas_espanol.htm
- [38] Instituto Cervantes – Diplomas de Español como Lengua Extranjera. “*Especificaciones generales de las pruebas DELE B1 y DELE B2*” [en línea]. Disponible en: http://difusionelectronica.institutocervantes.es/archivos/1/24134_dele-nuevas-especificaciones_2012.pdf
- [39] Instituto Cervantes – Diplomas de Español como Lengua Extranjera. “*Especificaciones generales DELE B1*” http://emagiustelli.wikispaces.com/file/view/presentazione_DELE_B1_.pdf
- [40] OpenmokoWiki. “*Google Voice Recognition*” [en línea]. Fecha:2012. Disponible en: http://wiki.openmoko.org/wiki/Google_Voice_Recognition
- [41] sgoliver.net blog. “*Pensamientos varios sobre programación, Android, .NET y Java*” [en línea]. Disponible en: <http://www.sgoliver.net/blog/?p=2035>
- [42] Wikipedia. [en línea]. Disponible en: <http://es.wikipedia.org/wiki/Wikipedia:Portada>