



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

MONITOREO DEL ESTADO ANÍMICO DE LOS ADULTOS MAYORES, MEDIANTE  
EL PROCESAMIENTO DE LA VOZ

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL EN COMPUTACIÓN

VÍCTOR ANDRÉS ROJAS HERNÁNDEZ

PROFESOR GUÍA:  
SERGIO OCHOA DELORENZI

MIEMBROS DE LA COMISIÓN:  
SERGIO OCHOA DELORENZI  
NANCY HITSCHFELD KAHLER  
ALEX BÓRQUEZ GRIMALDI

Este proyecto ha sido parcialmente financiado por el Proyecto Fondecyt Nro.  
1120207.

SANTIAGO DE CHILE  
2014

## Resumen

La sociedad ha ido cambiando, y envejeciendo también. A la par, los adultos mayores cada vez se encuentran más solos, ya sea por abandono o por un alejamiento progresivo de la familia, generado por el ritmo de vida actual. Este aislamiento puede llevar a una persona a sufrir algún trastorno emocional, como lo es la depresión.

La depresión no es fácil de detectar, es por ello, que en el trabajo de Memoria se desarrolló un sistema capaz de reconocer los estados anímicos de una persona mediante la voz, para así, poder detectar cuando la persona se encuentra en un estado anímico negativo.

Para poder decir que una persona se encuentra en un estado anímico negativo, no sólo basta con reconocer la emoción involucrada, también es necesario reconocer quién es esa persona. Por ello, además de detectar la emoción expresada, se agregó la detección del género y luego identidad.

La búsqueda para encontrar la mejor clasificación según género pasó por numerosas pruebas, incluyendo distintos tipos de muestras (grabaciones de voces leyendo y muestras de emociones). Después de varios intentos, se llegó a que la mejor forma de clasificar según género cuando hay emociones involucradas es clasificando con respecto a la emoción identificada.

También se hicieron pruebas para clasificar las emociones. Éstas se basaron en emociones y muestras de tres bases de datos emocionales, SAVEE, Emo-DB, y RekEmozio. Principalmente se buscó obtener una clasificación que permitiera reconocer las emociones negativas con mayor certeza. Las pruebas se vieron bastante enfocadas en lograr aislar la emoción Tristeza, emoción negativa de importancia para este trabajo, logrando reconocer si una muestra expresa una emoción de Tristeza o no.

Con respecto a la reconocimiento de la identidad, se hicieron pocas pruebas, las cuales tenían buenos resultados en un principio. Posteriormente a todas las pruebas hechas, se probó con muestras más reales o naturales, grabaciones de conversaciones, y se demostró que al menos la clasificación de emociones y género había dado buenos frutos.

Además de las pruebas de reconocimiento, se diseñó, implementó, y evaluó la aplicación móvil pensada para el monitoreo de estados anímicos. Esta aplicación fue diseñada para visualizar los datos que el monitoreo arroje sobre los estados del adulto mayor, y también para configurar de forma remota las alertas y el sistema de grabación.

## Agradecimientos

Agradezco a mis padres, quienes me apoyaron en todo momento, siempre dándome fuerzas para no rendirme, para seguir adelante todos estos años, y más aún en este trabajo final. A mi abuela, que siempre ha velado por el bienestar de mis hermanos y yo, y siempre ha deseado lo mejor para nosotros. Y como no agradecer a mis hermanos, sin su apoyo esto habría sido más difícil.

Quiero agradecer también a mis amigos, compañeros de Universidad, compañeros de trabajo, familiares, y a todos aquellos que, ya sea en persona o a distancia, me brindaron siempre su apoyo y nunca dudaron en que lograría llevar a cabo este proyecto.

Finalmente, agradezco enormemente a mi profesor guía Sergio Ochoa, por su constante apoyo y por la confianza que depositó en mí, esa confianza que me hizo también confiar más en mi mismo, confiar en que estaba haciendo bien las cosas.

Este proyecto ha sido parcialmente financiado por el Proyecto Fondecyt Nro. 1120207.

## Tabla de Contenido

1. Introducción.....	1
1.1. Justificación de la Propuesta.....	2
1.2. Objetivos de la Memoria.....	3
2. Trabajos Relacionados.....	5
2.1. Conjunto de Características Relevantes de la Voz.....	5
2.2. Programas Revisados.....	6
3. Pruebas Realizadas.....	7
3.1. Reconocimiento de emociones.....	8
3.2. Reconocimiento de género.....	10
3.3. Reconocimiento de identidades.....	11
3.4. Extracción de Frases y Palabras.....	11
4. Concepción de la Solución.....	14
4.1. Requisitos.....	14
4.2. Tecnologías y Productos de Software a utilizar.....	15
4.3. Arquitectura de la solución.....	16
4.4. Diseño de la solución.....	18
5. Implementación del Sistema.....	21
5.1. Scripting.....	21
5.1.1. Creación de modelos SVM.....	21
5.1.2. Predicción de modelos SVM.....	22
5.1.3. Procesamiento de resultados.....	23
5.1.4. Extracción de palabras y frases.....	23
5.1.5. Procesamiento de audio.....	24
5.2. Aplicación Móvil.....	25
5.2.1. La aplicación.....	25
5.2.2. Propósito de la evaluación de Usabilidad.....	27
5.2.3. Metodología de la evaluación de Usabilidad.....	27
5.2.4. Detalles de la implementación de la aplicación móvil.....	36
5.3. Resumen.....	38
6. Evaluación del Sistema.....	39
6.1. Evaluación de reconocimiento de emociones, género e identidad en conversaciones.....	39
6.2. Evaluación de Usabilidad de la Aplicación Móvil.....	40
6.2.1. Paper Prototyping.....	40
6.2.2. Evaluación Heurística.....	45
6.2.3. Observación.....	47
6.2.4. Cuestionario de usuario final.....	53

7. Conclusiones y Trabajo a Futuro.....	57
Bibliografía y Referencias.....	60
Anexo A: Pruebas Realizadas para Reconocimiento de Emociones, Género e Identidad.....	63
A.1. Pruebas de Distancia.....	63
A.2. Pruebas de Emociones.....	63
A.3. Pruebas de Género e Identificación de una persona.....	64
A.3.1. Prueba 1: Primera prueba.....	65
A.3.2. Prueba 2: Exactitud de la clasificación.....	69
A.3.3. Prueba 3: Clasificación según género en audio con emociones.....	72
A.3.4. Prueba 4: Kernel lineal.....	74
A.3.5. Prueba 5: Vector de características IS10_paraling.....	79
A.3.6. Prueba 6: Muestras de audio con emociones en conjunto de entrenamiento para reconocimiento de género.....	83
A.3.7. Prueba 7: Normalización del volumen.....	86
A.3.8. Prueba 8-A: Otra forma para clasificar voz según género - Extracción de características más específicas.....	89
A.3.9. Prueba 8-B: Otra forma para clasificar voz según género - Clasificación de género según emociones.....	97
A.3.10. Prueba 9: Clasificación de género por idioma.....	101
A.3.11. Prueba 10: Clasificación de género por emoción.....	103
A.4. Pruebas con Bases de Datos de Emociones.....	105
A.4.1. Prueba 1: Clasificación de cada emoción.....	106
A.4.2. Prueba 2-A: Clasificación de emociones por grupo - Grupo1.....	110
A.4.3. Prueba 2-B: Clasificación de emociones por grupo - Grupo2.....	113
A.4.4. Prueba 2-C: Clasificación de emociones por grupo - Grupo3.....	115
A.4.5. Prueba 2-D: Clasificación de emociones por grupo - Grupo3 - RekEmozio modificada.....	117
A.4.6. Prueba 2-E: Clasificación de emociones por grupo - Grupo3 - Unificación de bases de datos.....	119
A.4.7. Prueba 3-A: Reconocimiento de tristeza - Unificación de Grupo2.....	121
A.4.8. Prueba 3-B: Reconocimiento de tristeza - Arousal y Valence - Arousal +/- y Valence +/-.....	122
A.4.9. Prueba 3-C: Reconocimiento de tristeza - Arousal y Valence - 4 grupos .....	130
A.4.10. Prueba 3-D: Reconocimiento de tristeza - Arousal y Valence - 5 grupos .....	132

A.4.11. Prueba 3-E: Reconocimiento de tristeza - Arousal y Valence - 4 grupos sin Neutral.....	133
A.5. Pruebas de Conversaciones.....	134
A.5.1. Prueba 1: Conversación entre 2 hombres.....	134
A.5.2. Prueba 2: Conversación entre una mujer y un hombre.....	138
A.5.3. Prueba 3: Conversación entre una mujer y un hombre, con solapamiento de voces.....	142
Anexo B: Scripts y Resultados.....	144
B.1. Código del script test_model_folders.sh.....	144
B.2. Ejemplo de resultados en archivo predict.....	145
B.3. Ejemplo de resultados en archivo predict-results.....	145
Anexo C: Aplicación Móvil y su Estudio de Usabilidad.....	146
C.1. Pauta de Usuario Final.....	146
C.2. Pauta de Evaluación Heurística respondida por Experto1.....	149
C.3. Pauta de Evaluación Heurística respondida por Experto2.....	154
C.4. Pauta de tareas de Paper Prototyping.....	159
C.5. Pauta de Observación.....	165
Anexo D: Mockups diseñados.....	168
Anexo E: Pantallas de Segundo Prototipo de la Aplicación Móvil.....	169
Anexo F: Diagramas de Clases de Aplicación Móvil.....	171

## Índice de Tablas

Tabla 1: Media de edad para cada rango de usuarios finales.....	29
Tabla 2: Medias de tiempo en cada tarea observada.....	51
Tabla 3: Medias de tiempo en cada tarea observada, para cada rango de edad.....	51
Tabla 4: Medias de tiempo en cada tarea observada, para cada nivel de experiencia. .....	53
Tabla 5: Media de valoración de los ítems en Cuestionario de Usuario Final.....	54
Tabla 6: Media de valoración de los ítem en Cuestionario de Usuario Final, con respecto a rango de edad.....	55
Tabla 7: Tono e intensidad para grabación 1, con distintas emociones.....	63
Tabla 8: Tono e intensidad para grabación 2, con distintas emociones.....	64
Tabla 9: Tono e intensidad para grabación 3, con distintas emociones.....	64
Tabla 10: Resultados de pruebas con modelo Mujer-Hombre.....	67
Tabla 11: Resultados de pruebas modelo Mujer-Hombre, contrastando voces masculinas extras.....	67
Tabla 12: Resultados de pruebas modelo Mujer.....	67
Tabla 13: Resultados de pruebas modelo Hombre.....	67
Tabla 14: Resultados de clasificación según género en conjunto Mujer.....	70
Tabla 15: Resultados de clasificación según género en conjunto Hombre.....	70
Tabla 16: Resultados de clasificación según género en conjunto Hombre-Otro.....	70
Tabla 17: Resultados de reconocimiento de identidad en conjunto Mujer.....	71
Tabla 18: Resultados de reconocimiento de identidad en conjunto Hombre.....	71
Tabla 19: Resultados de pruebas en conjunto Hombre-SAVEE.....	73
Tabla 20: Resultados de pruebas en conjunto Mujer-Emo-DB.....	73
Tabla 21: Resultados de pruebas en conjunto Hombre-Emo-DB.....	73
Tabla 22: Resultados de clasificación de género en conjunto Mujer, usando kernel lineal.....	75
Tabla 23: Resultados de clasificación de género en conjunto Hombre, usando kernel lineal.....	76

Tabla 24: Resultados de clasificación de género en conjunto Hombre-Otro, usando kernel lineal.....	76
Tabla 25: Resultados de clasificación de género en conjunto Hombre-SAVEE, usando kernel lineal.....	76
Tabla 26: Resultados de clasificación de género en conjunto Mujer-Emo-DB, usando kernel lineal.....	77
Tabla 27: Resultados de clasificación de género en conjunto Hombre-Emo-DB, usando kernel lineal.....	77
Tabla 28: Resultados de reconocimiento de identidad en conjunto Mujer, usando kernel lineal.....	78
Tabla 29: Tabla 22: Resultados de reconocimiento de identidad en conjunto Hombre, usando kernel lineal.....	78
Tabla 30: Resultados de clasificación de género en conjunto Mujer, en base a IS10_paraling.....	79
Tabla 31: Resultados de clasificación de género en conjunto Hombre, en base a IS10_paraling.....	80
Tabla 32: Resultados de clasificación de género en conjunto Hombre-Otro, en base a IS10_paraling.....	80
Tabla 33: Resultados de clasificación de género en conjunto Hombre-SAVEE, en base a IS10_paraling.....	80
Tabla 34: Resultados de clasificación de género en conjunto Mujer-Emo-DB, en base a IS10_paraling.....	81
Tabla 35: Resultados de clasificación de género en conjunto Hombre-Emo-DB, en base a IS10_paraling.....	81
Tabla 36: Resultados de reconocimiento de identidad en conjunto Mujer, en base a IS10_paraling.....	82
Tabla 37: Resultados de reconocimiento de identidad en conjunto Hombre,, en base a IS10_paraling.....	82
Tabla 38: Resultados de clasificación de género en conjunto Mujer.....	84
Tabla 39: Resultados de clasificación de género en conjunto Mujer-Emo-DB.....	84
Tabla 40: Resultados de clasificación de género en conjunto Hombre.....	84
Tabla 41: Resultados de clasificación de género en conjunto Hombre-Otro.....	85
Tabla 42: Resultados de clasificación de género en conjunto Hombre-SAVEE.....	85

Tabla 43: Resultados de clasificación de género en conjunto Hombre-Emo-DB.....	85
Tabla 44: Resultados de clasificación según género en conjunto Mujer.....	87
Tabla 45: Resultados de clasificación según género en conjunto Mujer-Emo-DB.....	87
Tabla 46: Resultados de clasificación según género en conjunto Hombre.....	87
Tabla 47: Resultados de clasificación según género en conjunto Hombre-Otro.....	88
Tabla 48: Resultados de clasificación según género en conjunto Hombre-SAVEE....	88
Tabla 49: Resultados de clasificación según género en conjunto Hombre-Emo-DB...	88
Tabla 50: Resultados de clasificación según género en conjunto Mujer.....	95
Tabla 51: Resultados de clasificación según género en conjunto Mujer-Emo-DB.....	95
Tabla 52: Resultados de clasificación según género en conjunto Hombre.....	96
Tabla 53: Resultados de clasificación según género en conjunto Hombre-Otro.....	96
Tabla 54: Resultados de clasificación según género en conjunto Hombre-SAVEE....	96
Tabla 55: Resultados de clasificación según género en conjunto Hombre-Emo-DB...	97
Tabla 56: Características seleccionadas.....	98
Tabla 57: Resultados de clasificación según género en conjunto Mujer.....	99
Tabla 58: Resultados de clasificación según género en conjunto Mujer-Emo-DB.....	99
Tabla 59: Resultados de clasificación según género en conjunto Hombre.....	99
Tabla 60: Resultados de clasificación según género en conjunto Hombre-Otro.....	100
Tabla 61: Resultados de clasificación según género en conjunto Hombre-SAVEE..	100
Tabla 62: Resultados de clasificación según género en conjunto Hombre-Emo-DB.	100
Tabla 63: Resultados de clasificación según género en conjunto Mujer-Emo-DB.....	102
Tabla 64: Resultados de clasificación según género en conjunto Hombre-Emo-DB.	102
Tabla 65: Exactitud de clasificación y Verdaderos Positivos en el conjunto Emo.....	104
Tabla 66: Exactitud de clasificación y Verdaderos Positivos en el conjunto EmoNorm. .....	104
Tabla 67: Clasificación de emociones en SAVEE, usando modelo lineal.....	107
Tabla 68: Clasificación de emociones en Emo-DB, usando modelo lineal.....	108
Tabla 69: Clasificación de emociones en RekEmozio, usando modelo lineal.....	108

Tabla 70: Clasificación de emociones en SAVEE, usando modelo polinomial grado 2. .....	109
Tabla 71: Clasificación de emociones en SAVEE, usando modelo polinomial grado 3. .....	110
Tabla 72: Clasificación de grupos emociones en SAVEE, contrastado con todas las bases de datos emocionales.....	112
Tabla 73: Clasificación de grupos emociones en Emo-DB, contrastado con todas las bases de datos emocionales.....	112
Tabla 74: Clasificación de grupos emociones en RekEmozio, contrastado con todas las bases de datos emocionales.....	112
Tabla 75: Clasificación de grupos emociones en SAVEE, contrastado con todas las bases de datos emocionales.....	114
Tabla 76: Clasificación de grupos emociones en Emo-DB, contrastado con todas las bases de datos emocionales.....	114
Tabla 77: Clasificación de grupos emociones en RekEmozio, contrastado con todas las bases de datos emocionales.....	115
Tabla 78: Clasificación en tres grupos de emociones para SAVEE, contrastado con todas las bases de datos emocionales.....	116
Tabla 79: Clasificación en tres grupos de emociones para Emo-DB, contrastado con todas las bases de datos emocionales.....	116
Tabla 80: Clasificación en tres grupos de emociones para RekEmozio, contrastado con todas las bases de datos emocionales.....	117
Tabla 81: Contraste de modelo RekEmozio en muestras de SAVEE.....	118
Tabla 82: Contraste de modelo RekEmozio en muestras de Emo-DB.....	118
Tabla 83: Clasificación de emociones en RekEmozio, contrastado con todas las bases de datos emocionales.....	119
Tabla 84: Clasificación de grupos emociones en RekEmozio, contrastado con Todo-BD.....	120
Tabla 85: Clasificación de grupos emociones en Emo-DB, contrastado con Todo-BD. .....	120
Tabla 86: Clasificación de grupos emociones en RekEmozio, contrastado con Todo-BD.....	120
Tabla 87: Reconocimiento de muestras de Tristeza en el Grupo Neutral-Tristeza....	122

Tabla 88: Arousal y Valence de cada emoción, según los valores definidos para la base de datos de prueba de InterSpeech 2013. ?: Clasificado como otro y excluido de la evaluación. *: no aparece en la clasificación.....	125
Tabla 89: Clasificación de emociones en Arousal negativo, con Miedo en Arousal negativo.....	128
Tabla 90: Clasificación de emociones en Valence negativo, con Miedo en Valence positivo.....	128
Tabla 91: Clasificación de emociones en Arousal negativo y Valence negativo, con Miedo en Arousal negativo y Valence postivo.....	128
Tabla 92: Clasificación de emociones en Arousal negativo, con Miedo en Arousal positivo.....	129
Tabla 93: Clasificación de emociones en Valence negativo, con Miedo en Valence negativo.....	129
Tabla 94: Clasificación de emociones en Arousal negativo y Valence negativo, con Miedo en Arousal positivo y Valence negativo.....	129
Tabla 95: Clasificación de emociones en cuadrante Arousal negativo y Valence negativo, con Miedo en Arousal negativo y Valence positivo.....	131
Tabla 96: Clasificación de emociones en cuadrante Arousal negativo y Valence negativo, con Miedo en Arousal positivo y Valence negativo.....	131
Tabla 97: Exactitud de clasificación de muestras de Tristeza en cada prueba y base de datos.....	132
Tabla 98: Clasificación de emociones en cuadrante Arousal negativo y Valence negativo, con Miedo en Arousal negativo y Valence positivo, y Neutral aparte.....	133
Tabla 99: Clasificación de emociones en cuadrante Arousal negativo y Valence negativo, con Miedo en Arousal negativo y Valence positivo, y sin Neutral.....	134
Tabla 100: Clasificación según género en muestras de primera conversación, primer intento.....	135
Tabla 101: Clasificación según género en muestras de primera conversación, segundo intento.....	135
Tabla 102: Clasificación según emoción en muestras de primera conversación, primer filtro (Todo-BD).....	136
Tabla 103: Clasificación según emoción en muestras de primera conversación, segundo filtro (Arousal negativo y Valence negativo, AV4 nn).....	136

Tabla 104: Clasificación según género en muestras de primera conversación, que pasaron el segundo filtro de emoción. Se clasifica de acuerdo a emoción detectada. .....	136
Tabla 105: Reconocimiento de identidad en muestras de primera conversación, usando modelo basado en emobase2010.....	137
Tabla 106: Reconocimiento de identidad en muestras de primera conversación, usando modelo basado en IS10_paraling.....	137
Tabla 107: Clasificación según género en muestras de segunda conversación, primer intento.....	138
Tabla 108: Clasificación según género en muestras de segunda conversación, segundo intento.....	138
Tabla 109: Clasificación según emoción en muestras de segunda conversación, primer filtro (Todo-BD).....	139
Tabla 110: Clasificación según emoción en muestras de segunda conversación, segundo filtro (Arousal negativo y Valence negativo, AV4 nn).....	139
Tabla 111: Clasificación según género en muestras de segunda conversación, que pasaron el segundo filtro de emoción. Se clasifica de acuerdo a emoción detectada. .....	140
Tabla 112: Reconocimiento de identidad en muestras de segunda conversación, usando modelo basado en emobase2010.....	141
Tabla 113: Reconocimiento de identidad en muestras de segunda conversación, usando modelo basado en emobase2010.....	141
Tabla 114: Reconocimiento de identidad en muestras de segunda conversación, usando modelo basado en IS10_paraling.....	142
Tabla 115: Reconocimiento de identidad en muestras de segunda conversación, usando modelo basado en IS10_paraling.....	142
Tabla 116: Clasificación según emoción en muestras de tercera conversación, primer filtro (Todo-BD).....	143
Tabla 117: Clasificación según emoción en muestras de tercera conversación, segundo filtro (Arousal negativo y Valence negativo, AV4 nn).....	143

## Índice de Ilustraciones

Ilustración 1: Grabación, procesamiento del audio y proceso de reconocimiento.....	17
Ilustración 2: Almacenamiento de datos y alerta en caso de estado vulnerable.....	18
Ilustración 3: Diagramas de flujo para los procesos de entrenamiento y testeo de muestras.....	19
Ilustración 4: Diagrama de flujo de navegación de la aplicación móvil para Android. .....	20
Ilustración 5: Ejemplos de interfaces diseñadas. Login, Opciones e Historial de estados.....	32
Ilustración 6: Usuario final del rango 40-60 años, realizando una tarea de Paper Prototyping.....	33
Ilustración 7: Ejemplos de pantallas del primer prototipo. Lista de usuarios monitoreados y Configuración de Monitoreo.....	34
Ilustración 8: Usuarios expertos probando la aplicación y respondiendo la Pauta de Evaluación Heurística.....	34
Ilustración 9: Ejemplos de pantallas del segundo prototipo. Menú superior agregado e Icono de calendario en Historial de estados.....	35
Ilustración 10: Usuarios finales realizando las tareas asignadas en la Pauta de Observación. El primer Usuario final es del rango 20-35 años y el segundo es del rango 40-60 años.....	35
Ilustración 11: Diagrama de clases del paquete com.thesis.emostatus de la aplicación móvil.....	37
Ilustración 12: Flujo de navegación correcto para revisar el historial.....	41
Ilustración 13: Flujo de navegación incorrecto para revisar el historial.....	41
Ilustración 14: Mockup de pantalla de Configuración de Monitoreo.....	42
Ilustración 15: Flujo pensado para la modificación del correo electrónico.....	43
Ilustración 16: Comparación de interfaces para elegir opciones para un usuario monitoreado.....	44
Ilustración 17: Nueva funcionalidad de Calendario.....	47
Ilustración 18: Pantalla de login mostrando la opción Sig. en el teclado.....	48
Ilustración 19: Enfoque inicial del proceso de reconocimiento.....	65
Ilustración 20: MCR para muestras de los conjuntos MujerX y HombreX.....	92

Ilustración 21: MCR para muestras de los conjuntos MujerX-Low y HombreX-Low....	93
Ilustración 22: STE para muestras de los conjuntos MujerX y HombreX.....	94
Ilustración 23: STE para muestras de los conjuntos MujerX-Low y HombreX-Low.....	94
Ilustración 24: Nuevo enfoque del proceso de reconocimiento.....	105
Ilustración 25: Una de las versiones del Circumflex of Affect ideado por Russell [20]. .....	124
Ilustración 26: Arousal y Valence para emociones de bases de datos emocionales. .....	126

# 1. Introducción

Los problemas de salud en un adulto se acrecientan a medida que éste se acerca a la tercera edad, no sólo aquellos relacionados con la salud física que son parte de dicha etapa, sino también aquellos relacionados con su salud emocional [30]. Otro problema que afecta a muchos adultos mayores es el aislamiento social. Según un estudio hecho el año 2011, en base a una encuesta hecha a 3858 adultos mayores al azar (>75 años), se determinó que el 46,1% de ellos presentaba aislamiento social [29]. Este aislamiento puede deberse a muchos factores, como por ejemplo la falta de un trabajo al cual acudir diariamente, el deterioro físico, o la dinámica de la vida moderna, entre otros. Esta condición los hace mucho más propensos a sufrir trastornos emocionales o enfermedades psicológicas como por ejemplo la depresión [31]. En gran parte de los casos, la depresión en adultos mayores se da por el poco contacto que estas personas tienen con sus familiares. La vorágine de la vida moderna tampoco ayuda a crear estos espacios de interacción a través de los cuales los adultos mayores puedan mejorar su estado anímico.

Para los familiares o para aquellos que están a cargo de ancianos, a veces no es fácil reconocer si sus ancianos se encuentran en un estado emocional que necesita alguna atención. Por otro lado, tenemos la situación de que la mayoría de los familiares no tiene el tiempo suficiente para visitar a sus padres o abuelos, siendo el teléfono casi el único medio de comunicación con que pueden informarse sobre el estado emocional de sus adultos mayores. En tal caso, la voz es una parte fundamental en el reconocimiento de dicho estado.

Las distintas características de la voz cumplen un rol importante en la expresión de emociones, estados anímicos y actitudes. La combinación de su tono, intensidad y ritmo demuestran diferentes emociones como, tristeza, alegría, nerviosismo, etc. [1]. De esta manera, tales características pueden ser un indicio de una situación emocional vulnerable en un adulto mayor.

Dada la situación de aislamiento mencionada anteriormente, sumada a los problemas emocionales que esto puede acarrear, sería muy útil para los que están al cuidado de personas de la tercera edad, contar con un mecanismo que les permita saber cuándo los adultos mayores se encuentran en una situación frágil, ya sea de tristeza, enfermedad, u otros. Detectar esas situaciones en forma temprana permitirá tomar acciones paliativas en forma rápida, y así reducir el impacto que dicha situación podría tener en la vida de los abuelos.

Para ayudar manejar esta situación de manera no intrusiva, manteniendo la independencia de los adultos mayores, este trabajo de

memoria busca desarrollar un sistema de monitoreo remoto del estado anímico de los ancianos. Para ello habrá un dispositivo móvil, en principio una tablet, en la casa del adulto mayor, obteniendo información de su estado. Cuando el sistema detecte una situación emocional anómala, éste disparará una alarma a las personas que están al cuidado de esos abuelos; las cuales deberían tomar alguna acción remedial según sea el caso. De esa manera se busca ayudar a combatir situaciones de depresión y estados de ánimo vulnerables de los adultos mayores. Por otra parte, también se busca detectar casos de buen estado de ánimo, el cual será informado a los familiares de las personas monitoreadas, a fin de darles tranquilidad también a ellos.

Como se mencionó anteriormente, la voz cumple un rol muy importante en la expresión de emociones, y por ello, este sistema se basará en la detección de los diferentes estados de ánimo a través del reconocimiento de las características de la voz. Así, se podrá tener un sistema que informe a los familiares, mediante el envío de alertas o notificaciones, sobre la condición anímica en que se encuentran sus parientes de la tercera edad. De esta manera, las personas que actúan como monitores podrán actuar de forma oportuna ante un eventual problema.

### ***1.1. Justificación de la Propuesta***

El desarrollo del sistema de monitoreo del estado anímico de adultos mayores involucra distintos factores. Por ejemplo tenemos la captura del audio (voces captadas en el hogar donde viven los adultos mayores) y su respectivo análisis. Además tenemos el envío y registro de los resultados de ese diagnóstico a un servidor, el cual se comunicará con una plataforma de visualización de los datos para que los familiares o cuidadores puedan tomar acciones. El servidor también avisará mediante alertas (por ejemplo, enviando un sms) a las personas monitores, sobre el estado de quienes son monitoreados.

Claramente el desafío más importante tiene relación con la captura y análisis de la voz de las personas monitoreadas. Para realizar esta función es esencial contar con un buen sistema de grabación y procesamiento del audio, el cual sea capaz de separar la voz, del ruido ambiente (por ejemplo, un televisor encendido). Además, que pueda reconocer voces a pesar de que su intensidad no sea muy alta, dado que las personas monitoreadas no siempre estarán cerca del dispositivo que captura el audio. Posteriormente, damos paso al análisis de la voz, de sus diferentes características. Para realizar esto, se harán comparaciones con grabaciones anteriores, y se inferirá un estado de ánimo a partir de estos datos, con un margen de error razonable. Este desafío también incluye detectar si la voz de alguno de los adultos mayores a monitorear es parte del audio capturado, dado que la voz podría ser de otra persona, alguien que fue a visitarlos.

Asociado también a las grabaciones, es necesario controlar el espacio utilizado por éstas, para evitar el colapso de la memoria del dispositivo. Para eso, se deberá llevar un registro de las grabaciones, las cuales serían respaldadas en un servidor externo en caso de ser necesario. A este servidor, también se enviarían los resultados del análisis de la voz de las personas. La comunicación con este servidor deberá ser lo suficientemente confiable como para evitar la pérdida de datos. La solución que se proponga deberá también detectar cuándo no es posible establecer una comunicación con el servidor. En ese caso la aplicación mantendrá los datos en forma local hasta que se restablezca la comunicación con el servidor.

Por otro lado, se debe realizar un sistema de visualización que dé acceso a los familiares para ver los resultados del monitoreo de sus abuelos, y conocer sus variaciones del estado anímico. Dicho sistema contará con una sección de configuración, donde se podrá parametrizar los horarios de captura de datos, el canal para el envío de alertas, cuán seguido mandarlas, en qué casos, si se envían a una dirección de correo o no, etc. Para acceder a esta plataforma, se hará uso de un navegador Web y requerirá la autenticación del usuario. También existirá una versión móvil, para el teléfono celular, para así tener la aplicación disponible todo el tiempo que se requiera.

Tanto el software para la grabación del audio y su análisis, y también para la aplicación móvil, se desarrollarán para Android<sup>1</sup>. Se decidió esto, debido a que este sistema operativo es el que más se utiliza en los dispositivos móviles, y es de uso libre y gratuito.

## **1.2. Objetivos de la Memoria**

El objetivo principal que busca concretar esta memoria, es la implementación de un sistema remoto de monitoreo del estado anímico de adultos mayores, especialmente para los que se encuentran en una situación de aislamiento social. Para lograr ello, este sistema detecta el estado anímico mediante el procesamiento de la voz, y se encarga de alertar a las personas a cargo del cuidado de los ancianos, en caso de detectar que estos se encuentran en una situación vulnerable.

Para construir el sistema de monitoreo, hay distintos objetivos específicos que se deben alcanzar:

1. Desarrollar sistema de procesamiento y análisis de la voz, que sea capaz de capturar la voz del adulto mayor a monitorear, extraer su información e inferir su estado anímico.
2. Diseñar e implementar una aplicación móvil que permita la visualización de los estados detectados, y además, configurar de forma

---

<sup>1</sup> <http://www.android.com>

remota el sistema de monitorización y las alertas ligadas al usuario monitoreado.

## 2. Trabajos Relacionados

Este trabajo comenzó con la revisión de distintos artículos sobre temas relacionados a la detección de emociones mediante el procesamiento y estudio de la voz. Especialmente algunos que detallan distintos sistemas para monitorear pacientes con posible estado de depresión, siendo el análisis de la voz un factor importante. Es importante tomar en cuenta que a través de la voz se pueden reconocer distintos estados de una persona [7]. Básicamente son 3 estados: el primero, a corto plazo, contempla emociones, sentimientos forma de hablar, etc.; el segundo, se podría decir que de mediano plazo, engloba los estados medianamente temporales tales como una actitud (positiva o negativa), estado de salud, estado anímico (aquí entra en juego la depresión), entre otros; y por último, el de largo plazo, que abarca características biológicas y fisiológicas, como la edad y el género, como también la personalidad y la identidad de una persona. Todos estos estados son parte fundamental del estudio de esta memoria, porque no se puede reconocer el estado anímico de una voz, sin saber previamente de quién es y qué estados emocionales se encuentran involucrados.

Para llevar a cabo el reconocimiento de un estado, como los anteriormente nombrados, mediante la voz, es necesario primero contar con un conjunto de grabaciones, para posteriormente identificar los parámetros o rasgos que mejor describen este estado. Existen algunos programas que permiten extraer estos rasgos auditivos. Ya teniendo estos datos, es posible crear modelos, entrenarlos, con los cuales se pueden contrastar nuevas grabaciones (o distintas a las usadas para la creación del modelo) y así verificar si el modelo entrenado es lo suficientemente preciso, o revisar en qué se puede mejorar para aumentar su precisión. En resumen se identifican 4 principales pasos: 1-Grabación de audio; 2-Extracción de características; 3-Generación de un modelo que describa el estado; y 4-Pruebas con el modelo.

### 2.1. Conjunto de Características Relevantes de la Voz

Son muchas los descriptores o características auditivas que se pueden extraer de una pista de audio, pero sólo algunas son más descriptivas al momento de querer identificar una emoción. Entre los rasgos que describen mejor una emoción están el Pitch o F0, Energy, Loudness, MFCCs (Mel Frequency Cepstral Coefficients), Formants, entre otros [2][3]. El Pitch o F0, es la frecuencia fundamental que caracteriza a un sonido, y dependiendo de su valor permite clasificarlo como tono alto, bajo o medio, estando el tono y el Pitch estrechamente relacionados. Los MFCCs, son coeficientes que representan el habla basados en la percepción humana, y son altamente

utilizados en el reconocimiento de voz, reconocimiento de palabras, frases, y últimamente también en el reconocimiento de emociones.

Por sí solas estas características no suelen ser suficientes, por lo que siempre se agrupan con otras, y además se extraen de ellas valores a través del tiempo, cómo el máximo, mínimo, valor promedio, la desviación estándar, los cuales pueden dar una representación más fidedigna de una emoción, para diferenciarla de otra.

## **2.2. Programas Revisados**

Para extraer las características de un archivo de audio, existen algunos programas especializados en ello. Los programas probados fueron openSMILE [4] y Yaafe<sup>2</sup>. También se revisó HTK<sup>3</sup>, pero su propósito está enfocado en el reconocimiento de voz, es decir, transcripción de audio a texto, contemplando el proceso completo de ésta, sin dar mucha libertad de utilizar sólo algunas herramientas de éste, para pruebas básicas.

Yaafe está hecho especialmente para la extracción de características de cualquier audio, sea música, voz, etc., con una interesante arquitectura de procesamiento (en C++) y generación de resultados en formato CSV, un formato fácil de manejar. Su uso es bastante sencillo, amigable al usuario, además puede usarse con Python y Matlab, entre otras cosas. Es un programa de uso bastante general, en el sentido de que su tarea sólo apunta a la extracción de características de un audio.

Por otro lado, tenemos openSMILE. Este conjunto de herramientas se especializa en la extracción de características sonoras para el procesamiento de voz. Para ello, trae consigo un conjunto de descriptores de audio específicamente para el reconocimiento de emociones, y que pueden ser utilizados fácilmente. Además, entrega resultados en distintos formatos según lo que se necesite, entre ellos, archivos para procesar con máquinas de soporte vectorial (SVM). En su última versión, permite desde la grabación de audio, pasando por la extracción de descriptores del audio, la posterior generación de modelos usando SVM, hasta algunos scripts para graficar, entre otras cosas.

Teniendo ya un conjunto de programas seleccionados, pasamos a la parte de las pruebas. El programa o conjunto de herramientas que mejor se adapta al enfoque de esta memoria es openSMILE, especialmente por la base en que fue construido, para el procesamiento de voz, reconocimiento de emociones, y por su versatilidad al momento de hacer pruebas. Yaafe por su parte, es una buena opción como alternativa, o apoyo, en caso de querer extraer rasgos con algún formato diferente.

---

<sup>2</sup> <http://yaafe.sourceforge.net/>

<sup>3</sup> <http://htk.eng.cam.ac.uk/>

### 3. Pruebas Realizadas

En esta sección se presenta un resumen de las distintas pruebas realizadas en cuanto a detección de género, emoción, identidad, y extracción de trozos de audio hablado (frases y palabras). Aquí, se describen de forma más general los pasos que se realizaron y los resultados a los que condujeron.

En la sección Anexo A se puede encontrar con total detalle todas las pruebas que se llevaron a cabo, dando cuenta tanto de los hallazgos exitosos, como de los no exitosos, y las decisiones que se tomaron.

Para un mejor entendimiento del lenguaje técnico ligado a la clasificación de datos, se presenta a continuación un glosario con los términos más importantes [9]:

#### **Glorario de clasificación**

- SVM (Support Vector Machine): Técnica de clasificación de datos, que permite conocer a qué grupo pertenece una muestra de testeo, a partir de una función y las características extraídas de las muestras de entrenamiento.
- Clase: Es el objetivo que se quiere obtener al clasificar una muestra. Por ejemplo: Una muestra de voz masculina debería ser clasificada como perteneciente a la clase de voces masculinas.
- Característica o descriptor: Atributo que describe a una muestra.
- Vector o conjunto de características: Conjunto de atributos que permiten diferenciar a una clase de otra.
- Modelo: Predice, en base a los datos de entrenamiento, a qué clase pertenece una muestra del conjunto de testeo.
- Kernel: Función que permite separar el plano de los datos en diferentes hiper-planos, cada uno correspondiente a cada clase.
- Variable Gamma: Parámetro del Kernel.
- Variable C: Parámetro de penalización del error (cuán máximo puede ser el error obtenido).

Se añade también un glosario con los términos relacionados al uso del programa openSMILE:

## **Glosario de openSMILE**

- LLD (Low Level Descriptors): son aquellas características prosódicas o también llamadas supra-segmental, como el Pitch, la Intensidad, Duración, Calidad de la Voz (Voice Quality), entre otros [8].
- Functionals: Los Statistical Functionals o simplemente Functionals son funciones matemáticas que entregan algún valor específico, y proveen una especie de normalización a lo largo del tiempo en la muestra analizada [8].
- Archivos de configuración de openSMILE: archivos que permiten realizar alguna tarea sobre un archivo de audio, como es la extracción de características. Entre ellos están los siguientes, los cuales fueron utilizados durante las pruebas [5]:
  - IS10\_paraling: The INTERSPEECH 2010 Paralinguistic Challenge feature set. Contiene 1582 características que resultan de la base de diferentes LLD y sus Functionals. Dentro de los LLD utilizados están Pitch, Loudness (o Intensidad), Jitter, MFCC, MFB, LSP.
  - emobase2010: The openSMILE 'emobase2010' reference set. Este conjunto representa el actual estado del arte del conjunto de características para reconocimiento emocional y paralingüística. Está basado en el conjunto de características de IS10\_paraling. Es exactamente igual a IS10\_paraling, pero con la única diferencia que algunos LLD se encuentran normalizados.
  - emo\_large: The large openSMILE emotion feature set. Permite la extracción de 6552 características. Entre los LLD que contiene están Energy, MFCC, Mzcr, Pitch, Spectral, LSP.
  - prosodyShs: Este archivo extrae las siguientes características prosódicas: F0, Voicing Probability, y Loudness .

### **3.1. Reconocimiento de emociones**

#### **Objetivo**

El objetivo de estas pruebas fue encontrar la mejor forma de clasificar emociones, de tal manera que también fuera posible identificar aquellas emociones vulnerables, que por ahora solo incluyen a la Tristeza.

#### **Muestras**

Durante el proceso de trabajo, fue posible obtener acceso a tres bases de datos emocionales, SAVEE, Emo-DB y RekEmozio, con 7 emociones cada una. Las emociones son Alegría, Enojo, Miedo, Asco, Tristeza, Neutral y Sorpresa, excepto para Emo-DB, que en vez de traer Sorpresa, trae muestras de Aburrimiento. Los archivos de audio obtenidos de estas bases

de datos fueron utilizados tanto para entrenamiento, como para testeo, aproximadamente la mitad en cada conjunto.

## **Resultados**

Primero se partió separando cada una de las emociones a nivel de cada base de datos, y no se incluyó Aburrimiento. Esto arrojó resultados bastante bajos, pero ayudó a detectar las confusiones que había entre distintas emociones. Estas confusiones llevaron a agrupar las emociones. Al principio fueron dos grupos, Enojo-Alegría-Sorpresa-Miedo y Asco-Neutral-Tristeza. Los resultados mejoraron, pero la emoción Asco aún seguía generando mucho ruido, por lo que se decidió dejarla fuera, además de que es una emoción poco común y no caería en la categoría de vulnerable.

Después de haber sacado el Asco, seguía habiendo algo de ruido a causa de la emoción Miedo. Entonces se tomaron las muestras de esa emoción y se formó un grupo aparte. Ahora eran tres grupos. Al contrastar cada base de datos consigo misma y con las otras, los resultados eran algo mejores, pero no lo suficientemente buenos.

Se probó unir todas las muestras de entrenamiento de cada base de datos. Al hacer el contraste, los resultados fueron muy buenos, logrando clasificar casi todas las muestras correctamente.

Esta agrupación dio buenos resultados para los grupos. El siguiente paso era poder separar bien la Tristeza. Siguiendo la misma técnica, se probó uniendo todas las muestras de entrenamiento para Neutral y Tristeza, formando dos conjuntos por separado. Entonces, la idea era usar este modelo basado en Neutral y Tristeza y contrastarlo con las muestras que cayeron en el grupo Neutral-Tristeza. Los resultados fueron medianamente buenos, con más de la mitad verdaderos positivos, pero también con varios falsos positivos.

Se revisó la literatura en búsqueda de una mejor separación. Varios estudios mencionaban Arousal y Valence, conceptos definidos como dimensiones afectivas, el primero relacionado con la agitación o calma de las emociones, y el segundo relacionado con el placer, es decir, si la expresión de la emoción es placentera o no. Con esto se llegó al Circumflex of Affect creado por Russell, que separa a las emociones en cuadro cuadrantes en base a Arousal y Valence.

De las múltiples pruebas hechas en base a las dimensiones afectivas Arousal y Valence, la que arrojó mejores resultados fue en la cual se separaron las emociones según los cuatro cuadrantes, para el modelo de entrenamiento. Coincidió con que la Tristeza se encuentra sola en un cuadrante, haciendo más fácil su clasificación. Aunque arrojaba buenos resultados, al probarse directamente sobre las muestras ya clasificadas en Neutral-Tristeza por la clasificación de tres grupos anterior, los resultados

mejoraron mucho más, dejando un menor número de falsos positivos, y más verdaderos positivos.

Este buen efecto de doble filtro fue el que se dejó para el reconocimiento de emociones y clasificación específica de la Tristeza. No hay que olvidar mencionar que estos modelos se basaron en el conjunto de características *emobase2010*, usando el kernel lineal.

### **3.2. Reconocimiento de género**

#### **Objetivo**

Reconocer la emoción en una voz grabada, no es suficiente si no se sabe de quién es la voz. Para ello, el paso siguiente más fácil es reconocer el género. Fácil en el sentido que esto permite inmediatamente decir si la persona es de un grupo u otro.

#### **Muestras**

Inicialmente se contó con muestras de grabaciones de aproximadamente tres minutos, hechas por 5 hombres y 5 mujeres, cada uno leyendo algún texto. Posteriormente se agregaron algunas muestras de las bases de datos emocionales.

#### **Resultados**

Inicialmente se había probado en las muestras de los 5 hombres y 5 mujeres, con el conjunto de características *emobase2010* y kernel polinomial, y todo había salido bien, con resultados cercanos al 100%. Cuando se incluyeron las muestras con emociones, los resultados en estas muestras fueron diferentes, especialmente en muestras de voz femenina.

Se probó con otro conjunto de características, *IS10\_paraling*, y con el kernel lineal. La combinación de estos logró mejores resultados, pero aún se podía mejorar. La normalización del volumen logró una pequeña mejora, y se mantuvo.

Más tarde, al probar con muestras de audio de conversaciones, los resultados no fueron tan bueno como se esperaba. Había algo que generaba distorsiones en la clasificación. Para ello se revisó literatura específica sobre reconocimiento de género, se hicieron pruebas con lo encontrado en esos estudios, donde se mencionaban otras características a extraer, pero no se logró obtener mejoras, de hecho fueron peores. También se probó si el idioma afectaba, pero tampoco sirvió.

Aunque los resultados fueron desalentadores, se había detectado que la emoción, en especial las emociones más fuertes, generaban cierta distorsión en las clasificaciones. Por lo tanto, se decidió probar clasificando según género, a partir de una emoción dada. Es decir, se tomaron muestras de una emoción, se separaron en hombre y mujer, y se contrastaron. Esto se

hizo para todas las emociones. Los resultados fueron bastante buenos, y se comprobó que al normalizar el volumen de las muestras en las emociones más calmadas, Neutral y Tristeza, los resultados eran un poco mejores.

Finalmente se dejó esta última clasificación como la mejor opción, en la que el conjunto de características usado fue *IS10\_paraling*, y se usó el kernel lineal.

### 3.3. Reconocimiento de identidades

#### Objetivo

El último paso de reconocimiento es el del reconocimiento de identidad, después de ya haber reconocido el género de la persona. Este paso se podría decir que es casi el más importante, porque asegurará que el audio analizado corresponde a la persona monitoreada, y con eso se tendrá la seguridad al entregar el dato del estado reconocido.

#### Muestras

Las muestras utilizadas fueron las correspondientes a las grabaciones de los 5 hombres y 5 mujeres iniciales. Esto porque se contaba con hartas muestras para cada uno, lo que permite un mejor análisis.

#### Resultados

En este tipo de clasificación no se hicieron muchas pruebas. La primera fue usando el vector de características *emobase2010* y kernel polinomial. En general la clasificación fue buena, con resultados sobre el 70%. Al probar con kernel lineal, estos últimos mejoraron, siendo sobre el 80% en promedio.

También se probó con el vector *IS10\_paraling*, pero los resultados fueron mucho más bajos, la mayoría bajo el 50%.

Hasta ese momento se había decidido por usar el kernel lineal y *emobase2010*, pero las posteriores pruebas en conversaciones, demostraron que usando *IS10\_paraling* se lograban resultados un tanto mejores. Esto se debería a que al estar *IS10\_paraling* más enfocado en las características lingüísticas de la voz, sería más realista al momento de clasificar en el campo de las identidades, como también lo fue para el género.

### 3.4. Extracción de Frases y Palabras

El archivo de configuración *prosodyShs.conf*, además de incluir valores del Pitch e Intensidad por cada frame del audio (10ms), también calcula la probabilidad de voz, Voice Probabitlity o *voicingFinalUnclipped*, para cada uno de esos frames. El valor de *voicingFinalUnclipped* va entre 0 y 1, indicando el porcentaje de probabilidad de que ese frame el audio corresponda a voz.

En la sección A.2 del anexo, se menciona que cuando el valor de F0 es 0, indica que había una pausa en el habla. Esto fue sólo una aseveración con

respecto a los datos vistos en el CSV que entrega openSMILE al procesar un audio con *prosodoyShs.conf*, y la comparación directa con el espectro de audio. Pero si se cuenta con una medición de probabilidad, que ya está basada en estudios validados, convenía más utilizar esos valores, los valores de *voicingFinalUnclipped*.

Entonces, el paso siguiente después de obtener el valor de *voicingFinalUnclipped*, fue decidir qué valores o qué umbral indicaría que el audio grabado correspondía a voz. En [23] se encontró que usaban el valor de 0,55 como umbral, para discriminar voz y silencio. Al comparar el espectro de un archivo de audio de las muestras, y el resultado entregado en el archivo CSV, se pudo ver que cuando el valor de *voicingFinalUnclipped* superaba el 55%, también F0 era distinto de cero, por lo que no estaba tan lejos el razonamiento realizado anteriormente. Pero no coinciden exactamente, *voicingFinalUnclipped* incluye más frames en comparación con los que tienen valor F0 distinto de cero, así que se dejó *voicingFinalUnclipped* como el indicador oficial de voz.

Hay que considerar un factor muy importante al utilizar el valor de *voicingFinalUnclipped*. Los valores que entregan no son los mismos si se cambia la frecuencia de muestreo de la pista de audio. Se probó con 8 y 16 KHz, y los valores de *voicingFinalUnclipped* coincidían con los estimados, al igual que el valor de F0, pero al probar con audios de 44100 Hz, fue más difícil distinguir las pausas, los valores de *voicingFinalUnclipped* eran sobre el 55% incluso en pausas, y los mismos con F0, siendo distinto de cero en esas pausas. Por lo que se recomienda utilizar pistas de audio con una frecuencia de muestreo no tan alta, esto parece reducir ruidos, especialmente el ruido blanco, y mejora la detección de probabilidad de voz.

¿Por qué es tan importante conocer cuándo hay o no voz en una pista de audio? Cuando las muestras ya vienen procesadas, cortadas, no hay necesidad de preocuparse de esto. Pero al enfrentarse a una situación real, el sistema tendrá que grabar el audio, y de forma automática tendrá que procesarlo, y aquí es donde entra en juego el saber si hay o no voz en la grabación, para llevar a cabo la extracción de palabras y frases de forma precisa.

En el reconocimiento de emociones, la duración de las pistas de audio es un factor importante al momento de clasificar. A esto se le llama unidad de análisis, la que puede ser una palabra, una frase o *chunk*, o un turno [8] [24]. Un turno es cuando una persona termina de hablar y habla después la otra persona, si es que hay otra persona. Por lo visto en [24], la mejor unidad de análisis para la clasificación en emociones es el *chunk*, y para ello sería necesario conocer cuándo hay una pausa y cuando hay voz.

## **Extracción de palabras**

Lo primero que se hizo antes de probar separando frases o *chunks*, se separaron las palabras. En pocos detalles, cada vez que el valor de *voicingFinalUnclipped* de un frame superaba el 55%, se mantenía la cuenta de frames que también lo hacían, hasta que aparecía alguno con un valor menor al 55%, y se consideraba que hasta el anterior frame había una palabra.

## **Extracción de frases**

Ya habiendo extraído las palabras, para extraer las frases se calculó el tamaño de las pausas entre cada palabra, y cuando la pausa era mayor a cierto tiempo, se consideraba que una frase se había encontrado. ¿Pero qué valor de la duración de una pausa era suficiente para decir que con ello terminaba una frase, y posterior a la pausa comenzaba otra? En [8] se menciona que una pausa de 0,5 o 1s puede ser apropiada para este propósito. Al probar con 0,5 se observó que se obtenían más frases de las que se deberían, en su mayoría siendo sólo palabras sueltas. Probando con un número mayor, se llegó a que 0,6 era un tamaño de pausa apropiada para diferenciar una frase de otra.

Tanto para la extracción de frases, como de palabras, se creó un script especial que realiza estas tareas automáticamente, y se encuentra detallado de manera más completa en la sección 5.1.

## 4. Concepción de la Solución

En este capítulo se detallan los requisitos de la solución y las tecnologías a utilizar. Además se describe la arquitectura del sistema completo, con sus respectivos componentes, y al final, mediante unos diagramas, se muestra de forma más minuciosa el diseño de la solución desarrollada en este trabajo de memoria.

### 4.1. Requisitos

Los principales requisitos de la solución propuesta son:

- A partir del audio grabado, se debe poder detectar la emoción o el grupo de emociones al que pertenece el estado expresado mediante la voz.
  - Al reconocer la emoción da inmediatamente una primera impresión del estado general en que se encuentra el abuelo monitoreado. Esto puede decir mucho sobre su comportamiento y salud emocional.
- La solución debe determinar si la emoción detectada corresponde a un estado vulnerable en el usuario monitoreado.
  - Uno de los objetivos principales de este trabajo es determinar cuándo un estado detectado es vulnerable, para así poder llevar a cabo alguna acción que permite una mejoría de la persona monitoreada.
- Para el punto anterior es fundamental que el sistema reconozca fielmente de quién es la voz grabada, asegurándose de que es o no el usuario monitoreado en cuestión.
  - Es importante que se conozca realmente si la voz grabada corresponde al usuario monitoreado. Si por alguna razón se grabara otra voz, ya sea de la televisión o una visita, y no se reconociera la identidad de la persona, se podría estar generando falsas alertas o agregando ruido. Esto llevaría a que a mediano o largo plazo se tenga una percepción diferente y errónea del estado real del abuelo o abuela en monitoreo.
- El sistema de detección debe alertar, de forma oportuna, a los familiares cuando un usuario se encuentra en un estado vulnerable, asegurándose de que la alerta llegue y sea vista.
  - Los familiares son los principales responsables de velar por la salud de sus abuelos. Si se les puede avisar de forma oportuna una

situación vulnerable en sus adultos mayores, esto será de mucha ayuda.

- Se debe mantener un historial de los estados detectados para cada usuario.
- Esto es para que se pueda tener una visión general, tanto a corto como a largo plazo de la salud emocional de los adultos mayores monitoreados.
- Se debe disponer de un sistema de visualización que dé acceso a los familiares para ver el historial de estados de sus abuelos.
  - Toda información recopilada puede ser mostrada de distintas maneras. Al ser datos técnicos, es vital que el sistema pueda proveer una traducción de estos datos, para su sencilla lectura y fácil análisis. En base a esto, el sistema de visualización permitirá a cualquier usuario revisar el historial de estados, sin necesidad de grandes conocimientos.

#### **4.2. Tecnologías y Productos de Software a utilizar**

Durante el proceso de pruebas que se describen en el capítulo 3, diversas tecnologías se utilizaron para llevar a cabo esas pruebas. Al probar una y otra, solo algunas de estas tecnologías se eligieron para ser parte de la solución:

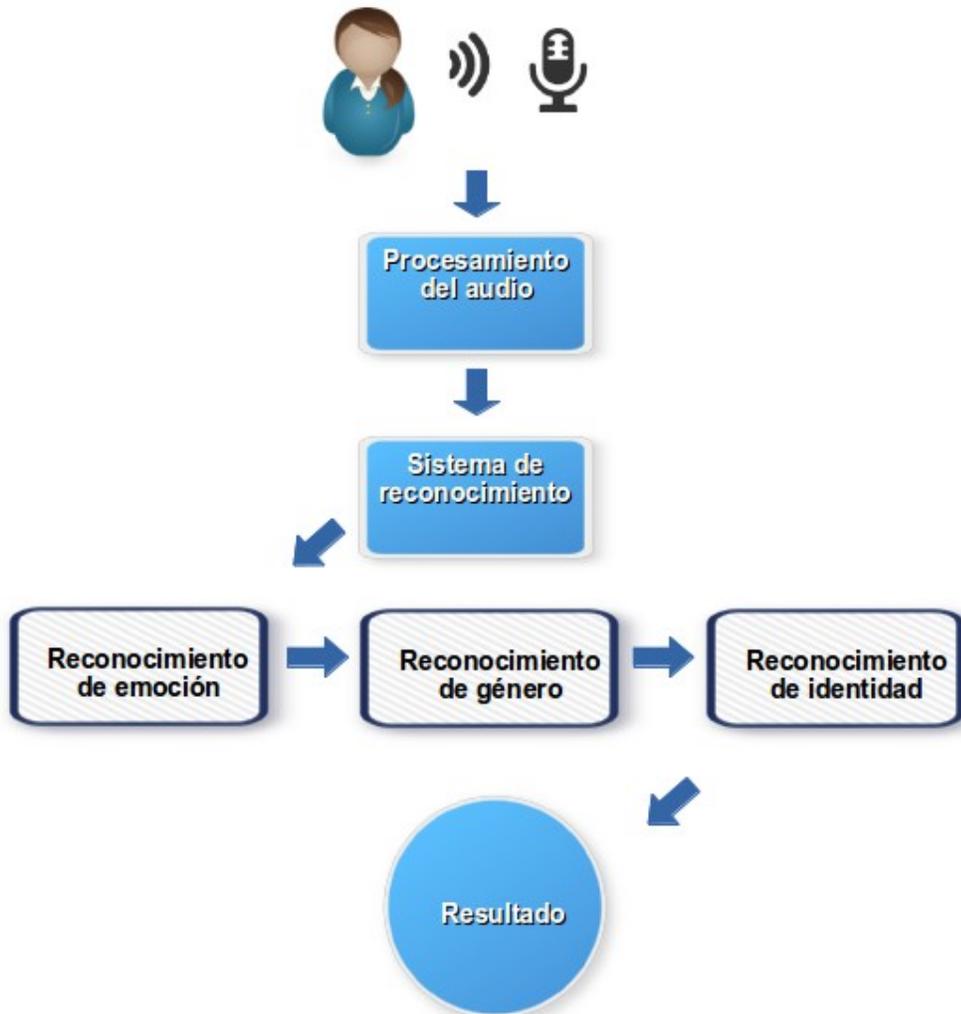
- openSMILE: Conjunto de herramientas de código abierto, para el análisis de la voz. Permite la grabación de audio y extracción de características auditivas de éste. Además, incluye pequeños programas para graficar, usar SVM, entre otros. Todas estas herramientas son solo accesibles mediante consola.
  - Este programa fue elegido principalmente por el enfoque por el cual fue creado, el reconocimiento de estados en la voz de una persona (emoción, género, edad, etc).
  - Además de traer un gran número de conjuntos de características, es posible crear fácilmente uno nuevo, solo eligiendo las opciones básicas. El cálculo respectivo para cada característica lo hará automáticamente al pedirle la extracción de ellas para algún audio. Todo esto asegura rapidez y confianza en los datos.
  - Este conjunto de herramientas ha sido probado y utilizado un sin número de veces en los distintos desafíos de InterSpeech y AVEC, mostrando que cumple con el propósito por el cual fue creado.
- LIBSVM: Conjunto de programas de consola que posibilitan el entrenamiento y predicción de modelos, utilizando distintos tipos de

kernel: lineal, polinomial, RBF, entre otros. La construcción de modelos SVM se realiza a partir de las características extraídas previamente. Luego, la herramienta de predicción permite contrastar un archivo de características de entrada con el modelo SVM ya construido.

- Es una de las herramientas más usadas para el entrenamiento de modelos SVM, lo que valida su buen rendimiento.
- Permite procesar una gran cantidad de datos de forma rápida y precisa.
- Además de poder crear modelos y testarlos, da la posibilidad de encontrar cuáles son los mejores parámetros para el modelo que se quiere crear. Esto se logra con el script *grid.py* que trae incorporado.
- Se encuentra disponible para múltiples lenguajes, por lo que si cambia el lenguaje utilizado en los scripts desarrollados (sección 5.1), no debería haber mayores dificultades.
- SoX (Sound eXchange): programa para línea de comandos que permite la grabación, reproducción y edición de archivos de audio. Ofrece una gran variedad de efectos de edición y soporta múltiples formatos de audio.
  - Al principio, la edición de los archivos de audio se hacía manualmente, usando Audacity. SoX reemplaza a Audacity, pudiéndose usar mediante línea de comandos.
  - Al ser un programa de consola posibilita su uso de forma automática para las distintas tareas de procesamiento de audio, como cortar, cambiar de formato, normalizar volumen, etc.

### **4.3. Arquitectura de la solución**

El proceso concebido para la solución requerida, se basa en el último enfoque del proceso de reconocimiento (ver Ilustración 24 en sección A.3.11). Antes de ir directamente a lo que es el reconocimiento propiamente tal, hay que recordar que primero está la grabación del audio. Posteriormente viene el procesamiento de éste, ya sea cambiando la frecuencia de muestreo, como también extrayendo las frases que la grabación contiene. Se recuerda que una frase fue la unidad de análisis elegida. Luego viene el proceso de reconocimiento, primero clasificando según emoción, le sigue la clasificación por género, hasta llegar al reconocimiento de identidad (Ilustración 1).



*Ilustración 1: Grabación, procesamiento del audio y proceso de reconocimiento.*

Cuando se ha logrado obtener un resultado satisfactorio, es decir, se pudo reconocer la emoción y la identidad de la persona que fue grabada, se prosigue con la actualización de los datos. Los nuevos datos se almacenan, y luego se hace la pregunta más importante ¿Es la emoción detectada vulnerable? Es primordial responder a esta pregunta lo más pronto posible. Si lo es, se debe alertar inmediatamente a los familiares o cuidadores que estén ligados al usuario monitoreado (Ilustración 2).



Ilustración 2: Almacenamiento de datos y alerta en caso de estado vulnerable.

#### 4.4. Diseño de la solución

Con el fin de poder determinar la emoción expresada, el género y la identidad de la voz grabada, el sistema de reconocimiento, para cada uno de estos estados (Ilustración 1), cuenta con una serie de pasos. Los pasos se dividen en dos procesos principales (Ilustración 3): Entrenamiento de muestras y Testeo de muestras. El primero genera el modelo que contiene las características que representan a cada clase, y el segundo predice,

gracias al modelo, a qué clase corresponde cada muestra, pudiendo reconocer así su estado respectivo (emoción, género o identidad).

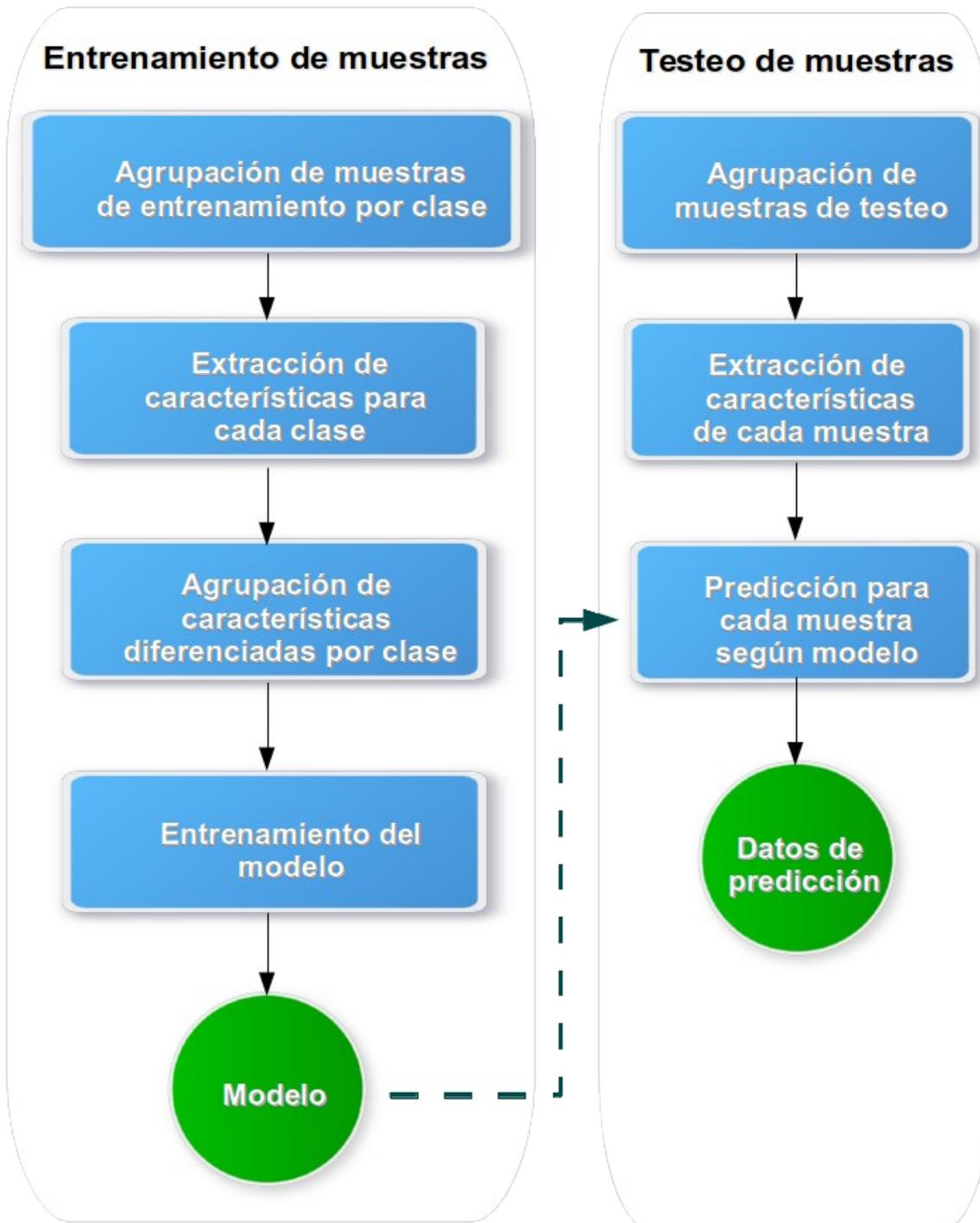


Ilustración 3: Diagramas de flujo para los procesos de entrenamiento y testeo de muestras.

Parte de la solución también concibe la aplicación móvil que forma parte del sistema de visualización de datos. El siguiente diagrama muestra el flujo de navegación para esta aplicación, donde aparecen las pantallas principales de ésta. Monitor contiene las opciones de configuración de grabación, y, Alertas las opciones de envío de alertas (sms, email, etc.).

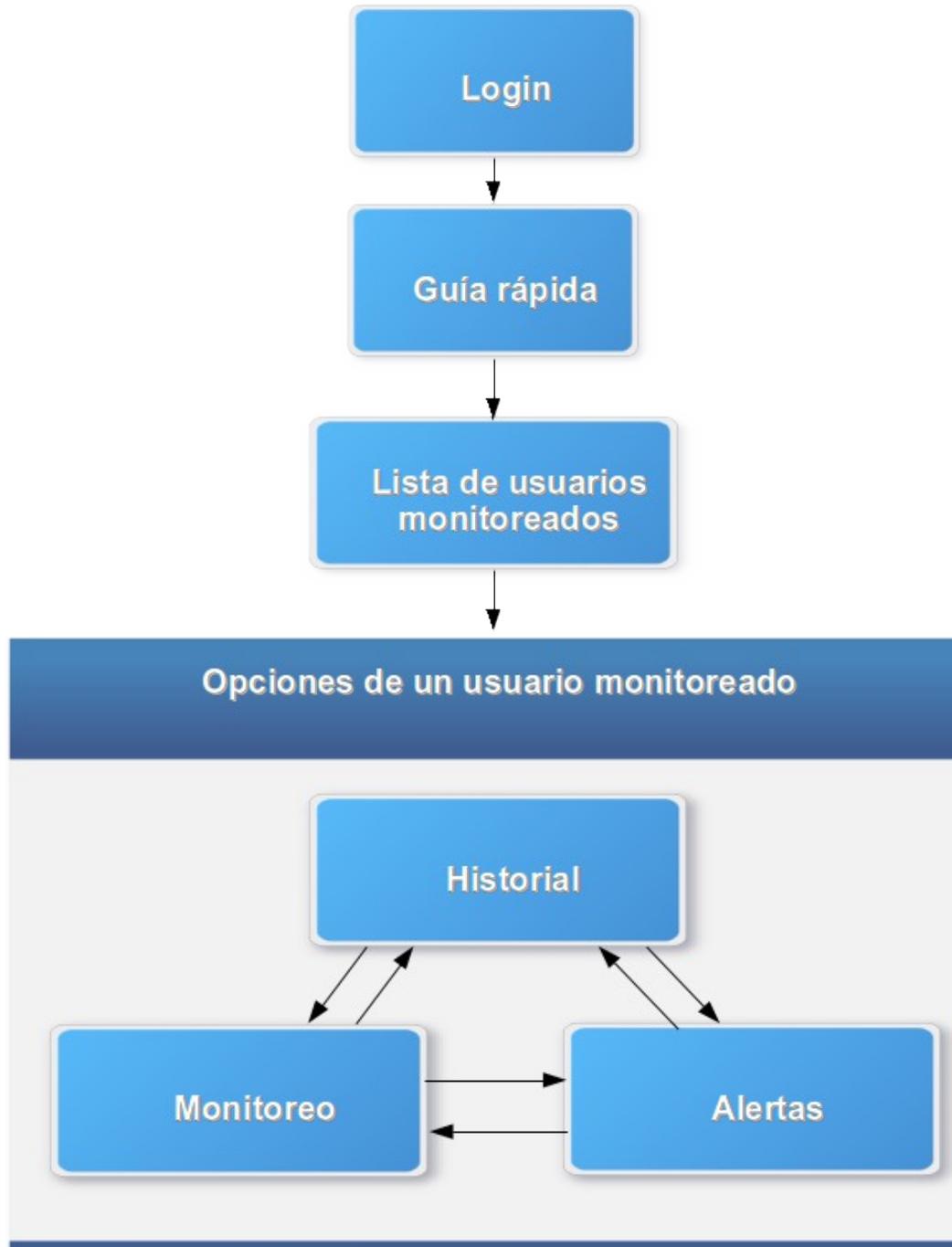


Ilustración 4: Diagrama de flujo de navegación de la aplicación móvil para Android.

## 5. Implementación del Sistema

Esta sección está dividida en dos partes. La primera está enfocada en los programas desarrollados para el entrenamiento y testeo de las muestras para las pruebas hechas y descritas en la sección 3 y Anexo A. La segunda parte contiene los detalles de la construcción de la aplicación móvil diseñada para el sistema de visualización de datos.

### 5.1. Scripting

En esta sección se detallan cada uno de los *scripts* o pequeños programas que se desarrollaron para la automatización de tareas llevadas a cabo en las distintas pruebas realizadas.

Hay variadas opciones en los distintos *scripts* que fueron configuradas a mano, y se encuentran directamente en el código. Esto se hizo por comodidad y facilidad de uso, porque tomaría mucho tiempo adaptar estos *scripts* para que aceptaran diferentes argumentos y que al ejecutarlos se supiera bien lo que se estaba haciendo.

#### 5.1.1. Creación de modelos SVM

La creación de un modelo usando SVM consta de varios pasos: Extracción de características, definición de las clases, definición de los parámetros y entrenamiento del modelo.

El script `create_models_from_folders.sh`, escrito en Bash, recibe como parámetros la ruta de la carpeta donde se encuentran las muestras de entrenamiento, y se asume que éstas se encuentran correctamente agrupadas en sus respectivas sub-carpetas, correspondiendo cada una de ellas a una clase. El otro parámetro que se recibe es el nombre del conjunto de características a utilizar. Se asume también que éste se encuentra junto a los archivos de configuración de openSMILE.

Primero se revisa si en el directorio de las muestras ya se extrajeron características para el conjunto definido. Si no es así, entonces, por cada sub-carpeta se extraen con openSMILE las características correspondientes de cada uno de sus archivos de audio, y todas éstas se almacenan en un único archivo que tendrá formato `arff`. Posteriormente se transforma cada uno de estos archivos a formato `lsvm`, que es el tipo de formato que acepta LIBSVM.

El siguiente paso es llamar a `create_models.sh`, script que recibe como parámetros todos aquellos que se necesitan para la creación de un modelo, como `svm_type`, `kernel`, `degree`, `gamma`, etc. Además recibe el nombre del conjunto de características, el directorio donde están los archivos `lsvm` y el nombre que tendrá el modelo. Recibiendo todos estos parámetros, toma

cada uno de los archivos lsvm, y les asigna una clase, definida con un número entre 1 y el número total de sub-carpetas, que es equivalente al número de clases. Finalmente se llama a al programa svm-train de LIBSVM con todos los parámetros necesarios.

### **5.1.2. Predicción de modelos SVM**

Se crearon dos scripts para predecir o contrastar los modelos SVM creados, ambos escritos en Bash.

- test\_model\_folders.sh:
  - Este script se hizo para predecir un modelo sobre muestras de prueba que se encuentran agrupadas con respecto a su correspondiente clase.
  - Recibe como parámetros la ruta del directorio donde se contrastará el modelo, la ruta del modelo y el nombre del conjunto de características.
  - Su principal función es extraer las características de los archivos de cada sub-carpeta, si es que aún no se ha hecho, y a partir de ellas crea un único archivo lsvm para cada sub-carpeta. Luego contrasta estos archivos con el modelo ingresado como parámetro.
  - Los resultados finales de predicción quedan en un archivo con sufijo "predict". Con un script detallado más adelante, se calcula el promedio de los porcentajes de exactitud por clase, guardándolos en otro archivo de sufijo "predict-results".
- test\_model\_audio\_files.sh:
  - En vez de testear el modelo sobre muestras separadas en sub-carpetas, lo que hace es contrastarlo sobre muestras sueltas dentro de una carpeta.
  - Recibe los mismos parámetros que el script anterior, es decir, la ruta del directorio a testear, la ruta del modelo, y el nombre del conjunto de características.
  - Extrae las características de cada uno de los archivos, almacenándolas todas en un único archivo lsvm. Este archivo es contrastado con el modelo, y los resultados son guardados en un archivo de sufijo "predict".
  - Con el mismo script que se calculan los promedios de los porcentajes de exactitud por clase, esta vez se crean sub-carpetas correspondientes a cada una de las clases, y se copian en cada una de ellas los archivos de audio que fueron clasificados en esas clases. Esto último permite que se puedan hacer pruebas posteriores en

estas sub-carpetas, lo que se hizo pensando específicamente en un proceso de reconocimiento que consta de varios pasos de clasificación.

### **5.1.3. Procesamiento de resultados**

En los programas para predecir un modelo se dijo que los archivos de los resultados eran procesados posteriormente. Esto lo hace el script escrito Python `read_predict.py`, el cual cuenta con tres funciones principales:

- Calcular promedios de porcentajes exactitud por clase:
  - Lee el archivo "predict" que tiene los resultados y agrupa estos por clase. Al final para cada clase calcula el promedio de los porcentajes y los escribe en un archivo de sufijo "predict-results".
- Añadir pista de audio a resultados:
  - Para cada resultado en el archivo "predict" se agrega al lado el nombre del archivo de audio que arrojó ese resultado al contrastar el modelo. Esto puede servir para analizar en detalle los porcentajes de predicción para cada muestra.
- Clasificación de muestras:
  - A partir de los resultados en "predict", se crean sub-carpetas para cada clase del modelo y se copian en ellas las muestras que resultaron clasificadas en esas sub-carpetas.

### **5.1.4. Extracción de palabras y frases**

Se creó un script llamado `read_prosody_csv.py`, escrito en Python también, por la facilidad que presenta este lenguaje para la lectura de archivos CSV y el manejo de objetos tan útiles como listas y diccionarios. Este programa recibe como parámetros principales la ruta del archivo CSV generado por openSMILE al usar el archivo de configuración `prosodyShs.conf`, la ruta del archivo de audio a procesar, lo que se quiere extraer y el directorio final donde dejar el audio extraído.

El primer paso es identificar todos los trozos de audio entre cada pausa. A estos trozos se les llama "cortes". Se les podría llamar directamente palabras, pero no se hizo porque no siempre lo son, a veces estos cortes son de una duración tan pequeña que no es posible definirlos como palabras. Entonces, los cortes se identifican leyendo el valor de `voicingFinalUnclipped` del archivo CSV, y diferenciando entre valores mayores y menores a 55% (explicado en 3.4). Para cada corte se guarda tiempo de inicio, fin y duración.

Ya teniendo los cortes, el siguiente paso es extraer frases o palabras. En el caso de palabras es casi inmediato, aunque antes se filtran los cortes

eliminando aquellos de duración menor a 0,2s, que en general son ruidos sin sentido. Los cortes restantes que quedan son las palabras. Luego de obtener las palabras, se procede a generar un archivo de audio por cada una de ellas. Esto se hace usando SoX, que gracias a la opción trim, permite extraer un trozo de audio de un archivo tan solo dándole el momento de inicio y la duración.

Para la extracción de frases lo que se hace es recorrer cada uno de los cortes y analizar la pausa que hay entre ellos, si la pausa es menor a 0,6s se van acumulando los cortes. Cuando se encuentra una pausa mayor a 0,6s se crea una frase con los cortes acumulados. Se repite el proceso hasta analizar todos los cortes y pausas. Con las frases conseguidas se procede finalmente a utilizar SoX para extraerlas del archivo de audio, de la misma manera que al extraer las palabras.

### **5.1.5. Procesamiento de audio**

Para procesar los archivos de audio con respecto a los efectos a aplicar en ellos, se crearon distintos scripts que realizaran estas tareas. Todos ellos reciben como parámetro la ruta del archivo de audio y el valor a utilizar.

- `convert_audio_file.sh`:
  - Modifica la frecuencia de muestreo de un archivo de audio.
  - Recibe como segundo parámetro la frecuencia de muestreo.
  - El resultado final es el mismo archivo, con un solo canal, Mono, y con la frecuencia de muestreo elegida.
- `normalize_audio_file.sh`:
  - Normaliza el volumen de audio de un archivo.
  - Recibe como segundo parámetro el nivel de normalización a utilizar.
  - Reemplaza el archivo original con el mismo, pero con volumen normalizado.
- `dc_shift.sh`:
  - En la sección A.3.8 del anexo se menciona que algunos archivos no se encontraban perfectamente centrados, estaban desfasados. Este script permite centrarlos.
  - Primero llama a SoX con los parámetros `-n stats`, esto entrega el DC offset, o desfase. A partir de este dato se aplica un desfase con el valor contrario, intentando centrar la señal en el archivo de audio. Para ello se usa SoX con el parámetro `dcshift` y el valor de desfase.

## 5.2. Aplicación Móvil

De forma complementaria al trabajo de Memoria, se llevó a cabo un proyecto de Usabilidad en el curso Taller de Usabilidad de Interfaces de Software, código CC6502, del Departamento de Ciencias de la Computación de la Universidad de Chile. En este proyecto de Usabilidad se diseñó, implementó y evaluó la interfaz móvil que se usará para el sistema de visualización propuesto en este trabajo de Memoria.

En las siguientes líneas se encuentra un glosario que describe los términos más importantes relacionados con el proyecto de Usabilidad llevado a cabo [28]:

### **Glorario**

- **Usuario experto:** Usuario que cuenta un nivel avanzado de conocimientos y experiencia en el ámbito de la Usabilidad. También tiene experiencia en el uso de los dispositivos o ambientes en que se prueba la interfaz.
- **Usuario final:** Tipo de usuario que será el que finalmente haga uso de la aplicación. Es el usuario para el cual la aplicación ha sido diseñada.
- **Paper prototyping:** Método de usabilidad que consiste en el uso de un prototipo de la aplicación, en papel, con el cual el usuario evaluador lleva a cabo ciertas tareas, simulando el uso real de la aplicación.
- **Evaluación Heurística:** Este método de usabilidad está enfocado en los usuarios expertos. Se les pide que inicialmente prueben la aplicación y posteriormente llenan una pauta de evaluación basada en los principios de usabilidad. Este método se puede usar antes, durante y después del desarrollo de la aplicación.
- **Cuestionario:** Serie de preguntas que son respondidas por el usuario evaluador después de haber probado la aplicación. En este cuestionario se evalúa la satisfacción del usuario al usar la aplicación.
- **Observación:** Procedimiento de evaluación de usabilidad, en que un usuario realiza ciertas tareas con la aplicación a evaluar, y el observador o evaluador se encarga de tomar nota de las acciones y resultados que genera el usuario.

### 5.2.1. La aplicación

#### Descripción general de la aplicación

El propósito de esta aplicación móvil es la visualización de los estados de una persona monitoreada, y también permitir realizar algunas configuraciones al sistema de forma remota. Este sistema de visualización y configuración consta de distintas secciones principales, las cuales se detallan a continuación:

- **Usuarios monitoreados:** Una lista de las personas que están siendo monitoreadas por el usuario de la aplicación. Incluye nombre, foto y último estado detectado. Permite también agregar o eliminar personas de la lista.
- **Historial de estados:** Para cada usuario monitoreado se puede revisar su historial de estados detectados, con su respectiva hora, medio de detección (Skype o Grabación) y fecha de detección.
- **Configuración de monitoreo:** Por cada usuario se puede activar o desactivar el modo de grabación, ya sea Skype o Grabación (usando el micrófono). El modo de Grabación permite configurar los días de la semana y horarios de grabación.
- **Configuración de envío de alertas:** Permite configurar los medios de envío de alerta, como son el SMS o correo electrónico.

Es importante mencionar que la aplicación desarrollada solo llegó hasta la etapa de prototipo funcional inicial, es decir, se puede navegar e interactuar con sus botones de acción y gestos, pero toda la información que muestra sobre los estados no es real, y cualquier cambio que se haga en la configuración no se mantendrá.

#### Usuarios Finales de la aplicación

Los usuarios finales de la aplicación son aquellos interesados en el cuidado de la salud emocional de las personas monitoreadas. Por el lado afectivo, estos usuarios son principalmente familiares de la persona, como hijos, nietos, sobrinos, hermanos, etc., y también los amigos. Por el lado técnico, podrían ser los cuidadores que hayan sido contratados para atender a las personas que están siendo monitoreadas.

Es importante mencionar que estos usuarios finales deben tener alguna experiencia en el uso de aplicaciones móviles, y ojalá en el uso de Android.

#### Contexto de Uso

Al ser una aplicación móvil, es posible acceder a ella en cualquier momento que se tenga el teléfono a mano. La dificultad podría estar en lo que a conexión a Internet se refiere. Para obtener la última información y para que los cambios realizados en la configuración surtan efecto, será

necesario que el teléfono esté conectado a Internet, ya sea mediante Wi-Fi o 3G/4G. Pero eso hoy en día es más fácil y común. Casi todas las personas mantienen sus teléfonos inteligentes conectados a Internet al menos unas horas al día.

En general, los contextos de uso serán cuando la persona se encuentre revisando su teléfono móvil, o también cuando reciba una notificación de la aplicación. Esto podría ser en el trabajo, en la casa, en el autobús, o en la calle.

### ***5.2.2. Propósito de la evaluación de Usabilidad***

El propósito principal de la evaluación de Usabilidad fue diseñar, implementar y evaluar la aplicación móvil para el sistema de monitorización de estados anímicos, con el fin de obtener una aplicación acorde a los principios básicos de Usabilidad, además de cumplir con un diseño apto para cualquiera de los usuarios finales, y que refleje el propósito final del trabajo de Memoria.

Para cumplir totalmente con el propósito principal, fue necesario cumplir con algunos objetivos específicos:

- Diseñar una interfaz para la aplicación móvil:
  - Simple
  - Fácil de interactuar
  - Con conceptos claros
  - Acorde a modelo mental actual de aplicaciones móviles
  - Que cumpla con las Guías de Diseño de Android
- Evaluar la interfaz diseñada, para corroborar su:
  - Entendimiento
  - Navegación rápida y simple
  - Facilidad de aprendizaje
  - Que lleve a pocos o ningún error

Todo esto para que se adapte a cualquier tipo de usuario final, sin importar su edad y nivel de experiencia en uso de aplicaciones móviles, específicamente en Android.

### **5.2.3. Metodología de la evaluación de Usabilidad**

Los atributos de usabilidad son la base para definir un buen estudio. Estos atributos permiten dar una medida de cuán usable es un programa o página web. Son cinco los atributos de usabilidad: Aprendizaje, Recuerdo, Eficiencia, Errores y Satisfacción.

#### **Justificación y Fundamentación Teórica**

A continuación se detallan qué atributos de usabilidad se midieron, por qué se midieron, y cómo se midieron (Métodos de usabilidad).

#### Atributos de usabilidad medidos

Para cumplir con los objetivos mencionados, se midieron cuatro de los cinco atributos de usabilidad. Se dejó de lado la eficiencia, ya que por ser un prototipo inicial de la aplicación, no se enfrenta a reales desafíos, como la descarga o envío de datos, que podría hacer más o menos eficiente el actual funcionamiento. Entonces, los atributos medidos fueron:

- **Aprendizaje:** Se midió el aprendizaje de las distintas acciones de configuración y opciones dentro de la aplicación, por ejemplo cómo activar y modificar el número de teléfono para las alertas por SMS.
- **Recuerdo:** Después de haber visitado las secciones por primera vez, se le pidió al usuario volver a las mismas secciones, para verificar que recordaba dónde se encontraba y dónde podía encontrar las opciones que debía buscar.
- **Errores:** Se tomó constancia de los distintos errores de navegación, como al intentar volver a otra sección, o al buscar una fecha en el historial de estados o en el calendario, o también al modificar una opción.
- **Satisfacción:** Al ser una aplicación para distintos rangos de edad y de experiencia en el uso de aplicaciones móviles, era importante medir cuán cómodos se habían sentido los usuarios al probar la aplicación. Es importante que la aplicación se adapte al usuario y no al revés.

#### Métodos de usabilidad que fueron usados

Con la intención de poder medir los atributos de usabilidad seleccionados, se eligieron cuatro métodos de usabilidad:

- **Paper Prototyping:** Permitted validar preliminarmente la interfaz diseñada, corroborando si cumplía con todas las necesidades de los usuarios, además de evaluar inicialmente los atributos Aprendizaje y Recuerdo.

- Evaluación Heurística: Método utilizado para analizar la usabilidad de la interfaz diseñada, y encontrar los defectos que pudiera tener, en base a las diez heurísticas de Usabilidad de Nielsen [25].
- Observación: Sirvió para detectar los posibles errores en el uso de la aplicación. Además, permitió medir el aprendizaje y recuerdo, mediante las tareas que le fueron asignadas al usuario.
- Cuestionario de Usuario Final: Midió la satisfacción de los usuarios sobre la interfaz de la aplicación al momento de usarla.

## Muestra

Los usuarios finales se dividieron en dos grupos según su rango de edad: 20-35 años y 40-60 años. La razón es porque estos dos rangos presentan diferencias en la experiencia en el uso de aplicaciones móviles, especialmente el último rango, en el cual la cantidad de usuarios de teléfonos inteligentes es menor y tienen menor experiencia usándolos.

En Paper Prototyping participaron 8 usuarios finales, 4 del primer rango de edad, y 4 del segundo rango. En ambos grupos hubo hombres y mujeres, con diferentes niveles de experiencia. La Evaluación Heurística fue realizada por dos usuarios expertos, ex-alumnos del curso Taller de Usabilidad de Interfaces de Software.

Por último, en la Observación y Cuestionario de Usuario Final participaron en ambos los mismos 20 usuarios finales, 10 de cada rango de edad, y con diferentes niveles de experiencia. El promedio de edad se puede ver en la siguiente tabla:

*Tabla 1: Media de edad para cada rango de usuarios finales.*

Rango	Edad promedio
20-35 años	24,9
40-60 años	50,3

## Instrumentos

Un instrumento de usabilidad es un documento que permite llevar a cabo una evaluación de usabilidad, donde se definen tareas y/o preguntas. Estos instrumentos son los que se se usan en los distintos métodos de usabilidad. Los instrumentos usados en este proyecto fueron:

- Pauta de tareas y observaciones para Paper Prototyping:
  - Constó de 5 tareas a realizar, tareas que los usuarios finales desarrollaron utilizando fotos impresas de la interfaz diseñada, una foto de cada pantalla.

- Las pautas de las tareas fueron creadas en base a la plantilla de tareas de usabilidad encontrada en el libro "*Paper Prototyping*" [26].
- Pauta de Evaluación Heurística:
  - Contó con 10 secciones de evaluación, con tres o cuatro ítem en cada sección, para evaluar la aplicación. Cada sección evaluó en una de las heurísticas de Usabilidad de Nielsen [25]. Se usó una escala Likert de 5 valores, desde *Muy de acuerdo* hasta *Muy en desacuerdo*.
  - Las evaluaciones fueron enfocadas en el uso de la aplicación y sus características de aplicación móvil, como son los gestos y la rotación de la pantalla.
  - Se basó en la pauta de evaluación heurística "*Evaluación de Usabilidad de Sitios Web*": *Método de Evaluación Heurística*, creada por el profesor Jaime Sánchez.
- Pauta de Observación:
  - Inicialmente constaba de 6 tareas, pero después de los cambios hechos posteriores a la Evaluación Heurística, se agregó una tarea más para evaluar una nueva funcionalidad de la aplicación, *Revisar el historial de algún día que tenga datos, en el calendario*, dado que se había agregado un calendario a la aplicación.
  - Basada en la "*Pauta de Observación Usuario Final*" creada por Fernanda Ramírez, ex-alumna del curso.
- Cuestionario de Usuario Final:
  - Incluyó 12 preguntas cerradas sobre la interfaz, colores, iconografía, imágenes, lenguaje utilizado en la aplicación, para conocer la satisfacción y comodidad al ser usada por los usuarios. Se usó la misma escala Likert usada en la Pauta de Evaluación Heurística.
  - También incluyó 3 preguntas abiertas.
  - Se basó en otros Cuestionarios de Usuario Final presentados en el curso: "*Evaluación de Usabilidad de Sitios Web*" y "*Pauta de evaluación de usabilidad de videojuegos*", ambas del profesor Jaime Sánchez.

## **Tareas**

En todas las evaluaciones, la primera etapa fue contarles a los usuarios de qué trataba la aplicación, su utilidad, su misión, y en general qué era posible realizar con ella. La etapa siguiente fue la aplicación de los métodos de usabilidad, la que se detalla inmediatamente para cada uno de ellos:

## Paper Prototyping

Inicialmente, en la evaluación de Paper Prototyping, se les pidió a los usuarios finales que miraran cada una de las interfaces, para tener una idea de lo que se encontraba en ellas. Luego, para cada una de las tareas definidas en la pauta de Paper Prototyping, los usuarios debían ordenar las imágenes impresas de las interfaces en el orden que ellos creían correcto para llevar a cabo la tarea asignada, solo seleccionando las necesarias. Además, debían simular el ingreso de datos o manejo de los elementos de acción, como presionar un botón. Después de cada tarea realizada se le hacían preguntas con respecto al diseño de la interfaz.

Listado de las tareas realizadas para el proceso de Paper Prototyping:

1. Navegar hasta el historial del Tío Pepe.
2. Configure la opción de grabación de monitoreo.
3. Ingresar una nueva dirección de correo para alerta por correo.
4. Modificar frecuencia de repetición de notificación de EmoStatus.
5. Comparar opciones de diseño para usuario monitoreado (aquí se comparan dos interfaces el usuario elige la que crea más apropiada).

## Evaluación Heurística

Se les pidió a los usuarios expertos que navegaran en la aplicación antes de responder la Pauta de Evaluación Heurística. Esto se hizo para que conocieran en más detalle la aplicación, antes de responder, y no se quedaran solo con la explicación inicial. Posteriormente procedieron a responder la pauta correspondiente.

## Observación y Cuestionario de Usuario Final

Después de la explicación inicial, se procedió inmediatamente a evaluar a los usuarios en cada una de las tareas que se les fue asignada. Después de haber terminado las tareas, se les solicitó navegar una última vez por la aplicación, para conocerla en más detalle y revisar aquellas secciones que no fueron visitadas durante las tareas. En este proceso no se les sugirió como hacerlo, quedó a criterio de los usuarios. Finalmente procedieron a responder el Cuestionario de Usuario Final.

Listado de las tareas realizadas para el proceso de Observación:

1. Leer historial de estados de hoy del Tío Pepe.
2. Revisar si el Tío Pepe estuvo triste la semana pasada.
3. Desactivar monitoreo por Skype para la Abuela Lorena.
4. Activar opción Grabación en monitoreo para la Abuela Lorena, y cambiar los días de la semana seleccionados.

5. Cambiar el número de teléfono para alertar por SMS, para el usuario Tío Pepe.
6. Activar las notificaciones de la aplicación como alerta.
7. Revisar el historial de algún día que tenga datos, en el calendario.

## Equipos

Al ser una aplicación móvil la que fue diseñada y evaluada, fue fundamental contar con un teléfono móvil *touch* para las pruebas realizadas. El teléfono fue necesario para la Evaluación Heurística, para el proceso de Observación y posterior respuesta del Cuestionario de Usuario Final. El teléfono utilizado fue un LG Optimus L5.

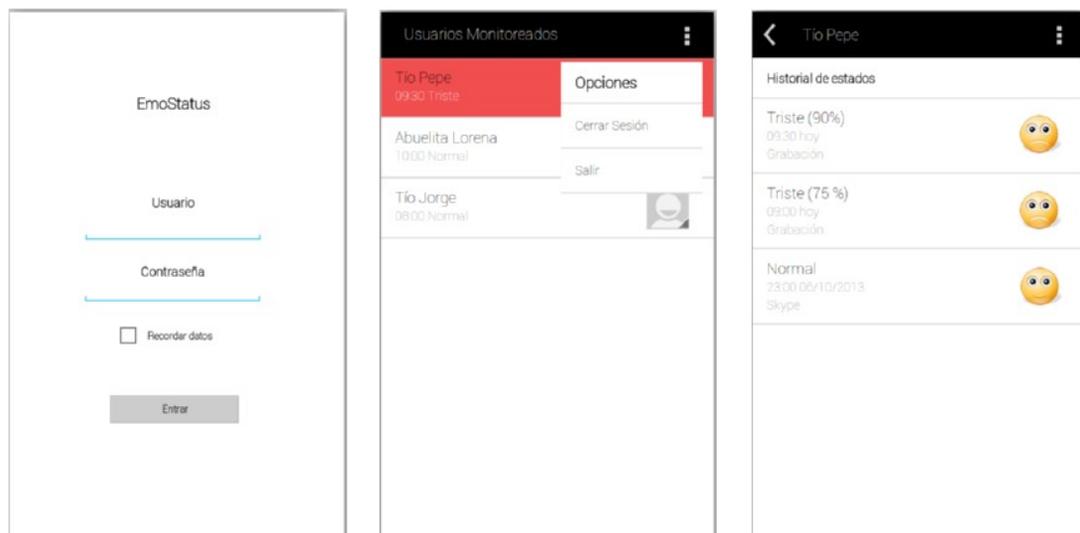
## Procedimiento

El procedimiento llevado a cabo en el proyecto consistió de distintas etapas de diseño, implementación y evaluación de usabilidad, las que se detallan a continuación, en el orden en que sucedieron:

### 1. Diseño de Interfaces

El diseño de las interfaces se realizó usando la aplicación en línea Fluid UI<sup>4</sup>, siguiendo las Guías de Diseño de Android [27]. Fluid UI permite diseñar *mockups* para distintas plataformas, iOS, Android, y Windows Phone. Cuenta con diversos elementos gráficos que se pueden arrastrar.

A continuación algunas imágenes de las interfaces diseñadas en esta etapa preliminar:



*Ilustración 5: Ejemplos de interfaces diseñadas. Login, Opciones e Historial de estados.*

<sup>4</sup> <http://www.fluidui.com>

## 2. Paper Prototyping

Como se dijo anteriormente, en la evaluación de interfaces diseñadas mediante Paper Prototyping, cada usuario final tuvo que realizar distintas tareas, haciendo uso de los *mockups* impresos de la aplicación.



*Ilustración 6: Usuario final del rango 40-60 años, realizando una tarea de Paper Prototyping.*

## 3. Implementación del primer prototipo

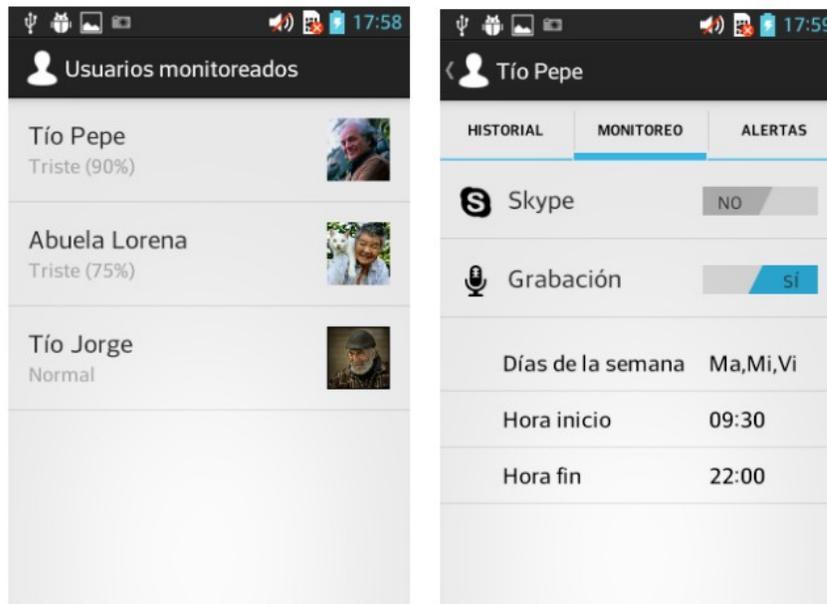
La implementación del primer prototipo de la aplicación se hizo en base a las interfaces diseñadas y a las correcciones hechas a partir de la evaluación con Paper Prototyping.

Se utilizó el programa Android Studio<sup>5</sup> para su implementación, un IDE desarrollado por Android, que incluye también el SDK. Para las pruebas, se usó el emulador de teléfonos móviles con Android, Genymotion<sup>6</sup>, y además la aplicación fue probada en el teléfono antes mencionado.

---

<sup>5</sup> <https://developer.android.com/sdk/installing/studio.html>

<sup>6</sup> <http://www.genymotion.com/>



*Ilustración 7: Ejemplos de pantallas del primer prototipo. Lista de usuarios monitoreados y Configuración de Monitoreo.*

#### 4. Evaluación Heurística

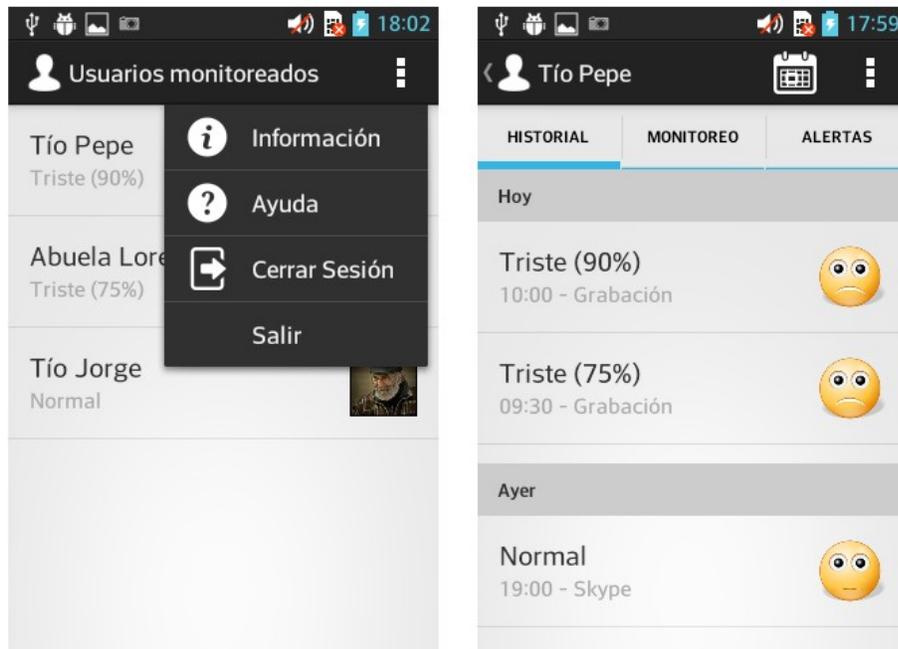
Dos usuarios expertos realizaron la Evaluación Heurística de la aplicación, haciendo uso de la aplicación ya instalada en el teléfono que se les proporcionó.



*Ilustración 8: Usuarios expertos probando la aplicación y respondiendo la Pauta de Evaluación Heurística.*

#### 5. Implementación del segundo prototipo

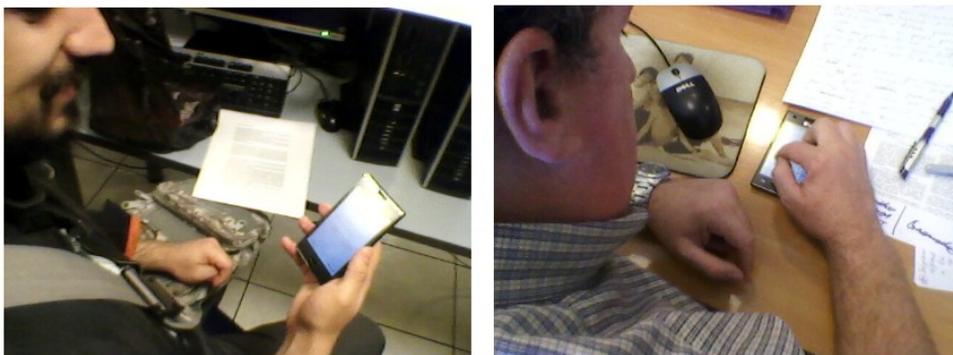
Se hicieron correcciones en la implementación de la aplicación, según la Evaluación Heurística, generando un segundo prototipo.



*Ilustración 9: Ejemplos de pantallas del segundo prototipo. Menú superior agregado e Icono de calendario en Historial de estados.*

## 6. Observación y Cuestionario de Usuario Final

La última etapa fue la realización de la Observación y Cuestionario de Usuario Final a los 20 usuarios finales, quienes llevaron a cabo primero algunas tareas y luego respondieron el cuestionario. Los usuarios hicieron uso del segundo prototipo de la aplicación, en el mismo teléfono ya mencionado.



*Ilustración 10: Usuarios finales realizando las tareas asignadas en la Pauta de Observación. El primer Usuario final es del rango 20-35 años y el segundo es del rango 40-60 años.*

#### 5.2.4. Detalles de la implementación de la aplicación móvil

Una aplicación Android consta básicamente de dos paquetes de archivos, *src* y *res*, donde se encuentran los archivos java que dan la funcionalidad a la aplicación y donde están los archivos asociados a la interfaz gráfica (maquetas xml, imágenes, colores, estilos, etc.), respectivamente. Para esta aplicación, los archivos implementados en el paquete *src* fueron separados en tres paquetes internos:

- *com.thesis.emostatus*: Contiene las *activities*, las cuales son los archivos que controlan y permiten la visualización de las interfaces gráficas. Un ejemplo de *activity* es el archivo *MainActiviy.java*, que se encarga de visualizar y manejar la pantalla de *Login*.
- *adapters*: En este paquete están las clases que permiten la construcción de distintas listas con información, recibiendo los datos desde las *activities* y entregando una lista ya completada. La lista de los usuarios monitoreados se construye gracias a un *adapter*.
- *persistance*: Este paquete almacena aquellas clases relacionadas con el modelo de datos, que puede ser la información del historial de estados, o el nombre del usuario monitoreado seleccionado, entre otros.

En la Ilustración 11 se pueden ver las principales clases implementadas en el paquete *com.thesis.emostatus*, con sus respectivos atributos y métodos.

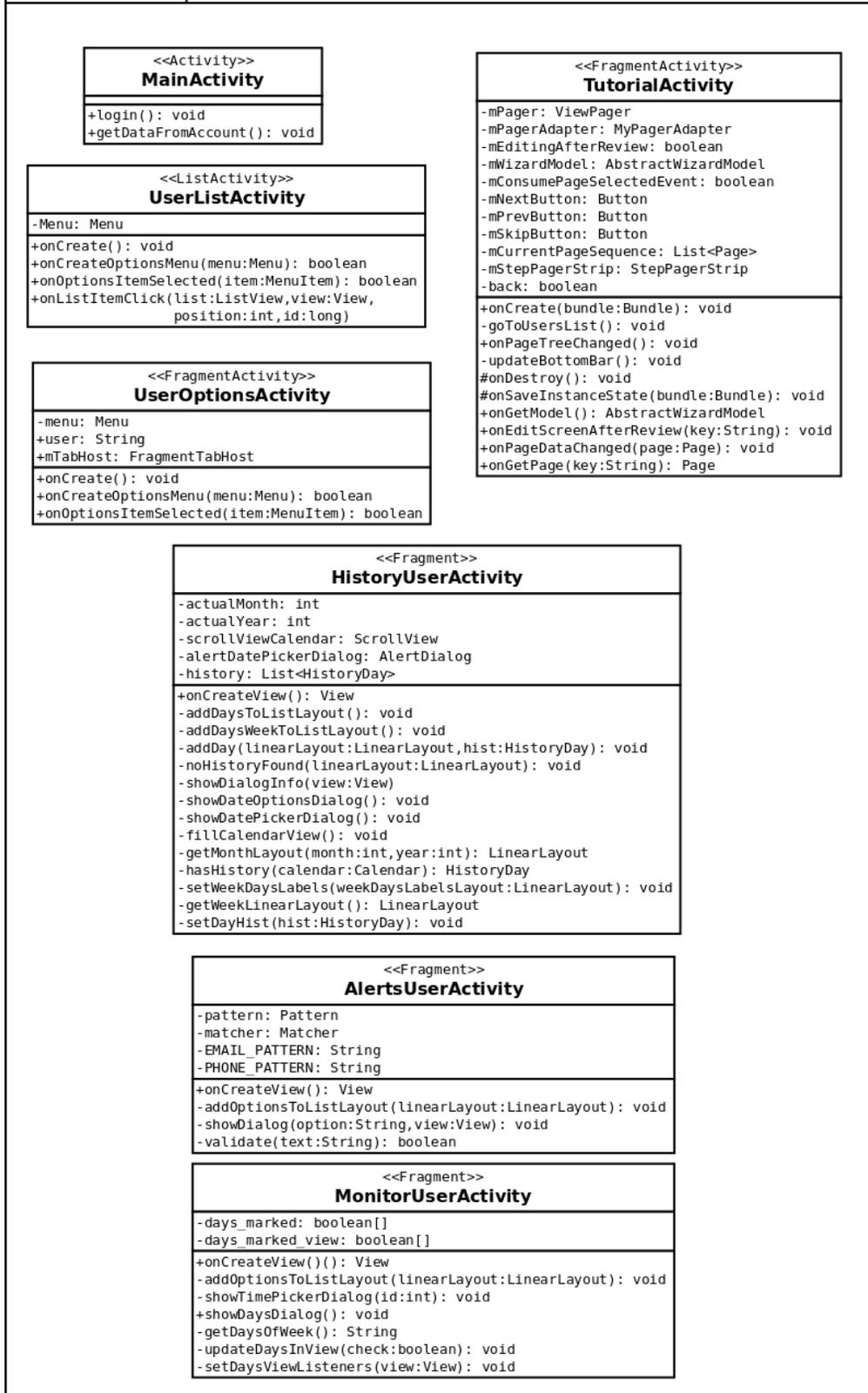


Ilustración 11: Diagrama de clases del paquete `como.thesis.emostatus` de la aplicación móvil.

### 5.3. Resumen

En este trabajo de memoria se implementaron dos partes del sistema. Una fue el conjunto de programas o *scripts* para el reconocimiento de emoción, género e identidad. La otra parte fue una aplicación móvil en Android, para la visualización de los datos, como aquellos datos que el sistema arroja después del reconocimiento, entre otros.

El conjunto de *scripts* desarrollados se enfocó en tres necesidades: procesamiento del audio (conversión de formato, normalización de volumen), análisis de datos (extracción de palabras, análisis de resultados) y, entrenamiento y testeo de muestras (generación de modelos, predicción, etc.). Los *scripts* fueron escritos en Bash y Python.

En el caso de la aplicación móvil, su construcción se basó principalmente en tres puntos: Diseño, Implementación de prototipos y Evaluación de Usabilidad. El diseño ayudó a plasmar las ideas iniciales con respecto a las distintas pantallas a utilizar, los elementos que éstas contendrían y la información que mostrarían. La implementación de los prototipos llevó a la aplicación móvil desde el diseño a la realidad, pudiendo probarla en un teléfono móvil. Por último, la Evaluación de Usabilidad consistió en la elaboración y aplicación de pautas de evaluación, enfocadas en encontrar problemas en la usabilidad tanto en el diseño como en los prototipos.

## 6. Evaluación del Sistema

En este capítulo se abordan dos tipos de evaluación. La primera es una evaluación del reconocimiento de emociones, género e identidad en grabaciones de conversaciones reales, para determinar la calidad de las soluciones encontradas a estas temáticas. La otra evaluación que se hace, es la relacionada con el estudio de Usabilidad y las herramientas creadas para este estudio, con respecto a la aplicación móvil construida.

### *6.1. Evaluación de reconocimiento de emociones, género e identidad en conversaciones*

Estas pruebas se hicieron para corroborar que las pruebas realizadas para reconocimiento entregaban buenos resultados en una situación más realista como lo es una conversación. Lo que aquí se presenta es un resumen de los resultados. En la sección A.5 del anexo se puede encontrar el detalle de estas pruebas.

#### **Muestras**

Se contaron con muestras de tres grabaciones. La primera entre dos hombres, la segunda entre un hombre y una mujer, y la tercera, igual que la segunda, pero ambos hablantes expresándose al mismo tiempo, con la intención de superponer las voces.

#### **Resultados**

En la primera y segunda conversación, los resultados fueron buenos, mejores en la primera, que en la segunda, logrando clasificar de forma correcta una gran parte de las muestras según emoción (Neutral), para luego clasificar según género. Por último vino el reconocimiento de identidades, y este fue el que falló, tal como se dijo en la sección anterior, generando una gran cantidad de falsos positivos.

Para la tercera conversación no se alcanzaron a hacer muchas pruebas. La idea era ver si era posible detectar muestras de audio no claras, en que hay más de una voz. En la segunda conversación se notó que algunas muestras eran con solapamiento, y éstas habían caído en otros grupos, y no donde estaba Neutral. Se pensó que podría ser como un filtro para evitar que caigan en Neutral, pero solo serviría para esa emoción. En la tercera conversación fue similar solo en algunos casos. Entonces se descartó totalmente esa opción, y se requerirá más pruebas y estudios de estos a futuro.

## 6.2. Evaluación de Usabilidad de la Aplicación Móvil

En esta sección se presentan los resultados y sus respectivos análisis, después de haber aplicado cada método de usabilidad elegido. Los resultados se basan en el flujo de navegación diseñado (ver Ilustración 4 en sección 4.4) y el que siguieron los usuarios.

### 6.2.1. Paper Prototyping

A continuación se presentan los resultados obtenidos y su análisis, para cada una de las tareas realizadas en Paper Prototyping:

#### 1. Navegar hasta el historial del Tío Pepe:

Hubo importantes diferencias en la percepción, "uso" y aprobación de la interfaz. Los mayores problemas de interpretación de la interfaz se vieron en el rango 40-60 años, especialmente en el flujo de navegación. Por otro lado, hay algo que no se explicó y se dejó a criterio de los usuarios su interpretación. Esto fue el porcentaje de exactitud del sistema al detectar si la emoción reconocida es Tristeza. La mayoría de las personas creyeron que indicaba el nivel de Tristeza reconocido, algo muy diferente a lo propuesto. A continuación las principales críticas al diseño:

- Los rangos de tristeza no son entendibles, son confusos, o simplemente no se sabe qué significan esos porcentajes.
- Los iconos utilizados para expresar emoción de Tristeza y Normal (Neutral), son muy iguales, la carita Normal podría tener la boca más recta, y el icono de Tristeza podría ser un poco más triste. Además para cada "nivel" de Tristeza, debería haber un rostro más o menos triste.

Estas críticas se debieron principalmente al no entendimiento de lo que el porcentaje indicaba, y lo segundo, los iconos de emociones, por un lado debido a la baja calidad de las impresiones, no permitiendo ver con claridad de las diferencias, y por otro lado a causa de creer que se representan diferentes niveles de Tristeza.

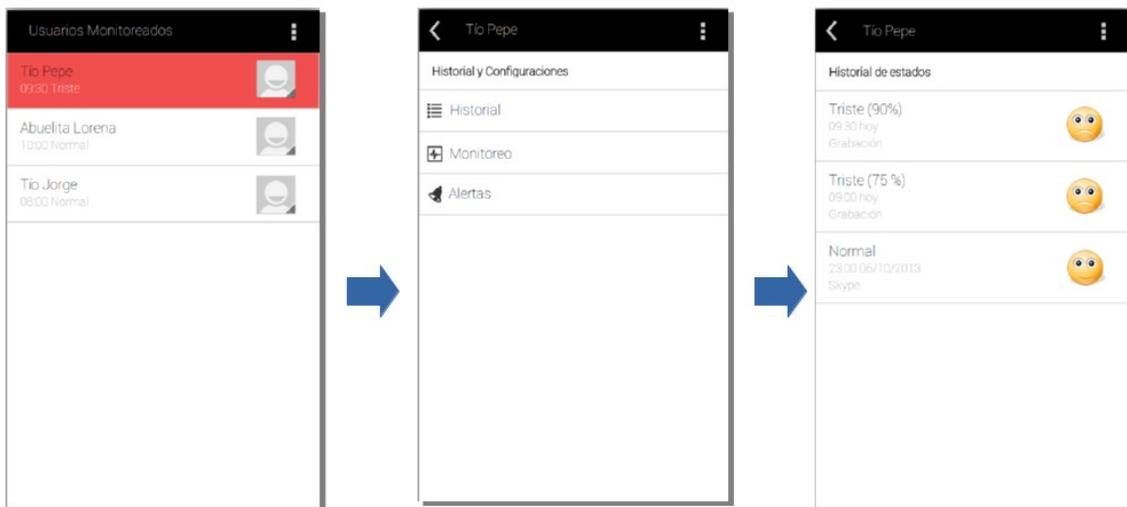
Otro problema que se observó, especialmente con usuarios finales del segundo rango de edad, fue el flujo de navegación:

- Uno de los flujos incorrectos que más se repitió fue: Historial y Configuraciones (seleccionó Historial) → Tío Pepe (en Usuarios Monitoreados) → Historial de Tío Pepe (ver más abajo Ilustración 12 e Ilustración 13).
- Otro flujo fue: Tío Pepe (en Usuarios Monitoreados) → Historial de Tío Pepe, sin pasar por el menú de opciones (Historial y Configuraciones). Este se debió a que el usuario no creyó que

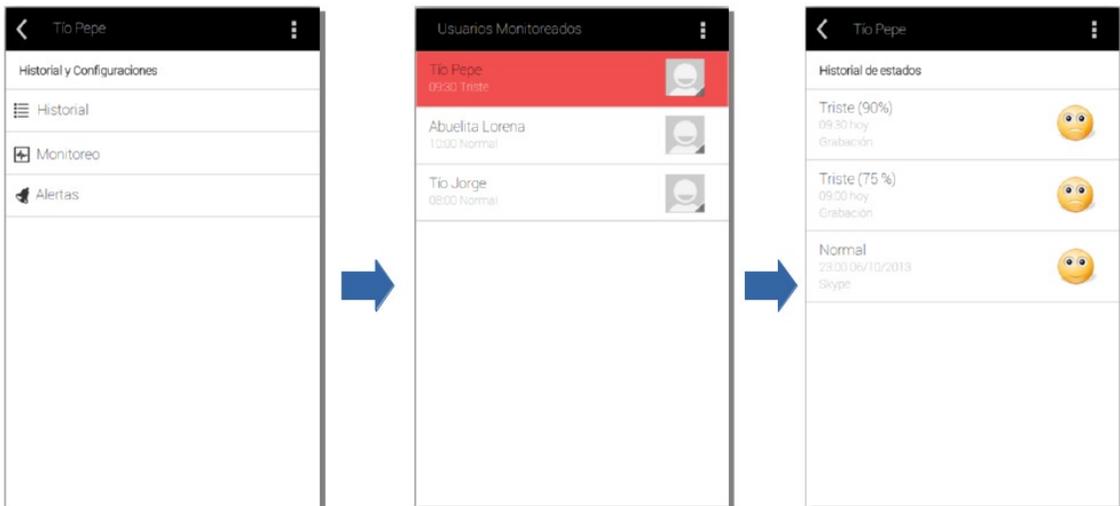
ese menú de opciones correspondiera a un menú de un usuario, si no que parecía algo general de la aplicación.

- Un flujo interesante fue que un usuario, para realizar esta tarea, pasó por casi todas las pantallas de configuración, como si tuviera que hacer una configuración inicial de la aplicación, antes de poder visualizar los datos.

Y por último, hubo dudas y confusiones sobre la información entregada relativo a la fecha, como ¿dónde queda la información de los otros días que no son hoy?



*Ilustración 12: Flujo de navegación correcto para revisar el historial.*



*Ilustración 13: Flujo de navegación incorrecto para revisar el historial.*

## 2. Configure la opción de grabación de monitoreo:

En esta tarea se presentaron menos problemas, después de haber entendido mejor cómo era la navegación entre las distintas interfaces. Aun así se repitió el incorrecto flujo de navegación, accediendo primero al menú y luego al usuario:

- Flujo incorrecto: Historial y Configuraciones (seleccionó Monitoreo) → Tío Pepe → Monitoreo.

El otro punto que no quedó claro, fue que las opciones de Grabación en Monitoreo estaban inhabilitadas, porque este modo estaba desactivado. Esto tendría dos causas, la primera es la no costumbre de notar esta diferencia entre habilitado e inhabilitado, y segundo, que la calidad de la imagen era poco clara, notándose muy poco las letras, pareciendo más mala calidad que elementos inhabilitados.



*Ilustración 14: Mockup de pantalla de Configuración de Monitoreo.*

En general, a pesar de los puntos mencionados, la mayoría de los usuarios finales realizó correctamente la tarea, y encontró apropiada la manera de escoger los días de la semana, además de comentar que les gustaba la forma de seleccionar la Hora Inicio y Hora Fin.

### 3. Ingresar una nueva dirección de correo para alerta por correo:

Todos los usuarios finales lograron “modificar” correctamente en la dirección de correo, y la mayoría de ellos siguió el flujo correcto (Ilustración 15), excepto la misma persona que siguió un flujo no correcto en las tareas anteriores. El flujo que siguió fue:

- Historial y Configuraciones (seleccionó Alertas) → Tío Pepe → Alertas (seleccionó Correo electrónico).

Cabe mencionar, que algunos creyeron que se podía modificar en la misma pantalla de Alertas, pero después se daban cuenta que estaba la otra pantalla para modificar el correo.

Una importante sugerencia fue que podría ser solo una pantalla en vez de dos, haciendo la tarea mucho más sencilla.

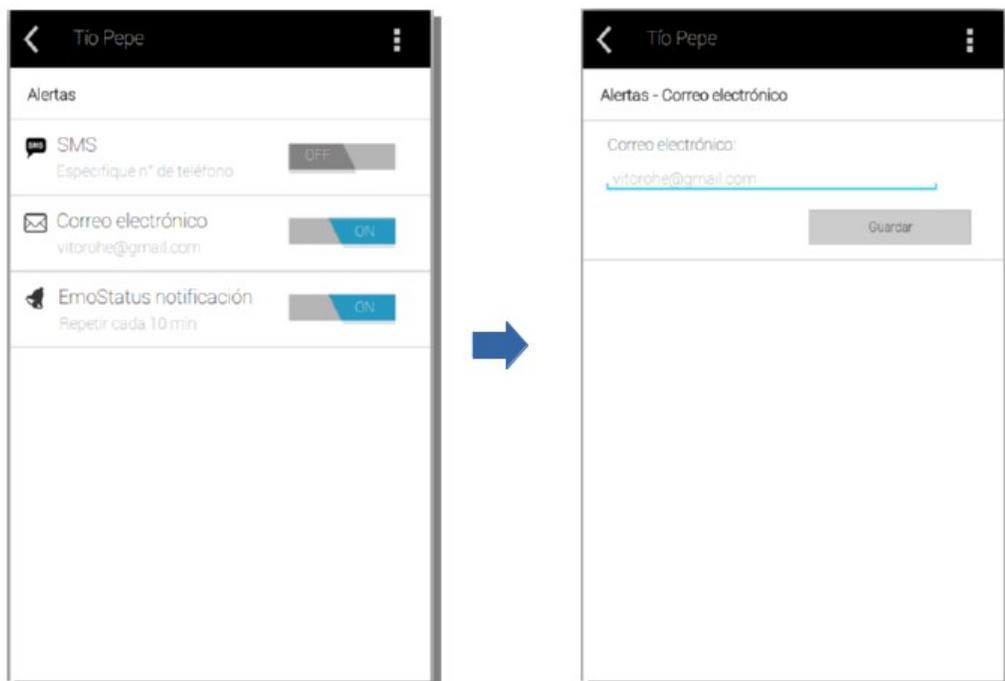


Ilustración 15: Flujo pensado para la modificación del correo electrónico.

### 4. Modificar frecuencia de repetición de notificación de EmoStatus:

Esta tarea también fue realizada sin mayores dificultades, pero hubo diferencias a la forma de cómo hacerlo. En el diseño se puso un menú *dropdown*, que al presionarlo debería mostrar las otras opciones hacia abajo. La mayoría creyó que era un *textfield*, es decir, un campo de edición de texto. Cuando se les contó lo que realmente era y se les preguntó qué pensaban de esa manera de elegir la frecuencia, casi todos prefirieron que fuera un campo de texto editable. Solo una

persona dijo que prefería usar *radio buttons*, con una lista de opciones predefinidas.

## 5. Comparar opciones de diseño para usuario monitoreado:

Esta tarea consistió básicamente en elegir de dos diseños de interfaces, la que le pareciera más cómoda y útil. Una de ellas es la ya vista interfaz de Historial y Configuraciones, y la otra es usando pestañas o *tabs*, mostrando inmediatamente todas las opciones disponibles. Todos, menos uno, prefirieron la segunda opción, por ser más directa, más fácil, y evita tener que ir atrás. El usuario al que no le gustó el uso de tabs, lo dijo por lo siguiente:

- Una lista es más fácil de interactuar, no siempre funciona el cambiar de tabs con gestos (*swipe*), y prefiere los submenús más que tener varias ventanas.

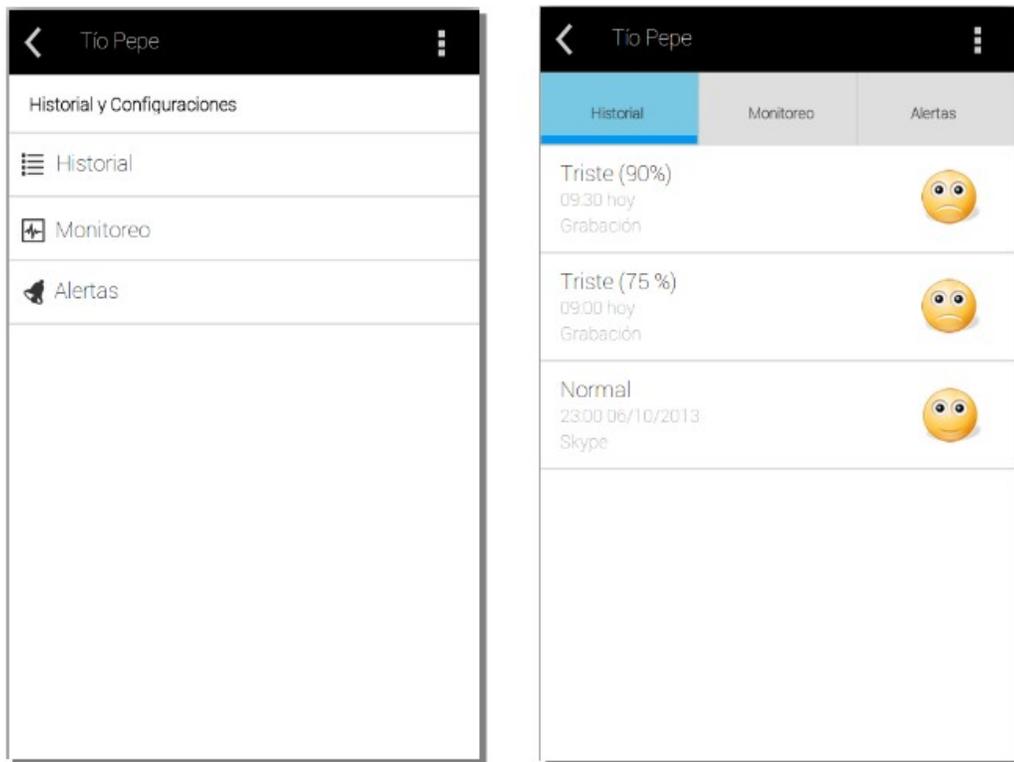


Ilustración 16: Comparación de interfaces para elegir opciones para un usuario monitoreado.

De los resultados obtenidos al llevar a cabo el Paper Prototyping, se tomaron distintas decisiones con respecto a los cambios que se harían en las interfaces antes de su implementación. Los cambios y decisiones tomadas fueron:

- Se dejó el uso de tabs para acceder a las opciones principales disponibles para cada usuario monitoreado.

- Permanecieron los iconos de caritas para describir las emociones, aunque se editó un poco la carita de Normal, dejándole una boca más pequeña. Se mantuvieron, porque se esperaba que al verlos directamente en la aplicación en el teléfono, se verían mejor y no causarían confusión.
- Además de explicar qué significan los porcentajes directamente a los usuarios finales, es decir, diciéndoselos cara a cara, también se ideó una manera de mostrarlo dentro de la aplicación, mediante diálogos.
- Se quitaron de la aplicación la frecuencia de monitoreo por Grabación y la frecuencia de las notificaciones de la aplicación. Esto se hizo porque se creyó que era mejor dejar esta decisión a la aplicación, y no al usuario final, para simplificar su uso.
- Se modificó la manera de editar el número de teléfono para alertar por SMS y el correcto electrónico. En vez de llevar a otra pantalla, se abre un diálogo con un campo de texto editable.
- Para el historial, se agregaron más días, siguiendo un orden vertical, más arriba los días más recientes, y más abajo los más antiguos. Usando *scroll* se puede navegar en el historial.

### **6.2.2. Evaluación Heurística**

En esta etapa se evaluó la aplicación en su primera implementación. En general, en las distintas secciones de la pauta, la aplicación fue bien evaluada, con valoración *Muy de acuerdo* o *De acuerdo*, y uno que otro Neutral. Las secciones peor evaluadas fueron las correspondientes a las heurísticas Prevención de Errores y Ayuda y Documentación, recibiendo principalmente valoraciones *En desacuerdo*.

Las falencias en Prevención de errores se debieron específicamente a que la aplicación no validaba los datos ingresados al modificar el número de teléfono y al cambiar el correo electrónico. Por su parte, Ayuda y Documentación recibieron mala valoración por la simple razón de que no se contaba con algún menú de ayuda o tutorial, o alguna sección de informaciones.

Otra sección en que hubo una mala valoración fue en Flexibilidad y eficiencia de uso, aunque no estaba cien por ciento funcional la aplicación, ya era posible notar algunas dificultades. La dificultad se encontraba en buscar una fecha específica en el historial. Si eran muchos estados detectados, y muchas fechas, tomaría harto tiempo buscando una fecha solo haciendo *scroll*. Para resolver este problema uno de los evaluadores sugirió el permitir el acceso al historial mediante un calendario, para visualizar más rápido fechas y estados.

De las sugerencias y problemas encontrados, los cambios realizados en la interfaz y su siguiente implementación fueron:

- **Ayuda y Documentación:**
  - Se agregó una guía rápida al inicio de la aplicación, describiendo en pocas palabras para qué sirve ésta.
  - Aunque el menú de opciones extras estaba en las interfaces diseñadas, para esta primera implementación no se alcanzó a desarrollar. Para el segundo prototipo, se incluyó este menú, contando con dos opciones, una de Información, y otra de Ayuda. En Información hay detalles de cada sección de la aplicación, además de dar la opción de volver a ver la guía rápida. En Ayuda se puede enviar un correo a los desarrolladores de la aplicación.
- **Prevención de errores:**
  - Los campos de número de teléfono y correo electrónico son validados en base a una expresión regular, y muestra un mensaje de error cuando se ingresan incorrectamente los datos.
- **Flexibilidad y eficiencia de uso:**
  - Se incluyó un icono de Calendario en la barra principal, que solo se ve estando en el historial. Este icono abre un diálogo permitiendo filtrar el historial, pudiendo ver todos los días, solo la semana actual, o un día específico en el calendario. Este último lleva a un calendario, donde los días con datos salen marcados, y se puede navegar entre un mes y otro (Ilustración 17).

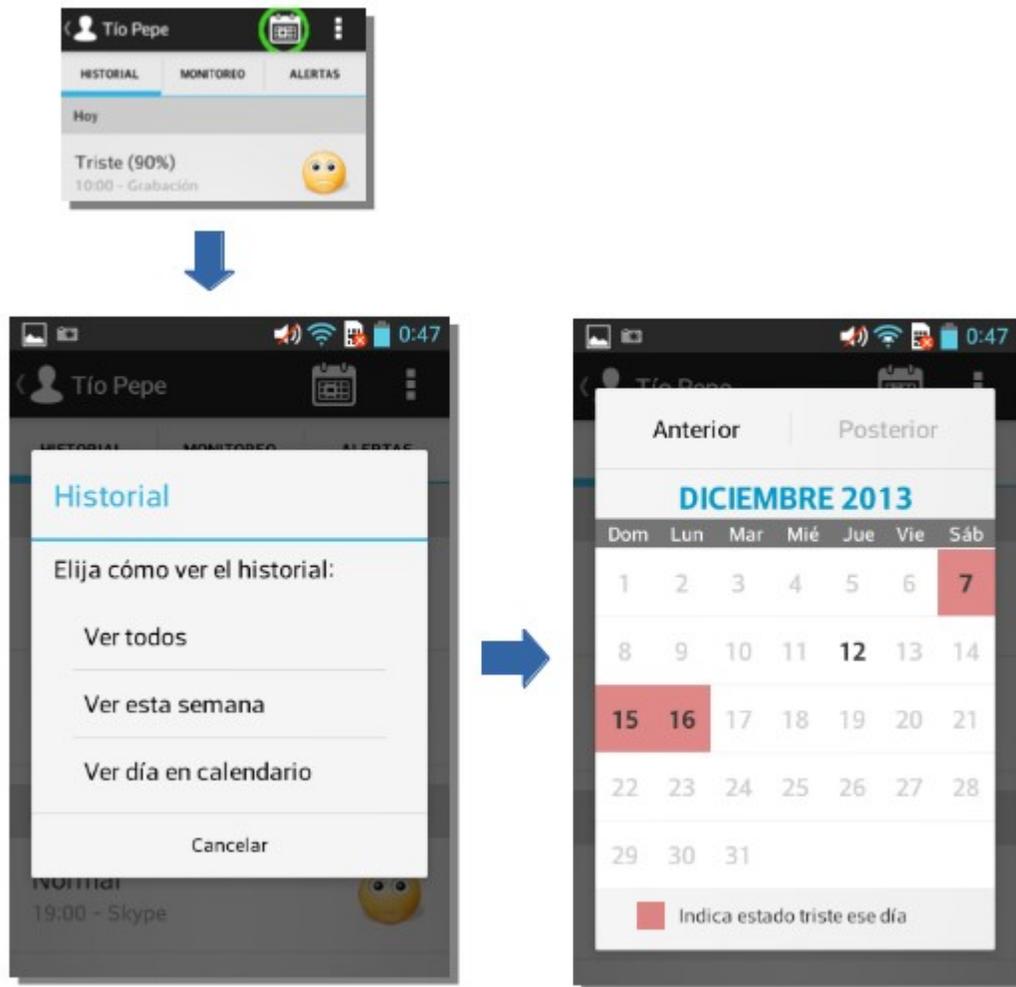
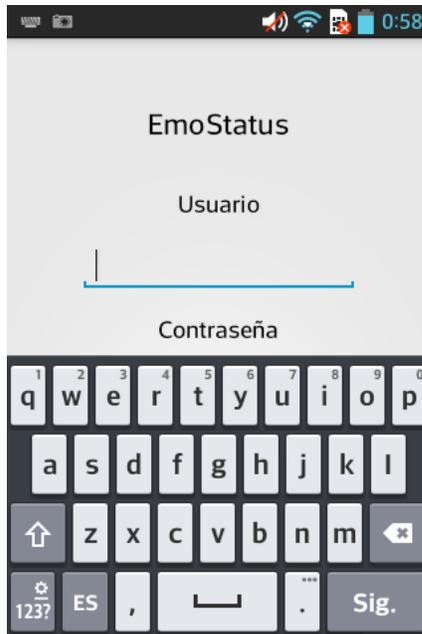


Ilustración 17: Nueva funcionalidad de Calendario.

### 6.2.3. Observación

Antes de presentar los resultados de cada tarea, hay un punto que destacó bastante al iniciar el uso de la aplicación. La primera pantalla que se ve es la pantalla de *Login*. Muchos tuvieron problema para ingresar la Contraseña, especialmente en el segundo rango de edad. Esto se debió a que la mayoría quería hacer scroll o presionar directamente sobre el campo de Contraseña después de rellenar el campo de Usuario, pero la aplicación no lo dejaba, y continuaba mostrando el teclado. Esto fue así porque se dejó la opción por defecto que trae Android. Esta opción es que si se está llenando un campo de texto, y viene otro más abajo, en el teclado aparece el botón *Sig.*, el cual permite saltar al siguiente campo, pero no deja presionar lo que está atrás. Una forma de salirse del teclado era presionando atrás, pero nadie lo hizo (Ilustración 18).



*Ilustración 18: Pantalla de login mostrando la opción Sig. en el teclado.*

Otro detalle que se destacó, fue en la navegación dentro de la Guía Rápida, que aparece antes de poder revisar la aplicación. Esta guía rápida contiene varias pantallas, algunas necesitando hacer scroll para ver toda su información. El problema fue que varios usuarios finales, navegaron cambiando de pantalla, pero nunca hicieron scroll, perdiendo información que se detallaba más abajo.

Ahora los resultados y sus análisis para cada una de las tareas realizadas en el proceso de Observación.

### **1. Leer historial de estados de hoy del Tío Pepe:**

Esta tarea fue realizada con éxito y sin problemas por la mayoría de los usuarios, sin importar el rango de edad o experiencia en aplicaciones móviles. Aún hubo confusiones con respecto al porcentaje mostrado.

Hubo dos usuarios que presionaron sobre "Hoy", sin darse cuenta que la información estaba directamente abajo, pero al segundo después la notaron.

### **2. Revisar si el Tío Pepe estuvo triste la semana pasada:**

La primera pregunta que casi todos se hicieron fue ¿qué día es hoy?, para luego preguntarse ¿qué días fueron la semana pasada?. En búsqueda de la respuesta, solo dos personas buscaron alguna forma de

saber qué fecha era el día de hoy (para día de la prueba, claro). Casi todos se conformaron con buscar en el historial, haciendo scroll, y encontrar, ya sea una fecha real de la semana pasada, o una que cierta fecha que creyeron era de la semana pasada. Esto claramente muestra un problema a resolver, de cómo mostrar los datos de fechas y cómo acceder a cada una de ellas. En la Ilustración 17 se puede ver que en el historial dice "Hoy", pero no indica la fecha.

Esta tarea estaba inicialmente pensada para usar scroll en el historial, que es lo que 18 de las 20 personas hizo. Solo dos usuarios probaron el calendario y encontraron alguna fecha de la semana pasada.

### **3. Desactivar monitoreo por Skype para la Abuela Lorena:**

Los problemas que se presentaron no fueron específicamente relacionados con el propósito final de la tarea. Estos se relacionaron con la navegación a la que están o no acostumbrados. Para realizar esta tarea, era necesario ir atrás. Finalizada la tarea anterior, el usuario se encontraba en las opciones de Tío Pepe, pero necesitaba esta vez ir atrás, a la lista de Usuarios Monitoreados, para poder seleccionar a la Abuela Lorena.

Al momento de ir atrás, para volver a la lista de Usuario Monitoreados, dos principales problemas se encontraron mientras algunos usuarios realizaban la tarea:

- Algunos apretaban la barra principal, pero no funcionaba. Esto porque apretaban sobre el nombre del usuario monitoreado, y no sobre el símbolo < para ir atrás. Esto causó un poco de caos en algunos, haciéndolos apretar otras secciones o menús.
- En el teléfono utilizado, el botón atrás se encuentra abajo a la izquierda, en algunos teléfonos actuales, especialmente de marca Samsung, tienen este botón abajo a la derecha. Entonces, para algunos usuarios finales, el modelo mental era diferente, y no correspondía al presentado.

A pesar de estos problemas, al volver atrás, ya sea después de varios intentos o siendo asistidos (caso extremo), llevaron a cabo la tarea de forma rápida y correcta.

### **4. Activar opción Grabación en monitoreo para la Abuela Lorena, y cambiar los días de la semana seleccionados:**

Después de haber llevado a cabo la tarea anterior, esta fue fácil de realizar para todos los usuarios, aunque algunos tomaron más tiempo que otros, pero nadie cometió error alguno.

## **5. Cambiar el número de teléfono para alertar por SMS, para el usuario Tío Pepe:**

En la aplicación, la alerta por SMS se encontraba desactivada por defecto, entonces casi todos al querer modificar el número, no podían, se daban cuenta que la opción estaba inhabilitada y luego activaban la alerta por SMS, para finalmente cambiar el número de teléfono.

En la edición del número, estuvo dividida la forma en que lo hicieron. Algunos ingresaron un número cualquiera, y otros se preocuparon que fuera un número correcto. Para los que ingresaban un número de longitud mayor o menor a 8, les mostraba un mensaje de error, de que el número debía ser de longitud 8. El problema es que muy pocos lo vieron. Para los que no lo vieron y volvían a fallar, se les explicaba el error, y la próxima vez se preocupaban de ingresar un número correcto.

## **6. Activar las notificaciones de la aplicación como alerta:**

Esta tarea fue la más confusa de entender. Fueron pocos los que entendieron inmediatamente que se refería a la opción *EmoStatus notificación* en la sección Alertas. A algunos usuarios hubo que repetirle varias veces la tarea. Solo unos pocos no lograron realizarla. De los que lo lograron, algunos la interpretaron a su manera, y lograron terminarla, y el resto logró entender de qué se trataba.

## **7. Revisar el historial de algún día que tenga datos, en el calendario:**

Ésta es la tarea que tuvo una mayor cantidad de usuarios que no la realizó correctamente, según lo esperado. En la Ilustración 17 se muestra el flujo que debía seguirse en esta tarea. Algunos de los flujos posteriores a presionar el icono de calendario fueron:

- Ver todos.
- Ver semanas actuales
- Ver día en calendario

Los que eligieron este último, que es el esperado, pudieron acceder al calendario. De las acciones siguientes, se destacó que:

- La mayoría presionó un día en rojo, sin leer antes la simbología.
- Otros presionaron cualquier día hasta que resultaba con alguno.
- Algunos tuvieron problemas para cambiar de mes, no se daban cuenta que podían usar el botón Atrás y Adelante que estaban arriba.

En general, la aplicación está bien, para ser un segundo prototipo, pero requiere de algunos cambios. Lo destacable, es que sin importar la experiencia o edad que tuvieran los usuarios finales, la mayoría logró llevar a cabo las tareas, quizás con uno que otro pequeño problema, pero lo lograron.

Son muchas las cosas que se pueden deducir sobre qué se debería mejorar a futuro en el diseño e implementación de la aplicación. Las más destacadas son:

- Pantalla de login en que se pueda fácilmente cambiar de un campo a otro. Aunque esta pantalla se vea solo una vez, solo hay una oportunidad para dar una primera buena impresión.
- Es mejor dejar habilitados los campos de número de teléfono y correo electrónico, sin importar si la opción de envío de alerta está activada o no.
- Hay que mejorar el calendario, tanto en la información que entrega, como en la prevención de errores.
- La opción de EmoStatus notificación en Alertas podría sacarse. Si el usuario final descarga esta aplicación, significa que desea ser alertado en cuanto se detecte un estado vulnerable. Entonces, esta opción quedaría por defecto activada.

Se hicieron algunos estudios estadísticos para visualizar si había alguna diferencia entre los dos rangos de edad.

*Tabla 2: Medias de tiempo en cada tarea observada.*

<b>Tarea</b>	<b>Tiempo promedio (s)</b>
<b>1</b>	3,45
<b>2</b>	18,90
<b>3</b>	14,20
<b>4</b>	9,15
<b>5</b>	12,85
<b>6</b>	8,20
<b>7</b>	10,55

Tabla 3: Medias de tiempo en cada tarea observada, para cada rango de edad.

<b>Tarea</b>	<b>Tiempo promedio Rango 1 (s)</b>	<b>Tiempo promedio Rango 2 (s)</b>
<b>1</b>	3,70	3,20
<b>2</b>	12,80	25,00
<b>3</b>	5,30	23,10
<b>4</b>	6,30	12,00
<b>5</b>	11,20	14,50
<b>6</b>	3,50	12,90
<b>7</b>	7,40	13,70

Se realizó una Prueba T sobre muestras independientes para comparar las medias de tiempo de las tareas, con respecto a los rangos de edad. Solo en la tarea 4 se encontró que las medias 6,30 y 12,00, para rango 1 y rango 2 , respectivamente, son estadísticamente diferentes. El mayor tiempo tomado al realizar esta tarea por parte de usuarios del segundo rango de edad, podría deberse específicamente a que ellos realizan las tareas con mayor calma. Se deduce esto porque todos ellos realizaron la tarea correctamente, entonces la mayor tardanza no se debería a algún problema con el su de la aplicación.

Se calculó la media de tiempo por nivel de experiencia, observando que los usuarios con mayor experiencia en aplicaciones móviles tomaron menor tiempo en realizar las tareas. Lamentablemente no se alcanzó a verificar si estas medias realmente son estadísticamente diferentes.

Tabla 4: Medias de tiempo en cada tarea observada, para cada nivel de experiencia.

Tarea	Exp. Baja (s)	Exp. Media (s)	Exp. Alta (s)
1	3,33	3,58	3,20
2	20,00	20,25	15,00
3	40,00	10,75	7,00
4	14,00	9,25	6,00
5	11,67	13,33	12,40
6	7,67	10,00	4,20
7	15,67	10,17	8,40

#### 6.2.4. Cuestionario de usuario final

Casi todas las valoraciones entregadas en este cuestionario de satisfacción, fueron *Muy de acuerdo* y *De acuerdo*. Muy pocas fueron *Neutral*, y ninguna cayó en *En desacuerdo* o *Muy en desacuerdo*. La siguiente tabla muestra los promedios de cada uno de los ítem. Las valoraciones se reemplazaron por números del 1 al 5, siendo *Muy de acuerdo* igual a 5 y *Muy en desacuerdo* igual a 1:

Tabla 5: Media de valoración de los ítems en Cuestionario de Usuario Final.

<b>Tarea</b>	<b>Promedio</b>
<b>La aplicación es fácil de utilizar.</b>	4,60
<b>Es fácil de navegar entre una sección y otra.</b>	4,85
<b>Es fácil encontrar la información u opción de configuración deseada.</b>	4,35
<b>Los botones o elementos de acción funcionan correctamente.</b>	4,75
<b>El uso del color es apropiado.</b>	4,55
<b>El diseño general de la aplicación es apropiado.</b>	4,60
<b>Los iconos de la aplicación son claramente identificables.</b>	4,50
<b>Me gustan los iconos de la aplicación.</b>	4,40
<b>Los iconos de la aplicación me transmiten información.</b>	4,65
<b>La información entregada es clara.</b>	4,45
<b>El contenido que se entrega es relevante.</b>	4,60
<b>La interfaz de la aplicación es placentera.</b>	4,65

Tabla 6: Media de valoración de los ítem en Cuestionario de Usuario Final, con respecto a rango de edad.

Tarea	Promedio Rango 1	Promedio Rango 2
<b>La aplicación es fácil de utilizar.</b>	4,60	4,60
<b>Es fácil de navegar entre una sección y otra.</b>	4,90	4,80
<b>Es fácil encontrar la información u opción de configuración deseada.</b>	4,40	4,30
<b>Los botones o elementos de acción funcionan correctamente.</b>	4,70	4,80
<b>El uso del color es apropiado.</b>	4,70	4,40
<b>El diseño general de la aplicación es apropiado.</b>	4,70	4,50
<b>Los iconos de la aplicación son claramente identificables.</b>	4,60	4,40
<b>Me gustan los iconos de la aplicación.</b>	4,50	4,30
<b>Los iconos de la aplicación me transmiten información.</b>	4,80	4,50
<b>La información entregada es clara.</b>	4,30	4,60
<b>El contenido que se entrega es relevante.</b>	4,70	4,50
<b>La interfaz de la aplicación es placentera.</b>	4,80	4,50

Se realizó una Prueba T de comparación de muestras independientes, con respecto a cada rango de edad, y ésta resultó en que no hay diferencias estadísticamente significativas en las medias correspondientes a cada ítem de evaluación del Cuestionario de Usuario Final, con respecto al rango de edad.

De las medias vistas en la Tabla 5, es posible notar la alta puntuación con la que la aplicación fue evaluada. Esto se refleja también en las respuestas de las preguntas abiertas, que mostraron una buena evaluación de la aplicación. Aunque también hubo respuestas sobre qué se podría mejorar en la aplicación, puntos importantes a tomar en cuenta.

A continuación un análisis más detallado por cada pregunta, de lo que más se comentó, y las respuestas que merecen más atención:

- **¿Qué le gustó de la aplicación?**
  - Es fácil de entender, aplicar y usar.

- Es cómoda, limpia y amigable.
- Entrega información oportuna y suficiente.
- Es simple y rápida de navegar.
- El resumen inicial es muy descriptivo.
- **¿Qué no te gustó de la aplicación?**
  - El historial y el calendario son poco claros, en cuanto a fecha y los datos entregados.
  - Los colores son algo apagados.
- **¿Qué agregarías a la aplicación?**
  - Más claridad en las fechas del historial.
  - Cambio de colores en iconos.
  - Distintos colores en cada sección.
  - Cuadro de Biografía, donde salgan las razones del seguimiento de cada usuario monitoreado.
  - Mayor claridad sobre los porcentajes en los estados tristes.
- **Observaciones o Comentarios**
  - Útil, interesante y novedoso el concepto. La idea de medir los estados es original.
  - Es una aplicación innovadora.

## 7. Conclusiones y Trabajo a Futuro

En este trabajo se logró desarrollar parte del objetivo principal *Implementación de un sistema remoto de monitoreo del estado anímico de adultos mayores*. Esto se logró mediante el cumplimiento de los objetivos específicos.

Con respecto al primer objetivo específico, procesamiento de la voz y reconocimiento de las emociones, fue posible encontrar un tipo de clasificación que permitiera identificarlas de forma más certera, basándose en grupos de emociones. Esto fue de gran ayuda, ya que se encontró que algunas emociones se confundían con otras, haciendo la clasificación mucho más incierta y errada.

Lo que no se previó fue el reconocimiento de género e identidad, no incluidos inicialmente, y que se vio durante el trabajo lo importante que eran al momento de identificar a la persona monitoreada. La identificación de género e identidad en emociones requirió de un arduo análisis, logrando reconocer de forma correcta género en las muestras con emociones. No así la identidad de la persona, quedando como un punto débil dentro del proceso general.

En cuanto al segundo objetivo específico, de contar con un sistema de visualización de los datos mediante una aplicación móvil, éste se logró mediante el diseño e implementación de dos prototipos. En el desarrollo se implementaron distintas clases dotando a la aplicación de un modelo básico de datos, pensando en la posterior recepción de información desde el servidor y luego la visualización de ésta.

Además de la implementación de la aplicación móvil, se realizó un estudio de Usabilidad mostrando que ésta, aun habiendo llegado a una segunda fase de prototipo, fue muy bien evaluada por los distintos usuarios. Se lograron los principales objetivos de simplicidad y facilidad de uso, adaptándose a distintos usuarios, sin importar su edad o experiencia con aplicaciones móviles.

A pesar de que no se haya logrado llevar a cabo todo lo que se había propuesto inicialmente, se lograron grandes resultados, dejando el sistema de monitorización en un estado bastante avanzado, principalmente en su punto fundamental que es el proceso de reconocimiento. Por su parte, el

sistema de visualización también contó con importantes avances gracias al diseño e implementación parcial de la aplicación móvil, dando una idea clara de qué se necesita y cómo se debe transmitir la información que se quiere entregar.

Dado esto, queda como trabajo futuro principalmente mejorar lo que es el proceso de reconocimiento, en especial lo que es la detección de la identidad. Por otro lado está todo lo relacionado con la automatización del proceso de monitoreo, desde la grabación, pasando por el reconocimiento, hasta las alertas y la visualización de los datos. En los siguientes puntos se plantea cuáles deberían ser los objetivos que se deberían llevar a cabo en un trabajo a futuro:

### **Procesamiento, análisis y reconocimiento**

- Mejorar el procesamiento de voz, descartando muestras de audio no claras, por ejemplo cuando se solapan voces, o hay un ruido de fondo.
- Clasificar según género basándose en el cuadrante del *Circumflex of Affect* en que se encuentra la emoción detectada. Esto permitirá clasificar cualquier emoción de cualquier cuadrante según su género, sean las mismas que se definieron en las pruebas, o cualquier emoción extra que se agregue.
- Contar con una mayor cantidad de muestras por usuario para obtener una mejor clasificación según identidad. Además, analizar cómo afectan las emociones, porque quizás la mejor solución sea similar a la obtenida para la clasificación según género.

### **Sistema de monitoreo**

- Desarrollar un servicio de envío de los resultados, desde el dispositivo donde se hará el procesamiento de la voz, hacia el servidor, y su posterior almacenamiento. Este servicio debe ser robusto, rápido y debe tener los resultados siempre actualizados.
- Completar la implementación del sistema de visualización móvil. Además, mejorar dos conceptos de Usabilidad en la aplicación: Claridad de los conceptos y Prevención de Errores, ambos relacionados con el Historial de estados, que es la sección que más utilizará un usuario.
- Desarrollar una plataforma de visualización de resultados Web que permita a los familiares tener conocimiento del estado de sus abuelos y

que también les permita configurar algunos aspectos de la captura del audio, además de las configuraciones respecto al sistema de alertas.

- Realizar de un estudio de Usabilidad para la aplicación web a diseñar e implementar, ya que es un herramienta muy importante para conocer las necesidades del usuario.
- Desarrollar un sistema de alertas que permita, por distintas vías, enviar mensajes de aviso cuando el estado de un adulto mayor es vulnerable. Estas notificaciones pueden realizarse vía correo electrónico, SMS, etc.

### **Privacidad de los datos**

- Elaborar una política de privacidad que se encargue de velar por la confidencialidad de las grabaciones, y asegurar de que solo serán con fines de monitoreo y no se extraerá más información de ellas que la emoción, género, e identidad de la persona.

## Bibliografía y Referencias

- [1] Shimura, M., Monma, F., Mitsuyoshi, S., Shuzo, M., Yamamoto, T., Yamada, I. 2010. Descriptive analysis of emotion and feeling in voice. *Proceedings of the 2010 International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE)*. pp. 1-4.
- [2] Alghowinem, S., Goecke, R., Wagner, M. 2012. From joyous to clinically depressed: Mood detection using spontaneous speech. *Proceedings of the International Florida Artificial Intelligence Research Society Conference (FLAIRS 2012)*, AAAI Press, Marco Island USA, pp. 141-146.
- [3] Dickerson, R., Gorlin, E., Stankovic, J. 2011. Empath: a continuous remote emotional health monitoring system for depressive illness. *Proceedings of the Wireless Health'11*, San Diego USA, pp. 5-5.
- [4] Eyben, F., Wöllmer, M., Schuller, B. 2010. openSMILE - The Munich Versatile and Fast Open-Source Audio Feature Extractor. *Proceedings of ACM Multimedia (MM)*, ACM, Florence, Italy, pp. 1459-1462.
- [5] Eyben, F. 2010. openSMILE: the Munich open Speech and Music Interpretation by Large Space Extraction toolkit. *openSMILE Book*, URL: [http://sourceforge.net/projects/opensmile/files/openSMILE\\_book\\_1.0.0.pdf/download](http://sourceforge.net/projects/opensmile/files/openSMILE_book_1.0.0.pdf/download).
- [6] Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C., Narayanan, S. 2010. The INTERSPEECH 2010 Paralinguistic Challenge. *Proceedings of the INTERSPEECH 2010*, Makuhari, Japan, pp. 2794-2797.
- [7] Schuller, B. 2012. The Computational Paralinguistics Challenge. *IEEE Signal Processing Magazine*, 29 (4), pp. 97-101.
- [8] Schuller, B., Batliner, A., Steidl, S., Seppi, D. 2011. Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech Communication (SPECOM), Special Issue: Sensing Emotion and Affect - Facing Realism in Speech Processing*, ELSEVIER, Vol. 53, No. 9/10, pp. 1062-1087.
- [9] Hsu, C., Chang, C., Lin, C. 2010. A Practical Guide to Support Vector Classification, Taiwan.
- [10] Lay, C. Y., James, N. H. 2005. Gender Classification from Speech, URL: <http://www.projapps.com/CS5240.doc>
- [11] Meena, K., Subramaniam, K., Gomathy, M. 2013. Gender Classification in Speech Recognition using Fuzzy Logic and Neural Network. *The International Arab Journal of Information Technology*.

- [12] Harb, H., Chen, L., Auloge, J.Y. 2001. Speech/Music/Silence and Gender detection algorithm. *Proceedings of the 7th International conference on Distributed Multimedia Systems DMS01*, Taipei, Taiwan, pp 257-262.
- [13] Sedaaghi, M. 2008. Gender Classification in Emotional Speech, *Speech Recognition, France Mihelic and Janez Zibert (Ed.)*, ISBN: 978-953-7619-29-9, InTech, DOI: 10.5772/6385, pp. 550, URL: [http://www.intechopen.com/books/speech\\_recognition/gender\\_classification\\_in\\_emotional\\_speech](http://www.intechopen.com/books/speech_recognition/gender_classification_in_emotional_speech)
- [14] Haq, S., Jackson, P.J.B., Edge, J.D. 2008. Audio-Visual Feature Selection and Reduction for Emotion Classification. *Proceedings of International Conference on Auditory-Visual Speech Processing*, Tangalooma, Australia, pp. 185-190.
- [15] Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W., Weiss, B. 2005. A Database of German Emotional Speech. *Proceedings of the INTERSPEECH 2005*, Lisboa, Portugal, pp. 1517-1520.
- [16] López, J.M., Cearreta, I., Garay, N., de López Ipiña, K., Beristain, A. 2006. Creación de una base de datos emocional bilingüe y multimodal. *Proceedings of the 7th Spanish Human Computer Interaction Conference, Interaccion'06*, Puertollano, España, pp. 55-66.
- [17] Valstar, M., Schuller, B., Smith, K., Eyben, F., Jiang, B., Bilakhia, S., Schnieder, S., Cowie, R., Pantic, M. 2013. AVEC 2013 – The Continuous Audio/Visual Emotion and Depression Recognition Challenge. *Proceedings of the 21st ACM International Conference on Multimedia (MM 2013)*, ACM, Barcelona, España.
- [18] McDuff, D., Karlson, A., Kapoor, A., Roseway, A., Czerwinski, M. 2012. AffectAura: An Intelligent System for Emotional Memory. *Proceedings of CHI 2012*, ACM.
- [19] Escalona, M. 2012. Emotion recognition from speech. *Erasmus exchange project work*, Ljubljana, Eslovenia, URL: [http://upcommons.upc.edu/pfc/bitstream/2099.1/15362/1/THESIS\\_EMOTION\\_SPEECH\\_RECOGNITION.pdf](http://upcommons.upc.edu/pfc/bitstream/2099.1/15362/1/THESIS_EMOTION_SPEECH_RECOGNITION.pdf)
- [20] Colibazzi, T., Posner, J., Wang, Z., Gorman, D., Gerber, A., Yu, S., Zhu, H., Kangarlu, A., Duan, Y., Russell, J.A., Peterson, B.S. 2010. Neural systems subserving valence and arousal during the experience of induced emotions. *Emotion*, Vol. 10(3), pp. 377-389.

- [21] Schuller, B., Steidl, S., Batliner, A., Vinciarelli, A., Scherer, K., Ringeval, F., Chetouani, M., Wenginger, F., Eyben, F., Marchi, E., Mortillaro, M., Salamin, H., Polychroniou, A., Valente, F., Kim, S. 2013. The INTERSPEECH 2013 Computational Paralinguistics Challenge: Social Signals, Conflict, Emotion, Autism. *Proceedings of INTERSPEECH 2013*, ISCA, Lyon, France.
- [22] Schuller, B., Vlasenko, B., Eyben, F., Rigoll, G., Wendemuth, A. 2009. Acoustic Emotion Recognition: A Benchmark Comparison of Performances. *Proceedings of Automatic Speech Recognition and Understanding Workshop (ASRU)*, IEEE, Merano, Italia, pp. 552-557.
- [23] Cummins, N., Epps, J., Kua, J.M.K. 2012. A Comparison of Classification Paradigms for Speaker Likeability Determination. *Proceedings of the INTERSPEECH 2012*, Portland, Estado Unidos, pp. 4.
- [24] Steidl, S. 2010. Vocal Emotion Recognition: State-of-the-Art in Classification of Real-Life Emotions. International Computer Science Institute (ICSI). Berkeley, Estados Unidos. URL: <http://www.stanford.edu/class/linguist287/materials/steidl.pdf>
- [25] Nielsen, J. 1995. 10 Usability Heuristics for User Interface Design. URL: <http://www.nngroup.com/articles/ten-usability-heuristics/>
- [26] Snyder, C. 2003. Paper prototyping. Elsevier, Morgan Kaufmann, ISBN 10: 0-08-051350-6.
- [27] Android Developers - Design URL: <https://developer.android.com/design/index.html>
- [28] Nielsen, J. 1993. Usability Engineering. ISBN-10: 0125184069.
- [29] Tilvis, R.S., Routasalo, P., Karppinen, H., Strandberg, T.E., Kautiainen, H., Pitkala, K.H., 2012. Social isolation, social activity and loneliness as survival indicators in old age; a nationwide survey with a 7-year follow-up. *European Geriatric Medicine*, Vol. 3(1), pp. 18-22.
- [30] You, K.S., Lee, H. 2006. The Physical, Mental, and Emotional Health of Older People Who Are Living Alone or With Relatives. *Archives of Psychiatric Nursing*, Vol. 20(4), pp. 193-201.
- [31] Cornwell, E.Y., WAITE, L.J. 2009. Social Disconnectedness, Perceived Isolation, and Health among Older Adults. *Journal of Health and Social Behavior*, Vol. 50(1), pp. 31-48.

## Anexo A: Pruebas Realizadas para Reconocimiento de Emociones, Género e Identidad

A continuación se detalla cada una de las pruebas realizadas para lo que es el reconocimiento de emociones, género e identidad, describiendo su respectivo muestreo, características extraídas, generación de modelos y los resultados obtenidos. Además, se incluye una prueba básica de la distancia de grabación, y pruebas de conversaciones, en las cuales se verifica el buen funcionamiento de la extracción de frases, y todo lo que es reconocimiento.

Hay que aclarar que estas pruebas fueron hechas con un computador, usando openSMILE v1.0.1 en las tres primeras. Desde la cuarta prueba en adelante, se utilizó la última versión de openSMILE, la versión 2.0-rc1, liberada en mayo del 2013.

### A.1. Pruebas de Distancia

Se grabó desde distintas distancias a una persona hablando, y aproximadamente un metro y medio sería un buen límite para obtener buenas grabaciones, para su posterior procesamiento y análisis.

### A.2. Pruebas de Emociones

La idea de esta prueba era ver si existen grandes diferencias entre distintas emociones, al analizar algunas características simples como el tono de voz y la intensidad de ésta.

La prueba se hizo grabando la voz de tres personas, dos mujeres y un hombre. A cada persona se le grabó diciendo una frase, expresando la emoción felicidad, normal (como leyendo un texto) y tristeza. Se utilizó la configuración *prosodyShs* de openSMILE. De cada archivo generado se extrajeron métricas como el promedio del valor del tono y de la intensidad, y también sus máximos. Las siguientes tablas muestran los valores obtenidos en las grabaciones realizadas.

Tabla 7: Tono e intensidad para grabación 1, con distintas emociones.

Emoción/ Característica	Tono promedio	Tono máximo	Intensidad promedio	Intensidad máxima
Feliz	51	380	0,19	0,86
Normal	29,63	289,69	0,22	0,75
Triste	28,33	228,50	0,19	0,65

Tabla 8: Tono e intensidad para grabación 2, con distintas emociones.

Emoción/ Característica	Tono promedio	Tono máximo	Intensidad promedio	Intensidad máxima
<b>Feliz</b>	60,58	261,78	0,31	1,28
<b>Normal</b>	34,46	154,12	0,25	1,08
<b>Triste</b>	34,01	146,89	0,40	1,01

Tabla 9: Tono e intensidad para grabación 3, con distintas emociones.

Emoción/ Característica	Tono promedio	Tono máximo	Intensidad promedio	Intensidad máxima
<b>Feliz</b>	62,68	359,65	0,37	1,02
<b>Normal</b>	62,34	323,82	0,26	0,81
<b>Triste</b>	83,43	292,84	0,25	0,70

Comparando los distintos resultados, se puede ver que la intensidad máxima disminuye siguiendo el orden feliz, normal y triste. Hay una mayor diferencia entre feliz y normal, pero ésta es más pequeña entre normal y triste, lo que hace que la intensidad como factor importante para diferenciar una emoción de otra no sea suficiente. En el caso del tono, su valor máximo refleja de forma similar lo hecho por la intensidad máxima, disminuyendo también desde feliz a triste.

Los valores promedios vistos en las tablas no son totalmente esclarecedores, lo que en parte se debe a que se tomaron todos los valores del tono e intensidad, siendo cero en algunos casos (tono), cuando hay pausas entre una palabra y otra. Si bien ambas características son importantes para diferenciar una emoción de otra, no son suficientes.

Un punto también importante a considerar después de estas pruebas, es que obtener una grabación de una emoción de una persona no es fácil si es en vivo. Al estar grabando en vivo es más difícil que aquella emoción resulte realmente natural. Es por esto, que para obtener resultados más confiables se requiere de un conjunto de grabaciones en que sea totalmente clara y natural la emoción a estudiar.

### **A.3. Pruebas de Género e Identificación de una persona**

Antes de detectar qué emoción ha sido expresada por la persona que está siendo monitoreada, se necesita identificar a esta persona. En primera

instancia, se debe conocer el género de la persona que está hablando, para luego conocer su identidad comparando con un conjunto de registros hechos previamente. De esta manera, se puede identificar si la persona en cuestión es alguno de los adultos mayores que están siendo monitoreados específicamente en el lugar donde se hace la grabación. Esto ayudará a descartar voces que sean externas, por ejemplo la voz de un vecino o amigo que esté visitando a los abuelos.

En el siguiente dibujo se puede ver gráficamente el enfoque inicial en este trabajo, al momento de querer reconocer género, identidad y emoción.



*Ilustración 19: Enfoque inicial del proceso de reconocimiento.*

### **A.3.1. Prueba 1: Primera prueba**

Esta prueba consistió en el reconocimiento de género e identidad de una voz. Para ello se grabaron aproximadamente 3 minutos de audio mientras una persona leía. El experimento contó con grabaciones de 5 mujeres y 7 hombres, la mayoría en el rango de edad de 23 y 26 años, con excepción de dos personas, una mujer de 56 y un hombre de 58 años de edad. Las voces fueron registradas con distintos dispositivos, algunas con teléfonos celulares y otras con el computador.

#### **Muestreo**

Para comenzar, se normalizaron los audios a una misma frecuencia (8000 KHz), a un solo canal (Audio Mono) y formato WAV. Se tomaron las 5 voces de las mujeres y 5 voces de los hombres. A cada grabación se le extrajeron 25 trozos de 5,5 segundos, tiempo necesario para generar los 25 trozos y también dejar algunos extras para pruebas posteriores. La normalización y muestreo se hicieron usando Audacity<sup>7</sup>.

#### **Extracción de Características**

Ya contando con una cantidad no tan pequeña de muestras, se procedió a la extracción de características de cada una de ellas. Las

<sup>7</sup> <http://audacity.sourceforge.net/>

características que fueron extraídas son las pertenecientes al conjunto *emobase2010* de openSMILE.

Por cada conjunto de 25 muestras de cada persona se generó un archivo de características, en formato ARFF. Estos archivos se convirtieron a formato *lsvm*, usando un script en Perl que trae openSMILE, *arffToLsvm.pl*, con una pequeña modificación para que no tome el primer campo del archivo ARFF, el cual es un string, y no es característica de audio. Esta conversión se hizo para poder utilizar el programa LIBSVM<sup>8</sup>, el cual permite entrenar modelos SVM en base a vectores de características. Teniendo los modelos ya entrenados, se pueden contrastar nuevos conjuntos de descriptores con estos modelos, y así reconocer si pertenecen o no a alguna de las clases que el modelo describe.

### **Modelos y Entrenamiento**

Se generaron 3 modelos diferentes: modelo de mujeres (5 clases, una por cada mujer), modelo de hombres (5 clases, uno por cada hombre seleccionado), y modelo mujer-hombre (2 clases, mujer y hombre). Cada modelo se generó a partir de la unión de los archivos *lsvm* correspondientes, etiquetando las clases en cada caso. El entrenamiento de los modelos se hizo con LIBSVM, utilizando los siguientes parámetros: `svm_type = C-SVC`, `kernel_type = polynomial`, `degree = 2`, `gamma = 2`.

### **Resultados y Análisis**

Las pruebas con estos modelos se realizaron de la siguiente manera: se eligieron entre 20 y 25 muestras que no fueron utilizadas para entrenar los modelos, y entre 2 y 7 muestras que ya habían sido parte del entrenamiento. Sólo hubo una excepción, ésta fue con las muestras de la mujer3, por no contar con suficientes muestras, teniendo que utilizar 25 ya previamente elegidas y 2 que no lo habían sido, al revés que los otros casos. En total suman 27 trozos de audio por persona. Cada uno de los trozos fue enfrentado a los modelos correspondientes, las grabaciones de las mujeres se contrastaron con los modelos Mujer-Hombre y Mujer, y las de los hombres con Mujer-Hombre y Hombre.

Cada uno de los modelos generados cuenta con clases, que permiten clasificar los datos de entrada. Mujer-Hombre tiene la clase mujer y la clase hombre, representadas numéricamente, lo mismo pasa con los otros modelos, teniendo 5 clases de mujeres y 5 clases de hombre, respectivamente, representando cada una de esas clases la voz de una persona.

- VP = Verdadero Positivo = Valores que corresponden a la clase y son considerados correctamente.

---

<sup>8</sup> <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

- FP = Falso Positivo = Valores que no corresponden a la clase, y son considerados positivos.

*Tabla 10: Resultados de pruebas con modelo Mujer-Hombre.*

Clase/Resultado	VP	FP
<b>Mujer</b>	135	2
<b>Hombre</b>	133	0

*Tabla 11: Resultados de pruebas modelo Mujer-Hombre, contrastando voces masculinas extras.*

Clase/Resultado	VP	FP
<b>Mujer</b>	0	0
<b>Hombre</b>	6	0

*Tabla 12: Resultados de pruebas modelo Mujer.*

Clase/Resultado	VP	FP
<b>Mujer1</b>	25	0
<b>Mujer2</b>	27	9
<b>Mujer3</b>	27	0
<b>Mujer4</b>	20	0
<b>Mujer5</b>	26	1

*Tabla 13: Resultados de pruebas modelo Hombre.*

Clase/Resultado	VP	FP
<b>Hombre1</b>	27	0
<b>Hombre2</b>	27	0
<b>Hombre3</b>	27	0
<b>Hombre4</b>	27	1
<b>Hombre5</b>	26	0

Como se mencionó previamente, de las 7 grabaciones masculinas, 5 se usaron para las pruebas recién descritas, especialmente para la generación de los modelos. Luego, las 2 restantes, que son de menor duración (20 segundos aproximadamente), que no intervinieron en el entrenamiento, fueron comparadas con el modelo Mujer-Hombre generando el resultado de la Tabla 11.

Es importante mencionar, que el conjunto de características utilizado toma en cuenta mediciones de tiempo, duración de la pista de audio. Esto genera resultados no muy consistentes cuando las pistas de audio son de distinta extensión de tiempo. Todos los trozos de audio utilizados en estas pruebas fueron de 5,5 segundos de duración.

Antes de analizar de forma global las métricas obtenidas, hay un punto interesante que evidencia esta prueba. Del grupo de mujeres, Tabla 12, la mujer4 es hija de la mujer2, y del grupo de hombres, Tabla 13, el hombre5 es hijo del hombre4. Para ambos hijos, se dio el caso en que una o más de sus muestras fueron confundidas con su ascendiente. Es importante evitar esa confusión al momento de implementar el sistema. Imaginando el caso en que los adultos mayores se encontraran con alguno de sus hijos, el sistema podría no reconocer de forma correcta la identidad de la persona, arrojando la identidad del padre o madre, en vez de la del hijo, o viceversa.

Los resultados de esta prueba de reconocimiento de género e identidad, arrojan un alto porcentaje de precisión con respecto a los resultados esperados. A pesar de ello, se puede notar que hay casos en que el modelo entrega resultados erróneos, especialmente en el modelo Mujer, pudiendo confundir casi cualquiera de las voces femeninas con la voz de la mujer2, lo que no ocurre con todas las muestras, pero igual es un aspecto relevante a considerar. Algo similar se puede evidenciar con el modelo Mujer-Hombre, donde dos trozos de audio de hombre fueron considerados como de mujer. Las causas que pueden llevar a estos resultados, pueden ser por un lado la poca cantidad de muestras utilizadas, siendo necesaria una cantidad superior, del orden de miles. Por otro lado, tenemos el vector de características, el cual puede estar considerando rasgos no realmente relevantes, o por el contrario, no estar considerando algunos que sí lo son. Otra causa posible son los parámetros dados al momento de entrenar el modelo.

Dados los resultados, los cuales reflejan un buen camino, parece conveniente seguir la misma línea en el siguiente experimento. Esta vez, utilizando más muestras, comparando diferentes modelos (cambiando parámetros), automatizando más el proceso, comenzando con la grabación de la voz, hasta su predicción con el modelo.

### **A.3.2. Prueba 2: Exactitud de la clasificación**

Para tener una idea de la exactitud con que un modelo predice a qué clase pertenecen los datos de prueba, se realizó el mismo procedimiento anterior, pero esta vez añadiendo el parámetro  $b = 1$  al entrenar el modelo. Esta variable indica que el modelo entregará los resultados con las probabilidades de que una muestra pertenezca a una clase o a la otra. Estos datos son mucho más útiles que sólo saber la clase predicha, porque el modelo podría decir que un dato es de la clase 1 y sólo porque su valor es mayor al de las otras clases, aunque ese valor sea bajo (especialmente cuando hay muchas clases). Con estos valores a mano se pueden hacer mejoras en búsqueda de un valor de predicción que sea aceptable. Por ejemplo: "los datos serán válidos si la predicción arroja un valor sobre el 80%".

#### **Muestreo**

Como el experimento es casi igual al anterior, se hizo uso de las mismas muestras de la prueba pasada, sin hacer cambios en ellas.

#### **Extracción de Características**

También se repitió el uso del conjunto de características *emobase2010*. Desde esta prueba en adelante, la extracción de características y el procesamiento de ellas, se hizo de forma automática, tanto para muestras de entrenamiento como de prueba, mediante scripts. En la sección 5.1 se puede revisar con detalle la composición y funcionamiento de estos scripts.

#### **Modelos y Entrenamiento**

Recordando, el entrenamiento de los modelos se hizo utilizando los siguientes parámetros: `svm_type = C-SVC`, `kernel_type = polynomial`, `degree = 2`, `gamma = 2`, y se añadió `b = 1`.

Desde esta prueba también se automatizó la creación y el posterior contraste de los modelos. Para más detalles, ver sección 5.1.

#### **Resultados y Análisis**

Resultados de clasificación según género y reconocimiento de identidad, pero esta vez visualizados a través del porcentaje de exactitud.

Tabla 14: Resultados de clasificación según género en conjunto Mujer.

Mujer/Clase	Mujer (%)	Hombre (%)
<b>Mujer1</b>	98,3	1,7
<b>Mujer2</b>	95,0	5,0
<b>Mujer3</b>	99,3	0,7
<b>Mujer4</b>	98,8	1,2
<b>Mujer5</b>	99,2	0,8
<b>Promedio Total</b>	98,12	1,88

Tabla 15: Resultados de clasificación según género en conjunto Hombre.

Hombre/Clase	Mujer (%)	Hombre (%)
<b>Hombre1</b>	1,7	98,3
<b>Hombre2</b>	3,7	96,3
<b>Hombre3</b>	3,3	96,7
<b>Hombre4</b>	1,0	99,0
<b>Hombre5</b>	1,9	98,1
<b>Promedio Total</b>	2,32	97,68

Tabla 16: Resultados de clasificación según género en conjunto Hombre-Otro.

Hombre-Otro/Clase	Mujer (%)	Hombre (%)
<b>Hombre-Otro1</b>	10,3	89,7
<b>Hombre-Otro2</b>	1,4	98,4
<b>Promedio Total</b>	5,85	94,05

Tabla 17: Resultados de reconocimiento de identidad en conjunto Mujer.

Mujer/Clase	Mujer1 (%)	Mujer2 (%)	Mujer3 (%)	Mujer4 (%)	Mujer5 (%)
<b>Mujer1</b>	67,5	6,4	6,8	3,9	15,5
<b>Mujer2</b>	11,4	67,8	1,9	14,7	4,3
<b>Mujer3</b>	1,3	4,1	88,6	2,8	3,2
<b>Mujer4</b>	5,3	32,1	1,6	55,5	5,5
<b>Mujer5</b>	3,3	9,3	2,5	7,0	77,9
<b>Promedio Total</b>					71,46

Tabla 18: Resultados de reconocimiento de identidad en conjunto Hombre.

Hombre/Clase	Hombre1 (%)	Hombre2 (%)	Hombre3 (%)	Hombre4 (%)	Hombre5 (%)
<b>Hombre1</b>	82,4	3,1	5,4	3,9	5,2
<b>Hombre2</b>	3,9	81,0	2,4	5,5	7,1
<b>Hombre3</b>	12,9	2,1	80,2	1,6	3,2
<b>Hombre4</b>	4,5	6,0	1,6	76,2	11,7
<b>Hombre5</b>	2,6	4,1	2,1	6,5	84,7
<b>Promedio Total</b>					80,9

Se puede apreciar que la mayoría de los resultados son bastante buenos, alcanzando una alta exactitud en la predicción. Pero también hay excepciones. La excepción está en la detección de identidad en el grupo de las mujeres, que ya se vio previamente en la Tabla 12. En esos resultados se notó una confusión entre algunas identidades. Esto se refleja también en la matriz de confusión de la Tabla 17, en que los porcentajes de predicción de los datos de entrenamiento de una mujer contra los de prueba de sí misma no superan en promedio el 72%.

Hasta ahora la clasificación según género marcha bien, pero para la detección de identidad hay que tener cuidado, especialmente para detectar una identidad femenina.

### **A.3.3. Prueba 3: Clasificación según género en audio con emociones**

Para las posteriores pruebas con emociones (ver sección A.4), fue posible conseguirse tres bases de datos de audio, las que serán descritas más en detalle en la sección de pruebas con emociones. Al momento de hacer esta prueba se contaba con dos de las bases de datos, una de ellas es SAVEE, la cual cuenta con grabaciones de 4 usuarios hombres, de iniciales DC, JE, JK y KL. La otra base de datos contrastada es Emo-DB. Las grabaciones que trae están en alemán y cuenta con 10 usuarios, 5 mujeres y 5 hombres, cada uno identificado con un número entre el 3 y el 16.

Se decidió probar el modelo de clasificación según género para probar su rendimiento en pistas de audio con emociones diversas y más reales. Entonces, se contrastó el modelo con parte de las muestras de las bases de datos emocionales antes descritas. Además, parte de este trabajo es poder reconocer género en la voz, sin importar qué lo que ésta expresa, o cómo lo expresa.

#### **Muestreo**

Para cada usuario de SAVEE se eligieron tres archivos de audio por emoción, los tres primeros impares (es decir, a01.wav, a03.wav, a05.wav). Además, para cada usuario de Emo-DB se eligieron los cuatro primeros archivos de audio.

#### **Extracción de Características**

Las características extraídas correspondieron a las mismas anteriormente extraídas, es decir, usando el vector emobase2010.

#### **Modelos y Entrenamiento**

Se utilizó el mismo modelo Mujer-Hombre entrenado previamente, incluyendo el parámetro  $b = 1$ , el cual se usará para el resto de las pruebas, por lo que no se volverá a nombrar su uso.

#### **Resultados y Análisis**

Los resultados obtenidos una vez contrastado el nuevo modelo se presentan en las siguiente tablas, mostrando solo aquellos que corresponden a las muestras de audio de emociones que fueron agregadas:

Tabla 19: Resultados de pruebas en conjunto Hombre-SAVEE.

Hombre-SAVEE/Clase	Mujer (%)	Hombre (%)
Hombre-SAVEE1 (DC)	0,0	100,0
Hombre-SAVEE2 (JE)	0,0	100,0
Hombre-SAVEE3 (JK)	0,0	100,0
Hombre-SAVEE4 (KL)	0,0	100,0
<b>Promedio Total</b>	<b>0</b>	<b>100</b>

Tabla 20: Resultados de pruebas en conjunto Mujer-Emo-DB.

Mujer-Emo-DB/Clase	Mujer (%)	Hombre (%)
Mujer-Emo-DB1 (08)	0,0	100,0
Mujer-Emo-DB2 (09)	0,0	100,0
Mujer-Emo-DB3 (13)	0,1	99,9
Mujer-Emo-DB4 (14)	0,2	99,8
Mujer-Emo-DB5 (16)	0,0	100,0
<b>Promedio Total</b>	<b>0,06</b>	<b>99,94</b>

Tabla 21: Resultados de pruebas en conjunto Hombre-Emo-DB.

Hombre-Emo-DB/Clase	Mujer (%)	Hombre (%)
Hombre-Emo-DB1 (03)	0,0	100,0
Hombre-Emo-DB2 (10)	3,4	96,6
Hombre-Emo-DB3 (11)	0,2	99,8
Hombre-Emo-DB4 (12)	0,1	99,9
Hombre-Emo-DB5 (15)	0,1	99,9
<b>Promedio Total</b>	<b>0,76</b>	<b>99,24</b>

Los buenos resultados en clasificación de género no duraron mucho. Se puede ver que mujeres de la base de datos Emo-DB fueron todas clasificadas como hombre, con casi un 100% de exactitud (ver Tabla 20).

Hay dos opciones que se vislumbran para explicar este mal resultado. Una es que el conjunto de características utilizadas no se enfoca realmente en las características lingüísticas que diferencian cada género, y la otra es que los parámetros de creación del modelo no son los más adecuados, haciendo que algunas muestras caigan en la clase incorrecta.

#### **A.3.4. Prueba 4: Kernel lineal**

Se decidió probar con otros tipos de kernel, para ver si se podía mejorar el resultado obtenido anteriormente, en el que claramente no se logró reconocer correctamente voces femeninas como del género femenino. Revisando la documentación de LIBSVM, se llegó a una guía rápida desarrollada por los autores, donde se explica en forma más detallada, qué kernels son más convenientes en ciertas ocasiones, y cómo comparar el rendimiento entre ellos [9].

En esta guía se mencionan varias cosas importantes. La primera es que el kernel lineal sería más preciso y útil para separar clases que poseen un gran número de características, como las que posee *emobase2010* (contiene 1582 características). Lo otro que se menciona es una herramienta que trae, llamada *grid.py*, que permite conocer los mejores parámetros gamma y C para un kernel específico en un modelo de datos específico. Esta herramienta lo que hace es probar con una cantidad especificada de valores para gamma y C, y encuentra la mejor combinación. Y por último, esta guía aconseja el escalamiento de los valores de las características, tanto para muestras de entrenamiento como de prueba, es decir, que estos valores vayan dentro de un rango apropiado, por ejemplo [-1 , 1].

Al parecer el kernel lineal sería una buena opción según esta guía. Se decidió probar con este kernel, pero para no dejar atrás el otro kernel no probado, RBF, se hizo una comparación de ambos utilizando *grid.py*. Al encontrar los mejores valores de parámetros para RBF, y comparar los resultados con el lineal, los resultados eran exactamente iguales. Por lo tanto, para una gran cantidad de descriptores, los valores de los parámetros que mejor funcionan en RBF, asimilan o igualan su comportamiento al del kernel lineal, que usa los valores por defecto (C=1 y gamma no se usa). Por consiguiente, si se comportaron similarmente, pero recomiendan más el lineal, se dejó el lineal en esta prueba, a ver qué tal funcionaba.

También, se probó el escalamiento de los datos, usando *svm-scale* de LIBSVM. Lamentablemente no sirvió para las muestras de este trabajo. Al probar con el kernel lineal, ya que todas las muestras eran clasificadas en un

solo género, a diferencia del kernel lineal sin escalar los datos, que funcionó un poco mejor.

Al probar con un nuevo kernel, no sólo se contrastó nuevamente el modelo de género, también se contrastaron de nuevo los modelos de identidad, para ver si también se veían afectados por este cambio.

### **Muestreo**

En esta prueba se sumaron las muestras iniciales para género e identidad de los conjuntos Mujer, Hombre y Hombre-Otro, más las muestras para reconocimiento de género de Hombre-SAVEE, Mujer-Emo-DB y Hombre-Emo-DB, descritas previamente.

### **Extracción de Características**

Con la intención de sólo identificar si es el kernel el que no funcionó bien, se siguió haciendo uso del vector *emobase2010*.

### **Modelos y Entrenamiento**

Por la recomendación mencionada, se usó el kernel lineal para entrenar el modelo Mujer-Hombre que es para clasificación de género, y para el modelo Mujer y Hombre, que son para el reconocimiento de identidad.

### **Resultados y Análisis**

Se presentan a continuación los datos obtenidos para clasificación según género y para clasificación de identidad, usando el kernel lineal:

*Tabla 22: Resultados de clasificación de género en conjunto Mujer, usando kernel lineal.*

<b>Mujer/Clase</b>	<b>Mujer (%)</b>	<b>Hombre (%)</b>
<b>Mujer1</b>	99,0	1,0
<b>Mujer2</b>	98,2	1,8
<b>Mujer3</b>	99,5	0,5
<b>Mujer4</b>	99,4	0,6
<b>Mujer5</b>	98,8	1,2
<b>Promedio Total</b>	98,98	1,02

Tabla 23: Resultados de clasificación de género en conjunto Hombre, usando kernel lineal.

Hombre/Clase	Mujer (%)	Hombre (%)
Hombre1	0,2	99,8
Hombre2	1,9	98,1
Hombre3	1,5	98,5
Hombre4	0,0	100,0
Hombre5	0,7	99,3
<b>Promedio Total</b>	<b>0,86</b>	<b>99,14</b>

Tabla 24: Resultados de clasificación de género en conjunto Hombre-Otro, usando kernel lineal.

Hombre-Otro/Clase	Mujer (%)	Hombre (%)
Hombre-Otro1	15,0	85,0
Hombre-Otro2	0,2	99,8
<b>Promedio Total</b>	<b>7,6</b>	<b>92,4</b>

Tabla 25: Resultados de clasificación de género en conjunto Hombre-SAVEE, usando kernel lineal.

Hombre-SAVEE/Clase	Mujer (%)	Hombre (%)
Hombre-SAVEE1 (DC)	0,0	100,0
Hombre-SAVEE2 (JE)	2,8	97,2
Hombre-SAVEE3 (JK)	0,7	99,3
Hombre-SAVEE4 (KL)	0,0	100,0
<b>Promedio Total</b>	<b>0,88</b>	<b>99,13</b>

Tabla 26: Resultados de clasificación de género en conjunto Mujer-Emo-DB, usando kernel lineal.

<b>Mujer-Emo-DB/Clase</b>	<b>Mujer (%)</b>	<b>Hombre (%)</b>
<b>Mujer-Emo-DB1 (08)</b>	0,0	100,0
<b>Mujer-Emo-DB2 (09)</b>	9,6	90,4
<b>Mujer-Emo-DB3 (13)</b>	2,4	97,6
<b>Mujer-Emo-DB4 (14)</b>	5,1	94,9
<b>Mujer-Emo-DB5 (16)</b>	11,6	88,4
<b>Promedio Total</b>	5,74	94,26

Tabla 27: Resultados de clasificación de género en conjunto Hombre-Emo-DB, usando kernel lineal.

<b>Hombre-Emo-DB/Clase</b>	<b>Mujer (%)</b>	<b>Hombre (%)</b>
<b>Hombre-Emo-DB1 (03)</b>	0,0	100,0
<b>Hombre-Emo-DB2 (10)</b>	2,3	97,7
<b>Hombre-Emo-DB3 (11)</b>	0,0	100,0
<b>Hombre-Emo-DB4 (12)</b>	0,0	100,0
<b>Hombre-Emo-DB5 (15)</b>	16,1	83,9
<b>Promedio Total</b>	3,68	96,32

Tabla 28: Resultados de reconocimiento de identidad en conjunto Mujer, usando kernel lineal.

Mujer/Clase	Mujer1 (%)	Mujer2 (%)	Mujer3 (%)	Mujer4 (%)	Mujer5 (%)
<b>Mujer1</b>	90,9	2,8	2,0	1,3	3,0
<b>Mujer2</b>	3,8	86,2	0,6	7,7	1,7
<b>Mujer3</b>	0,9	1,7	94,2	1,8	1,4
<b>Mujer4</b>	2,8	15,6	0,8	78,6	2,2
<b>Mujer5</b>	1,9	4,3	1,1	5,2	87,5
<b>Promedio Total</b>					87,48

Tabla 29: Tabla 22: Resultados de reconocimiento de identidad en conjunto Hombre, usando kernel lineal.

Hombre/Clase	Hombre1 (%)	Hombre2 (%)	Hombre3 (%)	Hombre4 (%)	Hombre5 (%)
<b>Hombre1</b>	92,3	1,2	3,5	1,1	1,9
<b>Hombre2</b>	1,1	89,6	1,0	2,0	6,3
<b>Hombre3</b>	7,8	1,1	88,7	1,0	1,4
<b>Hombre4</b>	1,1	1,7	0,8	93,2	3,2
<b>Hombre5</b>	1,2	2,3	1,0	4,2	91,3
<b>Promedio Total</b>					91,02

Al utilizar un modelo con kernel de tipo lineal, los resultados con respecto a la detección de identidad mejoraron sustancialmente, aproximadamente en un 10% para el caso de hombres y mujeres. Para el reconocimiento de género los cambios fueron muy pocos, pero al menos ya no era casi 100% en las muestras de Mujer-Emo-DB reconociéndolas como hombre, ahora fue un poco más bajo ese valor.

De todas maneras, el kernel lineal logró efectos positivos en las clasificaciones, así que podría ser el kernel definitivo a utilizar. Las próximas pruebas lo dirán.

### A.3.5. Prueba 5: Vector de características *IS10\_paraling*

Debido a que el kernel lineal produjo una leve mejoría en las clasificaciones, entonces la causa del mal reconocimiento de género en el conjunto Mujer-Emo-DB siguió siendo otra. Ya se cambió el kernel, entonces esta vez se probó con otro conjunto de características, *IS10\_paraling*, que quizás se acerca más al propósito de reconocimiento de género e identidad.

#### Muestreo

Se probó el nuevo vector de características en todas las muestras para clasificación de género y para reconocimiento de identidad, incluyendo las últimas muestras de emociones agregadas.

#### Extracción de Características

El conjunto de características se cambió a *IS10\_paraling*. Este conjunto, descrito en el libro de openSMILE [5], se enfocaría más en la detección de las características lingüísticas de la voz, que en la expresión de ésta (emociones). Este podría ser mejor que *emobase2010*, el cual se enfocaría más en emociones.

#### Modelos y Entrenamiento

Se entrenaron los modelos en base a *IS10\_paraling* y usando el kernel lineal.

#### Resultados y Análisis

Se puede ver en las siguientes tablas los datos obtenidos usando el conjunto de características *IS10\_paraling* y el kernel lineal, tanto para reconocimiento de género como de identidad:

Tabla 30: Resultados de clasificación de género en conjunto Mujer, en base a *IS10\_paraling*.

Mujer/Clase	Mujer (%)	Hombre (%)
Mujer1	91,8	8,2
Mujer2	67,9	32,1
Mujer3	98,2	1,8
Mujer4	86,6	13,4
Mujer5	91,1	8,9
<b>Promedio Total</b>	<b>87,12</b>	<b>12,88</b>

Tabla 31: Resultados de clasificación de género en conjunto Hombre, en base a IS10\_paraling.

Hombre/Clase	Mujer (%)	Hombre (%)
Hombre1	3,0	97,0
Hombre2	4,0	96,0
Hombre3	26,3	73,7
Hombre4	1,4	98,6
Hombre5	14,5	85,5
<b>Promedio Total</b>	<b>9,84</b>	<b>90,16</b>

Tabla 32: Resultados de clasificación de género en conjunto Hombre-Otro, en base a IS10\_paraling.

Hombre-Otro/Clase	Mujer (%)	Hombre (%)
Hombre-Otro1	10,4	89,6
Hombre-Otro2	21,2	78,8
<b>Promedio Total</b>	<b>15,8</b>	<b>84,2</b>

Tabla 33: Resultados de clasificación de género en conjunto Hombre-SAVEE, en base a IS10\_paraling.

Hombre-SAVEE/Clase	Mujer (%)	Hombre (%)
Hombre-SAVEE1 (DC)	9,4	90,6
Hombre-SAVEE2 (JE)	57,7	42,3
Hombre-SAVEE3 (JK)	64,9	35,1
Hombre-SAVEE4 (KL)	34,9	65,1
<b>Promedio Total</b>	<b>41,73</b>	<b>58,28</b>

Tabla 34: Resultados de clasificación de género en conjunto Mujer-Emo-DB, en base a IS10\_paraling.

<b>Mujer-Emo-DB/Clase</b>	<b>Mujer (%)</b>	<b>Hombre (%)</b>
<b>Mujer-Emo-DB1 (08)</b>	79,2	20,8
<b>Mujer-Emo-DB2 (09)</b>	81,2	18,8
<b>Mujer-Emo-DB3 (13)</b>	92,6	7,4
<b>Mujer-Emo-DB4 (14)</b>	71,3	28,7
<b>Mujer-Emo-DB5 (16)</b>	86,0	14,0
<b>Promedio Total</b>	82,06	17,94

Tabla 35: Resultados de clasificación de género en conjunto Hombre-Emo-DB, en base a IS10\_paraling.

<b>Hombre-Emo-DB/Clase</b>	<b>Mujer (%)</b>	<b>Hombre (%)</b>
<b>Hombre-Emo-DB1 (03)</b>	67,0	33,0
<b>Hombre-Emo-DB2 (10)</b>	50,0	50,0
<b>Hombre-Emo-DB3 (11)</b>	2,1	97,9
<b>Hombre-Emo-DB4 (12)</b>	25,7	74,3
<b>Hombre-Emo-DB5 (15)</b>	27,0	73,0
<b>Promedio Total</b>	34,36	65,64

Tabla 36: Resultados de reconocimiento de identidad en conjunto Mujer, en base a *IS10\_paraling*.

Mujer/Clase	Mujer1 (%)	Mujer2 (%)	Mujer3 (%)	Mujer4 (%)	Mujer5 (%)
<b>Mujer1</b>	47,1	13,2	9,5	9,9	20,3
<b>Mujer2</b>	14,1	34,8	3,6	30,4	17,0
<b>Mujer3</b>	2,7	2,8	85,9	4,4	4,3
<b>Mujer4</b>	10,2	23,2	6,4	35,7	24,6
<b>Mujer5</b>	10,5	15,3	11,3	24,4	38,4
<b>Promedio Total</b>					48,38

Tabla 37: Resultados de reconocimiento de identidad en conjunto Hombre,, en base a *IS10\_paraling*

Hombre/Clase	Hombre1 (%)	Hombre2 (%)	Hombre3 (%)	Hombre4 (%)	Hombre5 (%)
<b>Hombre1</b>	48,3	16,1	11,3	3,4	20,9
<b>Hombre2</b>	14,3	41,8	4,7	14,0	25,2
<b>Hombre3</b>	23,9	2,9	56,1	0,9	16,2
<b>Hombre4</b>	5,5	16,1	1,4	66,6	10,5
<b>Hombre5</b>	21,4	19,9	9,6	5,9	43,1
<b>Promedio Total</b>					51,18

Se puede observar que la clasificación de género mejoró notoriamente para las muestras de Mujer-Emo-DB (Tabla 34). Para el resto de muestras, esta clasificación bajó, en algunos casos, como Hombre-SAVEE y Hombre-DB, casi llegando al 50% de promedio de clasificación como hombre. Esto genera dudas con el uso de *IS10\_paraling*, pero no siempre una mejora particular genera una mejora global. Hay que ser conscientes de que cada voz es diferente, y una clasificación tan cercana a 100% era casi imposible de tener para todos los casos. Por eso, hay que buscar una manera de que las nuevas mejoras permitan mantener un equilibrio en la clasificación para cualquier muestra, aunque no se llegue al 100%, pero que

sí se pueda clasificar una muestra con seguridad, por ejemplo, sobre el 75% de exactitud.

Para las pruebas de reconocimiento de identidad los resultados fueron bastante pobres. Por ahora se seguirá usando *emobase2010* para este tipo de pruebas. El desafío se mantiene en obtener una buena clasificación según género, para cualquier muestra, sea cual sea la emoción que exprese.

#### **A.3.6. Prueba 6: Muestras de audio con emociones en conjunto de entrenamiento para reconocimiento de género**

Con la intención de ver qué tanto influye la emoción al momento de clasificar según género, se probó incluir archivos de la base de datos de emociones Emo-DB en las muestras de entrenamiento. Desde esta prueba en adelante no se incluye la detección de identidad.

##### **Muestreo**

La diferencia con el experimento anterior es que se incluyeron al conjunto de entrenamiento 6 archivos de audio de cada usuario de Emo-DB, que no hubieran sido seleccionados para el conjunto de prueba. Estos archivos pertenecen a las distintas emociones grabadas por cada usuario.

##### **Extracción de Características**

El vector de características *IS10\_paraling* se mantuvo como el vector elegido.

##### **Modelos y Entrenamiento**

Se entrenó el modelo Mujer-Hombre, ahora incluyendo las nuevas muestras, usando como parámetro el kernel lineal.

##### **Resultados y Análisis**

Las tablas de los resultados fueron reagrupadas, dejando primero los datos de los grupos de mujeres, y luego los de los hombres, esperando que así se pueda visualizar mejor la clasificación en cada género.

Tabla 38: Resultados de clasificación de género en conjunto Mujer.

Mujer/Clase	Mujer (%)	Hombre (%)
<b>Mujer1</b>	81,0	19,0
<b>Mujer2</b>	73,9	26,1
<b>Mujer3</b>	89,2	10,8
<b>Mujer4</b>	68,3	31,7
<b>Mujer5</b>	77,3	22,7
<b>Promedio Total</b>	77,94	22,06

Tabla 39: Resultados de clasificación de género en conjunto Mujer-Emo-DB.

Mujer-Emo-DB/Clase	Mujer (%)	Hombre (%)
<b>Mujer-Emo-DB1 (08)</b>	83,5	16,5
<b>Mujer-Emo-DB2 (09)</b>	55,3	44,7
<b>Mujer-Emo-DB3 (13)</b>	67,8	32,2
<b>Mujer-Emo-DB4 (14)</b>	72,4	27,6
<b>Mujer-Emo-DB5 (16)</b>	80,4	19,6
<b>Promedio Total</b>	71,88	28,12

Tabla 40: Resultados de clasificación de género en conjunto Hombre.

Hombre/Clase	Mujer (%)	Hombre (%)
<b>Hombre1</b>	23,3	76,7
<b>Hombre2</b>	28,9	71,1
<b>Hombre3</b>	40,2	59,8
<b>Hombre4</b>	18,7	81,3
<b>Hombre5</b>	23,5	76,5
<b>Promedio Total</b>	26,92	73,08

Tabla 41: Resultados de clasificación de género en conjunto Hombre-Otro.

Hombre-Otro/Clase	Mujer (%)	Hombre (%)
<b>Hombre-Otro1</b>	56,5	43,5
<b>Hombre-Otro2</b>	8,5	91,5
<b>Promedio Total</b>	32,5	67,5

Tabla 42: Resultados de clasificación de género en conjunto Hombre-SAVEE.

Hombre-SAVEE/Clase	Mujer (%)	Hombre (%)
<b>Hombre-SAVEE1 (DC)</b>	33,4	66,6
<b>Hombre-SAVEE2 (JE)</b>	26,7	73,3
<b>Hombre-SAVEE3 (JK)</b>	10,4	89,6
<b>Hombre-SAVEE4 (KL)</b>	26,7	73,3
<b>Promedio Total</b>	24,3	75,7

Tabla 43: Resultados de clasificación de género en conjunto Hombre-Emo-DB.

Hombre-Emo-DB/Clase	Mujer (%)	Hombre (%)
<b>Hombre-Emo-DB1 (03)</b>	10,0	90,0
<b>Hombre-Emo-DB2 (10)</b>	14,1	85,9
<b>Hombre-Emo-DB3 (11)</b>	13,7	86,3
<b>Hombre-Emo-DB4 (12)</b>	8,4	91,6
<b>Hombre-Emo-DB5 (15)</b>	13,7	86,3
<b>Promedio Total</b>	11,98	88,02

Como se mencionó anteriormente, algunas mejoras para algunos, pueden traer perjuicios para otros. Eso se puede ver en estos resultados al añadir muestras de emociones en el grupo de entrenamiento. Las muestras de los grupos masculinos que antes se veían más afectados, Hombre-SAVEE y Hombre-DB, subieron su exactitud de clasificación. Por otro lado, esta

exactitud bajó en los grupos femeninos. Esto podría deberse a que hay ciertas características en las muestras con emociones que tienden a mejorar la clasificación en hombres, en perjuicio de la clasificación en mujeres. Más adelante se tomará una decisión de si mantiene o no este nuevo modelo Mujer-Hombre.

### **A.3.7. Prueba 7: Normalización del volumen**

Aunque los resultados anteriores no son malos, siendo casi todos sobre el 70%, hay un caso en que falla rotundamente, y es en la detección de género del Hombre-Otro1 en la Tabla 41. Al revisar algunas archivos de audio, que eran parte de las muestras, se notó que no todos tenía el mismo volumen, y se pensó que esto podría estar generando alguna distorsión al momento de crear el modelo y posteriormente al evaluarlo. Además, si se analiza bien, el volumen del audio grabado no debería influenciar en los rasgos de la voz para un género específico, por lo que normalizar el volumen en todas las muestras podría significar una mejora al momento de probar el modelo.

#### **Muestreo**

En esta ocasión se utilizaron las mismas muestras para entrenamiento y para testeo, utilizadas en la prueba anterior. La diferencia está en que todos los archivos de audio utilizados fueron normalizados a un mismo nivel de audio. Esto se realizó mediante el programa SoX<sup>9</sup>, programa de consola para edición de audio, y el nivel de volumen elegido fue -10, siendo un término medio entre los archivos de audio con volumen más fuerte (cercano a 0) y los con volumen más bajo (cercano a -20).

#### **Extracción de Características**

Se usó el conjunto de características *IS10\_paraling*.

#### **Modelos y Entrenamiento**

El modelo Mujer-Hombre se entrenó usando el kernel lineal.

#### **Resultados y Análisis**

Las tablas posteriores muestran los resultados para clasificación según género, luego de haber normalizado el volumen de todas las muestras.

---

<sup>9</sup> <http://sox.sourceforge.net/>

Tabla 44: Resultados de clasificación según género en conjunto Mujer.

Mujer/Clase	Mujer (%)	Hombre (%)
<b>Mujer1</b>	80,9	19,1
<b>Mujer2</b>	73,9	26,1
<b>Mujer3</b>	89,0	11,0
<b>Mujer4</b>	68,2	31,8
<b>Mujer5</b>	77,3	22,7
<b>Promedio Total</b>	77,86	22,14

Tabla 45: Resultados de clasificación según género en conjunto Mujer-Emo-DB.

Mujer-Emo-DB/Clase	Mujer (%)	Hombre (%)
<b>Mujer-Emo-DB1 (08)</b>	83,5	16,5
<b>Mujer-Emo-DB2 (09)</b>	55,3	44,7
<b>Mujer-Emo-DB3 (13)</b>	67,8	32,2
<b>Mujer-Emo-DB4 (14)</b>	72,4	27,6
<b>Mujer-Emo-DB5 (16)</b>	80,4	19,6
<b>Promedio Total</b>	71,88	28,12

Tabla 46: Resultados de clasificación según género en conjunto Hombre.

Hombre/Clase	Mujer (%)	Hombre (%)
<b>Hombre1</b>	23,3	76,7
<b>Hombre2</b>	28,8	71,2
<b>Hombre3</b>	40,3	59,7
<b>Hombre4</b>	18,7	81,3
<b>Hombre5</b>	23,5	76,5
<b>Promedio Total</b>	26,92	73,08

Tabla 47: Resultados de clasificación según género en conjunto Hombre-Otro.

Hombre-Otro/Clase	Mujer (%)	Hombre (%)
<b>Hombre-Otro1</b>	35,0	65,0
<b>Hombre-Otro2</b>	3,6	96,4
<b>Promedio Total</b>	19,3	80,7

Tabla 48: Resultados de clasificación según género en conjunto Hombre-SAVEE.

Hombre-SAVEE/Clase	Mujer (%)	Hombre (%)
<b>Hombre-SAVEE1 (DC)</b>	33,4	66,6
<b>Hombre-SAVEE2 (JE)</b>	26,2	73,8
<b>Hombre-SAVEE3 (JK)</b>	10,9	89,6
<b>Hombre-SAVEE4 (KL)</b>	26,7	73,3
<b>Promedio Total</b>	24,3	75,83

Tabla 49: Resultados de clasificación según género en conjunto Hombre-Emo-DB.

Hombre-Emo-DB/Clase	Mujer (%)	Hombre (%)
<b>Hombre-Emo-DB1 (03)</b>	10,0	90,0
<b>Hombre-Emo-DB2 (10)</b>	14,1	85,9
<b>Hombre-Emo-DB3 (11)</b>	13,7	86,3
<b>Hombre-Emo-DB4 (12)</b>	8,5	91,5
<b>Hombre-Emo-DB5 (15)</b>	15,9	84,1
<b>Promedio Total</b>	12,44	87,56

Si comparamos los resultados de esta prueba, con respecto a la prueba anterior, los cambios son casi imperceptibles, algunos bajando o subiendo una o dos décimas. Pero en el grupo en que sí hubo cambios sustanciales fue el mismo caso de bajos resultados que se mencionaron en la prueba anterior (ver Tabla 41 y Tabla 47), donde los valores de exactitud para predecir el género cambiaron de 43,5% a 65,0% en un sujeto, y de 91,5% a 96,4% en

el otro sujeto de aquel conjunto de muestras. En esta prueba, se pudo ver que el volumen sí afecta de alguna manera el reconocimiento de género en un audio.

### ***A.3.8. Prueba 8-A: Otra forma para clasificar voz según género - Extracción de características más específicas***

Más adelante, en la sección A.5, se puede ver el análisis y los resultados de la clasificación de emoción, género e identidad a partir de la grabación y el posterior procesamiento de distintas conversaciones entre dos personas. Al momento de clasificar las muestras de acuerdo al género, los resultados fueron bastante más bajos con respecto a los obtenidos en los experimentos previos. Muchas de las muestras fueron mal clasificadas, en especial cuando el género era masculino. ¿Qué estuvo mal? ¿qué influyó para obtener una mala clasificación?

Dentro de las posibles causas que se identificaron, que llevaran a una mala clasificación de género, se encuentra el conjunto de características, duración del audio y el idioma. El conjunto de características es una pieza fundamental en lo que se refiere a clasificación, por lo que fue el primer elemento a analizar.

Se revisaron distintas investigaciones sobre clasificación de género a partir de la voz, las cuales describen las características de audio que mejor diferencian a un género de otro. En un principio, es esperable imaginar que el Pitch es uno de los rasgos principales que diferencian la voz femenina de la masculina. Generalmente el pitch de una voz femenina es más alto que el de una voz masculina, pero esto no siempre ocurre así, siendo a veces el Pitch de la voz de una mujer más bajo que el pitch de la voz de un hombre [11]. Una de las razones son las emociones. Los efectos de una emoción en el habla tienden a alterar algunos rasgos de la expresión acústica, como la calidad de la voz, duración, entre otros [13]. Entonces son otros los rasgos que determinan de mejor manera el género de una voz, al momento de clasificar pistas auditivas.

Antes de mencionar, los rasgos que mejor representan a un género mediante la voz, según las investigaciones analizadas, es importante recalcar lo dicho en [13]: la clasificación de género mediante la voz es independiente de la duración, de lo que se exprese, y de la identidad de la persona. Esto nos indica que la forma en que las pruebas se han hecho, ha sido la correcta, al asumir estas invariantes. A continuación, se describe en detalle cada una de las características mencionadas en los estudios analizados:

- Short-Time Energy (STE): Es el incremento repentino en la señal de energía. Su valor es más alto en voces femeninas que en masculinas [11].

- Zero Crossing Rate (ZCR): Es la medida del número de veces en un intervalo de tiempo (frame) en que la amplitud de la señal del habla pasa a través del eje cero. Su valor es más alto en voces femeninas que en masculinas [10][11][12].
- Centro de gravedad del espectro (CG): Es el centro de gravedad de un vector acústico. Para voces femeninas su valor es cercano a las altas frecuencias, y para las masculinas, cercano a las bajas frecuencias [12][10].
- Energy Entropy (EE): Es definida como los diferentes cambios repentinos en el nivel de energía de una señal del habla. En hombres, es baja y distribuida, mientras que en mujeres es alta y por un período corto [11].

## **Muestreo**

Para las características individuales, se extrajeron 40 archivos al azar de las muestras de entrenamiento para clasificación de género anterior, que incluye archivos de audio de bases de datos, y todas las muestras se encontraban con volumen normalizado. De los 40 archivos, 20 fueron pistas de audio de mujer y 20 de hombre.

Para cada conjunto de audios de hombre y de mujer, se crearon dos grupos, uno conteniendo los archivos de audio tal como se describieron, y el segundo conteniendo los mismos archivos de audio, pero se les había aplicado previamente un filtro pasa bajo, de 1 KHz. Esto se hizo para eliminar cualquier ruido de frecuencia mayor a 1 KHz, debido a que la voz humana, tanto para hombre como para mujer, se encuentra en rangos de frecuencia menor a 1 KHz.

Llamaremos MujerX, MujerX-Low, HombreX y HombreX-Low a los distintos conjuntos de muestras.

También se utilizaron las muestras de entrenamiento y prueba para la clasificación de género anterior, con el fin de generar un modelo usando todas las características.

## **Extracción de Características**

Se extrajeron características de dos formas: individualmente y en grupo. Se creó un archivo de configuración de openSMILE para cada característica individual y uno para el grupal. Los archivos de configuración para las características individuales generan un archivo en formato CSV y los datos de todos los archivos procesados se guardan en el mismo archivo de salida CSV (opción de append es igual a 1). El archivo CSV que se genera muestra los valores de los rasgos extraídos para cada frame.

La configuración para STE, CG y EE utiliza frames de 60 ms, con un paso de 10 ms. Para ZCR, los frames son de 10 ms y el paso también. En

este último caso se calculó el ZCR y MCR (Mean-crossing rate). MCR es la cantidad de veces que la señal cruza su promedio, para las señales con promedio 0 éste es idéntico al ZCR [5]. Se agregó MCR porque no todas las pistas de audio se encontraban perfectamente centradas, entonces el promedio no era 0. Se probó con SoX para centrar el audio (opción *dcshift*), pero no siempre se pudo hacer con 100% de precisión, por lo que se prefirió calcular el valor de MCR.

Por cada conjunto de muestras, para ambos géneros, se extrajeron las características correspondientes. Se generó un archivo CSV por cada conjunto, el cual contiene los datos correspondientes para todos los archivos de ese conjunto. Para STE, ZCR y MCR, el archivo CSV muestra un único valor por frame, lo que permite una mejor visualización, pero para CG y EE, muestra más de un valor, según la banda de frecuencia en que se analice.

En el archivo de configuración grupal se calculó cada una de las características (LLD), calculando su contorno suavizado (cContourSmoother) y finalmente para cada contorno se calcula su valor máximo, mínimo y promedio (Statistical Functionals).

### **Modelos y Entrenamiento**

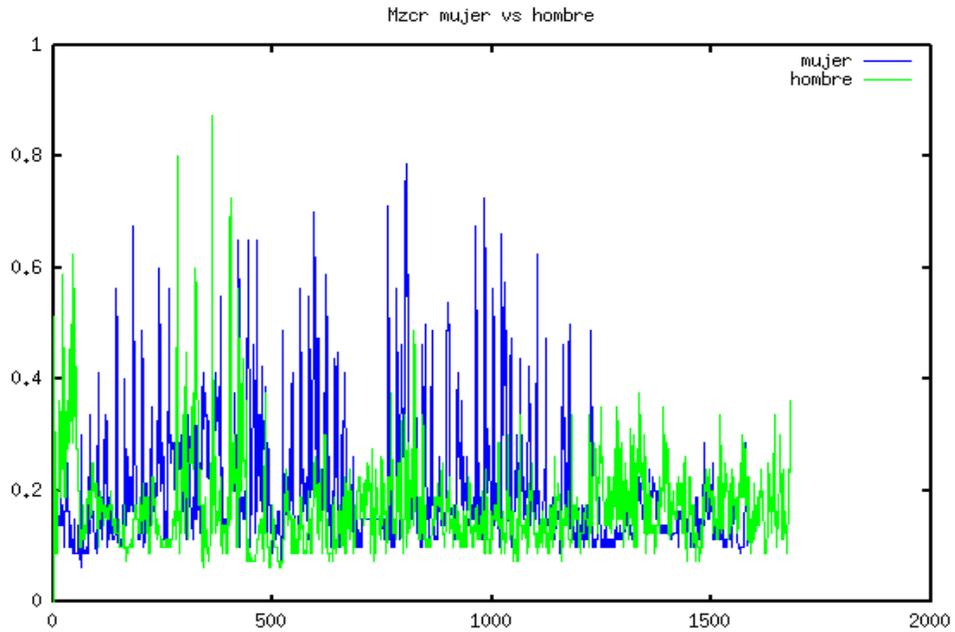
Se intentó generar un modelo para configuración grupal, utilizando las muestras de entrenamiento y prueba del punto anterior. Pero el modelo no logró ser procesado, el entrenador de LIBSVM no terminaba de realizar iteraciones, y mientras más iteraciones se hagan, significa que no logra separar una clase de otra según las características elegidas.

Se entrenó otro modelo, pero que esta vez sólo incluyera STE y MCR. Se utilizaron los mismos parámetros de entrenamiento. Este modelo sí logró generarse. Ambos fueron entrenados con el kernel lineal.

### **Resultados y Análisis**

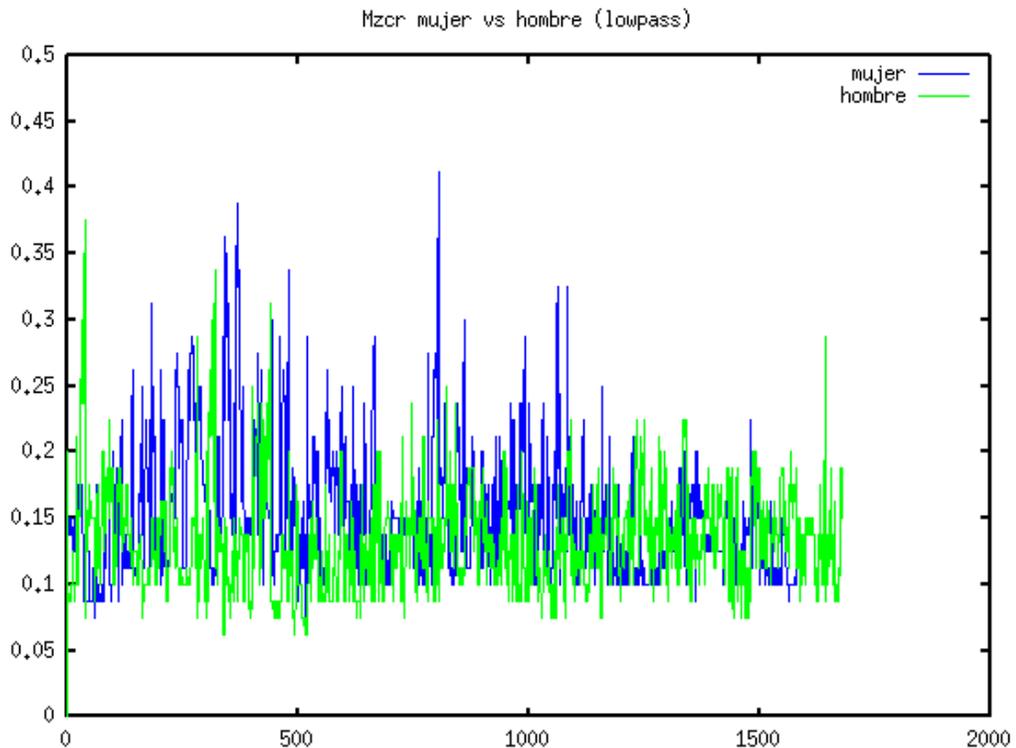
En la Ilustración 20 se muestran gráficamente los valores de MCR para las muestras de HombreX y MujerX. Se puede ver que hay una similitud en las muestras de mujeres, siendo valores más altos, y para los hombres valores más bajos, cumpliendo con la descripción de ZCR. Pero hay casos extremos en los valores de las voces masculinas son mucho más altos, incluso sobrepasando el máximo de MCR para las voces femeninas. Estos casos son precisamente muestras de audio que expresan la emoción de enojo.

Para las voces femeninas hay valores que caen por debajo del promedio, incluso más bajo que los valores de las voces masculinas. Aunque en estos casos son emociones neutrales, también hay otros casos de voces neutrales en que los valores sí se encuentran en rangos más altos, entonces no hay una clara influencia de la emoción.



*Ilustración 20: MCR para muestras de los conjuntos MujerX y HombreX*

Analizando los valores de MCR para las muestras de HombreX-Low y MujerX-Low, los resultados son bastante similares a los de MCR en HombreX y MujerX, no tanto por los valores (los que cambian de rangos), sino por la tendencia que ellos siguen. Pero también hay algunas excepciones, que son los peaks muy altos vistos para las voces masculinas anteriormente, los cuales se reducen un poco, pero para las voces femeninas no hay mayores cambios.



*Ilustración 21: MCR para muestras de los conjuntos MujerX-Low y HombreX-Low*

Si se analizan los valores para STE, ya sea en las muestras sin y con filtro pasa bajo (ver Ilustración 22 e Ilustración 23), los valores para hombres y mujeres son al revés de lo que deberían ser, los valores para hombres son más altos en general que los valores para las mujeres. Para esta característica se puede ver que para las mujeres hay unos valores fuera del promedio, y corresponden a muestras de emociones más fuertes, como el enojo.

El uso del filtro pasa bajo no generó demasiados cambios en la tendencia de las características, pero sí cambió sus valores, en mayor o menor medida. Para MCR los valores cayeron en un rango menor, mucho más cercano al cero. Al ser los valores más pequeños, esto podría afectar en la generación de los modelos, pero de una buena manera, haciendo más fáciles los cálculos (relacionado con el escalamiento visto en A.3.4). A pesar de su posible efecto positivo, se posterga a futuro un análisis más detallado de sus verdaderos efectos.

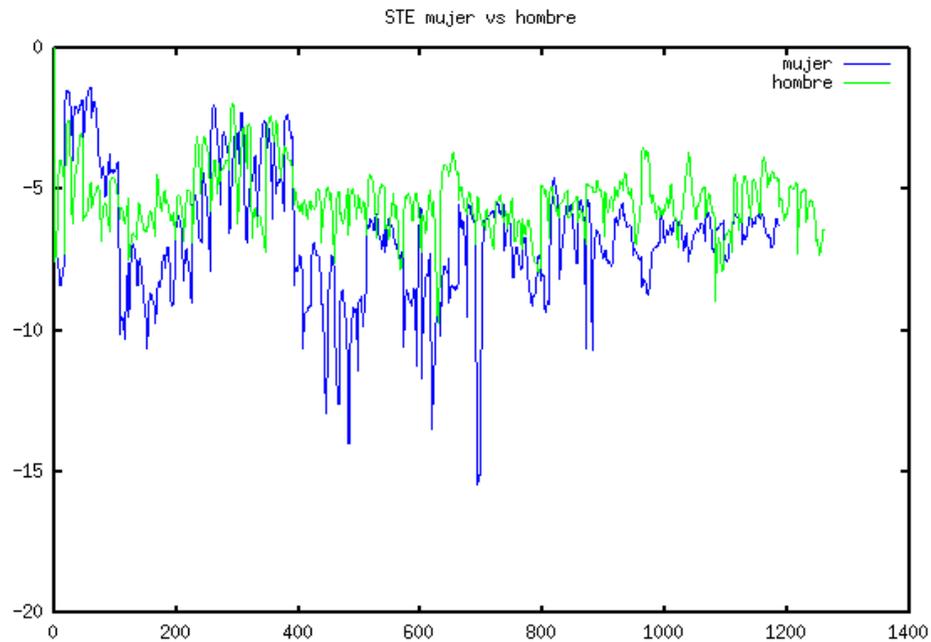


Ilustración 22: STE para muestras de los conjuntos MujerX y HombreX

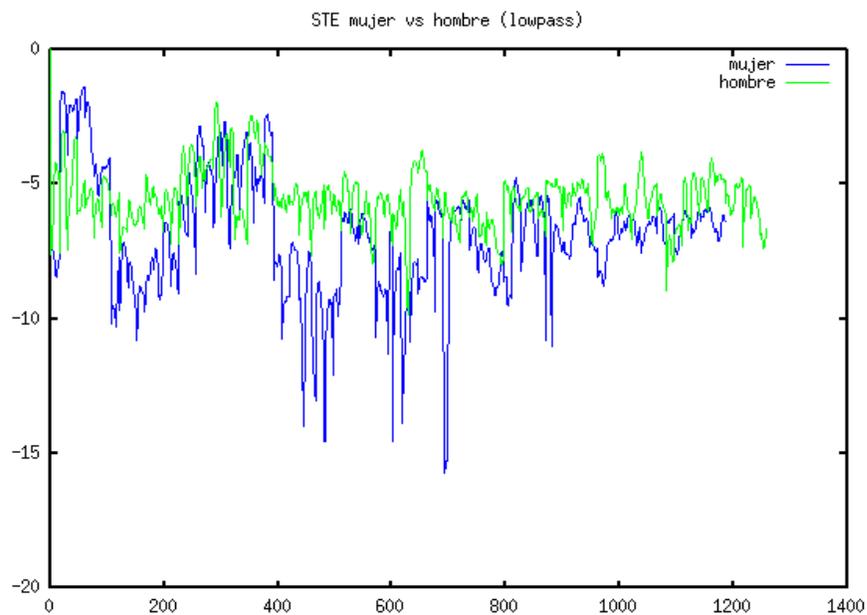


Ilustración 23: STE para muestras de los conjuntos MujerX-Low y HombreX-Low

Como los valores de CG y EE no son únicos por frame, eso los hace más difícil de analizar, y además de que los valores para STE y MCR ya mostraron no cumplir con lo que los estudios señalaban, entonces quizás la mezcla de ellos podría generar mejores resultados. Pero, ya se mencionó antes, que al entrenar el modelo grupal, éste no logró generarse, por lo que

se probó con uno que sólo tuviera los rasgos STE y MCR. A continuación se muestran los resultados al contrastar este modelo con los grandes grupos de prueba que se utilizaron para clasificación de género del punto anterior:

*Tabla 50: Resultados de clasificación según género en conjunto Mujer.*

<b>Mujer/Clase</b>	<b>Mujer (%)</b>	<b>Hombre (%)</b>
<b>Mujer1</b>	47,9	52,1
<b>Mujer2</b>	65,5	34,5
<b>Mujer3</b>	75,8	24,2
<b>Mujer4</b>	65,5	34,5
<b>Mujer5</b>	62,2	37,8
<b>Promedio Total</b>	63,38	36,62

*Tabla 51: Resultados de clasificación según género en conjunto Mujer-Emo-DB.*

<b>Mujer-Emo-DB/Clase</b>	<b>Mujer (%)</b>	<b>Hombre (%)</b>
<b>Mujer-Emo-DB1 (08)</b>	26,9	73,1
<b>Mujer-Emo-DB2 (09)</b>	42,7	57,3
<b>Mujer-Emo-DB3 (13)</b>	21,6	78,4
<b>Mujer-Emo-DB4 (14)</b>	20,4	79,6
<b>Mujer-Emo-DB5 (16)</b>	43,4	56,6
<b>Promedio Total</b>	31	69

Tabla 52: Resultados de clasificación según género en conjunto Hombre.

Hombre/Clase	Mujer (%)	Hombre (%)
Hombre1	60,8	39,2
Hombre2	42,0	58,0
Hombre3	43,1	56,9
Hombre4	45,2	54,8
Hombre5	33,8	66,2
<b>Promedio Total</b>	<b>44,98</b>	<b>55,02</b>

Tabla 53: Resultados de clasificación según género en conjunto Hombre-Otro.

Hombre-Otro/Clase	Mujer (%)	Hombre (%)
Hombre-Otro1	45,2	54,8
Hombre-Otro2	26,4	73,6
<b>Promedio Total</b>	<b>35,8</b>	<b>64,2</b>

Tabla 54: Resultados de clasificación según género en conjunto Hombre-SAVEE.

Hombre-SAVEE/Clase	Mujer (%)	Hombre (%)
Hombre-SAVEE1 (DC)	23,5	76,5
Hombre-SAVEE2 (JE)	26,5	73,5
Hombre-SAVEE3 (JK)	10,4	89,6
Hombre-SAVEE4 (KL)	82,3	17,7
<b>Promedio Total</b>	<b>35,68</b>	<b>64,33</b>

Tabla 55: Resultados de clasificación según género en conjunto Hombre-Emo-DB.

Hombre-Emo-DB/Clase	Mujer (%)	Hombre (%)
Hombre-Emo-DB1 (03)	15,1	84,9
Hombre-Emo-DB2 (10)	34,7	65,3
Hombre-Emo-DB3 (11)	28,5	71,5
Hombre-Emo-DB4 (12)	35,6	64,4
Hombre-Emo-DB5 (15)	50,1	49,9
<b>Promedio Total</b>	32,8	67,2

Los resultados muestran que no hay una clara clasificación de las voces según género. De los resultados obtenidos, hay un buen número de ellos que parecen ser buenos, pero el resto, como el caso de las voces femeninas de Emo-DB (ver Tabla 51), son contrarios a lo esperado.

Tanto para los resultados para características separadas, como en grupo, la influencia de las emociones en el habla es lo que estaría afectando en la clasificación. Lamentablemente estas características no dieron resultado, y si la duración del audio no es un problema al momento de clasificar, quedan pocas opciones.

#### **A.3.9. Prueba 8-B: Otra forma para clasificar voz según género - Clasificación de género según emociones**

Dentro de las estudios que se encontraron, también estaba la clasificación de género en bases de datos emocionales, en que se clasificaba el género en cada base de datos individualmente, no conjuntamente. En el estudio en que se hace esta clasificación, se describen las características del audio utilizadas (LLD) y sus respectivos valores estadísticos (Functionals), para cada base de datos de emociones [13]. Debido a que el propósito de la clasificación de género en este trabajo busca ser transversal al idioma, emoción u otro factor posiblemente influyente, se tomaron las distintas características que coincidían para cada una de las bases de datos utilizadas en el estudio mencionado. Al utilizar características que fueron probadas en emociones, podría dar mejores resultados.

#### **Muestreo**

Se utilizaron las mismas muestras de entrenamiento y prueba para la clasificación de género donde se normalizó el volumen (sección A.3.7).

## Extracción de Características

Como se dijo inicialmente, de todas las características que mejor funcionaron para la clasificación de género en las muestras con emociones según el estudio, se eligieron las que coincidían para las tres bases de datos utilizadas en aquel estudio. A continuación el detalle del vector de características:

Tabla 56: Características seleccionadas.

LLD	Statistical Functionals
<b>Formants</b>	Máximo del cuarto formante.
<b>Pitch</b>	Máximo, media, mediana, rango intercuartil de las duraciones de los mínimos del contorno suavizado.
	Máximo, media, rango intercuartil de los valores del pitch en los mínimos del contorno suavizado.
	Rango intercuartil, límite superior (90%) de las duraciones de los mínimos del contorno suavizado.
	Rango intercuartil de las duraciones de las pendientes ascendentes de los contornos del pitch.
<b>Intensity (Energy)</b>	Rango intercuartil de los máximos del contorno suavizado.
	Rango intercuartil de las duraciones de las pendientes ascendentes del contorno de la energía.
	Rango intercuartil de las duraciones de las pendientes descendentes del contorno de la energía.
<b>Spectral</b>	Energía en las bandas de frecuencia 250-600, 1000-1500, 2100-2800 Hz.
	Energía en las bandas de frecuencia 1000-2100, 1500-2800, 2100-2500 Hz.
	Energía en las bandas de frecuencia 1000-2800, 1500-3500, 2100-3950 Hz.
	Radio de Energía entre las bandas de frecuencia (3950-2100 Hz) y (2100-0 Hz) y entre las bandas de frecuencia (2100-1000 Hz) y (1000-0 Hz).

## Modelos y Entrenamiento

El modelo entrenado se basó en las características de la tabla anterior, y en el uso del kernel lineal. Hasta ahora se ha mantenido el uso de kernel lineal por ser el que ha dado mejores resultados.

## Resultados y Análisis

Tabla 57: Resultados de clasificación según género en conjunto Mujer.

Mujer/Clase	Mujer (%)	Hombre (%)
Mujer1	94,9	5,1
Mujer2	98,1	1,9
Mujer3	0,0	100,0
Mujer4	100,0	0,0
Mujer5	79,0	21,0
<b>Promedio Total</b>	<b>74,4</b>	<b>25,6</b>

Tabla 58: Resultados de clasificación según género en conjunto Mujer-Emo-DB.

Mujer-Emo-DB/Clase	Mujer (%)	Hombre (%)
Mujer-Emo-DB1 (08)	0,0	100,0
Mujer-Emo-DB2 (09)	22,8	77,2
Mujer-Emo-DB3 (13)	25,2	74,8
Mujer-Emo-DB4 (14)	10,4	89,6
Mujer-Emo-DB5 (16)	0,5	99,5
<b>Promedio Total</b>	<b>11,78</b>	<b>88,22</b>

Tabla 59: Resultados de clasificación según género en conjunto Hombre.

Hombre/Clase	Mujer (%)	Hombre (%)
Hombre1	0,0	100,0
Hombre2	100,0	0,0
Hombre3	7,1	92,9
Hombre4	99,9	0,1
Hombre5	100,0	0,0
<b>Promedio Total</b>	<b>61,4</b>	<b>38,6</b>

Tabla 60: Resultados de clasificación según género en conjunto Hombre-Otro.

Hombre-Otro/Clase	Mujer (%)	Hombre (%)
<b>Hombre-Otro1</b>	0,0	100,0
<b>Hombre-Otro2</b>	4,4	95,6
<b>Promedio Total</b>	2,2	97,8

Tabla 61: Resultados de clasificación según género en conjunto Hombre-SAVEE.

Hombre-SAVEE/Clase	Mujer (%)	Hombre (%)
<b>Hombre-SAVEE1 (DC)</b>	9,5	90,5
<b>Hombre-SAVEE2 (JE)</b>	0,0	100,0
<b>Hombre-SAVEE3 (JK)</b>	0,0	100,0
<b>Hombre-SAVEE4 (KL)</b>	1,5	98,5
<b>Promedio Total</b>	2,75	97,25

Tabla 62: Resultados de clasificación según género en conjunto Hombre-Emo-DB.

Hombre-Emo-DB/Clase	Mujer (%)	Hombre (%)
<b>Hombre-Emo-DB1 (03)</b>	0,4	99,6
<b>Hombre-Emo-DB2 (10)</b>	25,0	75,0
<b>Hombre-Emo-DB3 (11)</b>	0,0	100,0
<b>Hombre-Emo-DB4 (12)</b>	25,0	75,0
<b>Hombre-Emo-DB5 (15)</b>	17,1	82,9
<b>Promedio Total</b>	13,5	86,5

Nuevamente los resultados en la clasificación son confusos. Se puede notar que las mismas voces femeninas de Emo-DB vuelven a ser clasificadas más como hombre que como mujer. Son tales las características, que tienden a clasificar la mayoría de las muestras con mayor porcentaje en hombres que en mujeres, por lo que no son las mejores para clasificar según género en el conjunto de muestras elegido.

### **A.3.10. Prueba 9: Clasificación de género por idioma**

Después de haber probado distintas opciones para clasificar según género, el idioma es probablemente la última opción que queda por revisar. Las actuales muestras de audio están en Inglés, Alemán, Español de España, y Vasco. Con la idea de confirmar o descartar la influencia del idioma en la clasificación según género, se probó inicialmente con un idioma, el Alemán. Si los resultados eran buenos, se seguiría la prueba con los otros idiomas, en caso contrario, se descartaría la clasificación de esta manera.

#### **Muestreo**

Con el fin de tener una variedad de muestras de audio en alemán, y no sólo las que forman parte de Emo-DB, se buscó en YouTube<sup>10</sup> vídeos que estuvieran en este idioma, y que la gente en ellos expresara distintas emociones. La mayoría de los vídeos encontrados y que sirvieran a este propósito fueron vídeos sobre noticias, debates, entrevistas, reportajes, charlas y venta de productos.

Para descargar los videos y extraer su audio, se hizo uso del programa youtube-dl<sup>11</sup>, que además de permitir la descarga de vídeos de YouTube, también permite hacerlo de otras páginas web. Incluye la opción de elegir el formato de descarga, extracción del audio y el formato de éste, entre una amplia variedad de opciones.

Una vez obtenidos los archivos de audio de cada video (en formato WAV), se separaron de forma manual las voces femeninas y masculinas, mediante el programa Audacity. Se obtuvieron aproximadamente 10min con 10s de audio para cada género.

En las pruebas anteriores se usaron muestras de la base de datos Emo-DB para el entrenamiento del modelo, y esas mismas muestras fueron incluidas en el conjunto para esta sección. La cantidad de muestras eran alrededor de 70 por género. Aunque esta cantidad no es poca, siempre es ideal tener una mayor cantidad, para tener más datos y así poder clasificar mejor.

Entonces, en vez de buscar más vídeos en Alemán y extraer el audio de ellos, se extrajeron las palabras de las muestras que ya se tenían. Esto se refiere a que se toma una muestra, se detectan entre qué frames se dijo una palabra, y luego se corta ese rango de frames, obteniendo un nuevo archivo de audio, que contiene sólo una palabra. En la sección 3.4 se puede ver en detalle cómo se logra extraer las palabras o las frases de un audio.

Después de haber extraído las palabras de las muestras iniciales, el nuevo total de muestras quedó en aproximadamente 960 para cada género. Esto constituyó el grupo de entrenamiento del experimento. Para las pruebas

<sup>10</sup> <http://www.youtube.com>

<sup>11</sup> <http://youtube-dl.org/>

se utilizaron las mismas muestras de audio de Mujer-Emo-DB y Hombre-Emo-DB ya vistas anteriormente.

### Extracción de Características

Debido a que el último conjunto de características que dio buenos resultados fue *IS10\_paraling*, fue este mismo el que se utilizó en esta experimento.

### Modelos y Entrenamiento

El entrenamiento del modelo Alemán se hizo también con el kernel lineal.

### Resultados y Análisis

Los resultados que se muestran en las tablas posteriores corresponden a las muestras de Mujer-Emo-DB y Hombre-Emo-DB, las que son las muestras de emociones que están en Alemán.

Tabla 63: Resultados de clasificación según género en conjunto Mujer-Emo-DB.

Mujer-Emo-DB/Clase	Mujer (%)	Hombre (%)
<b>Mujer-Emo-DB1 (08)</b>	87,4	12,6
<b>Mujer-Emo-DB2 (09)</b>	88,3	11,7
<b>Mujer-Emo-DB3 (13)</b>	84,7	15,3
<b>Mujer-Emo-DB4 (14)</b>	91,8	8,2
<b>Mujer-Emo-DB5 (16)</b>	89,5	10,5
<b>Promedio Total</b>	88,34	11,66

Tabla 64: Resultados de clasificación según género en conjunto Hombre-Emo-DB.

Hombre-Emo-DB/Clase	Mujer (%)	Hombre (%)
<b>Hombre-Emo-DB1 (03)</b>	63,3	36,7
<b>Hombre-Emo-DB2 (10)</b>	55,5	44,5
<b>Hombre-Emo-DB3 (11)</b>	61,8	38,2
<b>Hombre-Emo-DB4 (12)</b>	70,8	29,2
<b>Hombre-Emo-DB5 (15)</b>	55,6	44,4
<b>Promedio Total</b>	61,4	38,6

En esta prueba la clasificación siguió una tendencia inversa a las dos pruebas anteriores. Esta vez tendió a clasificar las muestras en mayor medida como si fueran del género femenino, ya sea para voces femeninas o masculinas. Esto puede querer decir que las características escogidas tienden a enfatizar los rasgos de la voz de una mujer más que los de la voz de un hombre, pero no queda realmente clara su influencia en ello.

Debido a que los resultados obtenidos no fueron satisfactorios, se descartó continuar clasificando muestras para los otros idiomas. Por esta vez, se dejó de lado el idioma como variable dependiente para la clasificación de voz según género, quizás estudios posteriores podrían dar una mejor perspectiva. Siguiendo el propósito de este trabajo, conviene que sea una variable independiente.

### ***A.3.11. Prueba 10: Clasificación de género por emoción***

Después de una variedad de intentos por clasificar de forma correcta, aunque no sea 100% precisa, un audio en hombre o mujer, las opciones ya eran casi nulas. Hasta ese momento, el enfoque era clasificar una pista de audio según género, sin importar la emoción y el idioma (aunque con éste último igual se probó para confirmar o descartar su influencia). Se pudo ver que de alguna manera la emoción generaba distorsiones en las características analizadas (sección A.3.6 y A.3.8). Con esto surgió una idea o pregunta: ¿Por qué no clasificar primero el audio según su emoción y luego de haber identificado la emoción hacer la clasificación por género?

#### **Muestreo**

Al cambiar el enfoque, esta vez las muestras para entrenamiento y prueba debían ser diferentes, debían ser en su totalidad muestras en que se expresara una misma emoción. Más adelante, en la sección A.4, se describen en detalle las numerosas pruebas para lograr clasificar correctamente las distintas emociones con las que se contó, y principalmente lograr identificar la tristeza. Entonces, las mismas muestras utilizadas en las pruebas de clasificación de emociones se utilizaron en esta prueba.

Además de dividir las muestras por emoción, éstas también fueron separadas por género. Y con el fin de reafirmar el efecto positivo de la normalización del volumen de las muestras, se crearon dos conjuntos, uno con todas las muestras de emociones sin normalizar, y otro con todas las muestras normalizadas. A estos conjuntos los llamaremos Emo y EmoNorm, respectivamente.

Es importante destacar que no existe una uniformidad en la cantidad de muestras para entrenamiento y prueba. Esto se debe a que las bases de datos no vienen con una misma cantidad de muestras de hombre y de mujer, en algunos casos conteniendo sólo muestras de un género para una emoción.

## Extracción de Características

Nuevamente el vector de características utilizado es *IS10\_paraling*, que es el que mejores resultados ha entregado en las pruebas de reconocimiento de género.

## Modelos y Entrenamiento

El modelo fue entrenado con el kernel lineal.

## Resultados y Análisis

Las siguientes tablas muestran los resultados al clasificar las muestras de emociones según género, tanto para el conjunto Emo como para EmoNorm. Los valores en porcentaje de la columna Mujer y Hombre representan la exactitud promedio con que las muestras de mujer o de hombre de cada emoción fueron correctamente clasificadas. Las columnas Mujer (VP) y Hombre (VP) muestran la cantidad Verdaderos Positivos con respecto al total de muestras en cada caso.

Tabla 65: Exactitud de clasificación y Verdaderos Positivos en el conjunto Emo.

Emoción	Mujer (%)	Hombre (%)	Mujer (VP)	Hombre (VP)
<b>Enojo</b>	83,7	80,9	32/35	50/59
<b>Miedo</b>	72,4	90,3	15/18	48/48
<b>Felicidad</b>	80,5	72,1	12/14	41/52
<b>Neutral</b>	63,8	86,8	14/20	80/80
<b>Tristeza</b>	74,9	83,8	16/19	40/41
<b>Sorpresa</b>	51,9	93,1	1/2	30/30

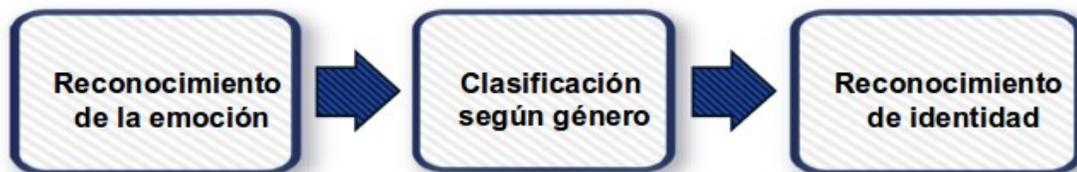
Tabla 66: Exactitud de clasificación y Verdaderos Positivos en el conjunto EmoNorm.

Emoción	Mujer (%)	Hombre (%)	Mujer (VP)	Hombre (VP)
<b>Enojo</b>	79,0	76,7	30/35	47/59
<b>Miedo</b>	80,2	82,1	17/18	42/48
<b>Felicidad</b>	83,4	73,7	11/14	41/52
<b>Neutral</b>	64,5	85,3	15/20	80/80
<b>Tristeza</b>	74,9	85,1	17/19	41/41
<b>Sorpresa</b>	54,6	91,8	1/2	30/30

El reconocimiento de género en esta prueba parece entregar buenos resultados. Al comparar los resultados para Emo y EmoNorm, es posible detectar que para emociones más suaves, más calmadas, conviene normalizar antes de reconocer género. Para emociones más fuertes, más agitadas, no funciona tanto y aumenta la probabilidad de que una muestra se reconozca como mujer, sin importar el verdadero género. En el caso de la Sorpresa, es difícil ver con exactitud el comportamiento, ya que son muy pocas muestras de prueba y entrenamiento por género.

Si se establece que para las emociones Neutral, Tristeza y quizás Miedo, se deben normalizar las pistas de audio antes de clasificar según género, y que para el resto de las emociones no se necesiten normalizar, se tendría un buen método para reconocer género en emociones.

Esta es la última prueba realizada en el intento de reconocer género correctamente. Por ahora quedará como la solución a esta problemática. Como el enfoque cambió, el flujo del proceso cambió a:



*Ilustración 24: Nuevo enfoque del proceso de reconocimiento.*

#### **A.4. Pruebas con Bases de Datos de Emociones**

Como se mencionó en la sección A.2, para poder diferenciar una emoción de otra, la naturalidad o realismo de la grabación es un factor fundamental para una detección confiable y con un alto grado de exactitud. Para este trabajo fue posible acceder al contenido de tres bases de datos de emociones, que se acercan un poco más a la realidad, dado que algunas de las grabaciones fueron hechas por actores profesionales, pero aún no son del todo reales.

Las bases de datos fueron conseguidas ya sea por descarga directa de la web, o firmando una petición para poder acceder a ellas. Estas tres son:

- SAVEE (Surrey Audio-Visual Expressed Emotion) [14]: Base de datos creada en la Universidad de Surrey, con el propósito de detectar emociones de forma multimodal (audio y video). La sección de audio está compuesta por grabaciones de 4 hombres de habla inglesa (identificados como DC, JE, JK, KL), de entre 27 y 31 años. Estos usuarios eran estudiantes e investigadores de la universidad. Por cada hombre se grabaron frases con y sin sentido semántico, con respecto a la emoción. El conjunto de grabaciones se divide en 7 emociones:

Alegría, Enojo, Asco, Miedo, Neutral, Sorpresa y Tristeza. En total son 120 archivos de audio por usuario, 15 por cada emoción, exceptuando Neutral, la cual tiene 30.

- Emo-DB (Berlin Database of Emotional Speech) [15]: Forma parte de una investigación de la Universidad Técnica de Berlín, entre los años 1997 y 1999. Esta base de datos cuenta con grabaciones de 10 actores, 5 hombres y 5 mujeres, todos hablantes de Alemán, y a cada uno de ellos le corresponde un número entre el 3 y el 16. Por cada actor se almacenaron grabaciones de frases sin sentido semántico y en 7 emociones diferentes (Alegría, Enojo, Asco, Aburrimiento, Neutral, Miedo y Tristeza). La cantidad de archivos de audio por emoción y por usuario no es igual para todos. La cantidad total ronda entre 35 y 60 archivos para cada usuario.
- RekEmozio [16]: Base de datos de emociones desarrollada en la Universidad del País Vasco, la cual también fue diseñada para la detección de emociones de forma multimodal (análisis de la voz y la expresión facial). La base de datos completa tiene grabaciones de audio en Castellano (o Español de España) y en Vasco, realizadas tanto por actores profesionales como amateur. Actualmente no se encuentra disponible para todo público, pero fue posible conseguir un extracto de ella. Este extracto o sub-base de datos incluye archivos de audio y video de dos actores, Actor006 (hombre que habla en Vasco) y Actor009 (mujer que habla en Español). Contempla las mismas emociones que la base de datos SAVEE. La sección de audio contiene palabras, frases y párrafos, con y sin sentido semántico. Para las pruebas se utilizaron los archivos de audio de frases. El total de pistas de audio de frases son 5 por cada emoción, para cada usuario.

#### ***A.4.1. Prueba 1: Clasificación de cada emoción***

La primera prueba que se hizo para clasificar emociones fue la más básica de todas. Primero se separaron todas las emociones, en caso de que no lo estuvieran, para cada base de datos, y se dividió cada emoción en dos grupos, entrenamiento y testeo.

##### **Muestreo**

De la base de datos SAVEE, por cada usuario, se eligieron los archivos de numeración impar para entrenamiento y los de numeración par para testeo. Para RekEmozio y Emo-DB la elección fue similar, aunque sus archivos no están numerados, sí están ordenados alfabéticamente, entonces se eligieron según su posición, los impares para entrenamiento y los pares para testeo. Este sencillo tipo de elección asegura que la cantidad de archivos de entrenamiento sea igual o uno más que los de prueba. Siempre es ideal que la cantidad de archivos de entrenamiento sea grande, algo que no se logra mucho por no contar con tantos archivos.

En total, se cuenta con audio de ocho emociones, pero se dejó el Aburrimiento de lado, por estar presente solo en una de las bases de datos. En resumen, se crearon grupos para las siete emociones, alegría, asco, miedo, enojo, neutral, tristeza y sorpresa, para cada una de las bases de datos, excepto para Emo-DB, la que no cuenta con grabaciones de sorpresa, por lo que se crearon sólo seis grupos (no se incluyó Aburrimiento).

### Extracción de Características

El conjunto de características escogido para esta prueba fue *emobase2010*. Ya fue utilizado anteriormente en las pruebas de reconocimiento de género, y se observó que para ese tipo de pruebas no era el más apto, pero sí lo sería para las pruebas de esta sección, ya que se enfocaría más en las características emocionales de la voz.

### Modelos y Entrenamiento

Se entrenaron tres tipos de modelos, uno con el kernel lineal, y los otros con kernel polinomial (el que dio mejores resultados en SAVEE), uno de grado 2 y el otro de grado 3. Para los polinomiales se probaron distintos valores de gamma y C, para verificar si había alguna influencia al utilizar valores grandes o pequeños de estos parámetros. Los diversos valores elegidos fueron gamma:1 C:1, gamma:1 C:1000, gamma:1000 C:1, gamma:0,0001 C:1000, gamma:1000 C:0,0001.

### Resultados y Análisis

En las siguientes tablas se plasman los primeros resultados de clasificación de emociones, correspondientes a una clasificación individual de éstas:

Tabla 67: Clasificación de emociones en SAVEE, usando modelo lineal.

Emoción/ Clase	Enojo (%)	Asco (%)	Miedo (%)	Felicidad (%)	Neutral (%)	Tristeza (%)	Sorpresa (%)
<b>Enojo</b>	33,2	11,1	9,9	20,0	8,7	9,7	7,5
<b>Asco</b>	12,4	23,1	6,9	9,9	23,9	16,9	6,9
<b>Miedo</b>	7,8	7,5	28,6	19,8	6,0	13,6	16,7
<b>Felicidad</b>	19,8	9,0	13,2	34,7	3,9	7,9	11,5
<b>Neutral</b>	5,1	12,3	3,3	3,5	57,3	15,4	3,0
<b>Tristeza</b>	5,6	14,8	11,5	6,9	23,6	29,1	8,5
<b>Sorpresa</b>	4,8	6,6	27,6	19,2	2,0	6,7	33,1

Tabla 68: Clasificación de emociones en Emo-DB, usando modelo lineal.

Emoción/ Clase	Enojo (%)	Asco (%)	Miedo (%)	Felicidad (%)	Neutral (%)	Tristeza (%)
<b>Enojo</b>	62,5	4,8	5,4	22,7	3,7	0,9
<b>Asco</b>	16,7	41,0	9,7	10,1	15,1	7,4
<b>Miedo</b>	13,5	10,0	40,4	10,2	18,0	7,9
<b>Felicidad</b>	40,0	6,3	9,7	37,0	5,6	1,4
<b>Neutral</b>	6,5	14,3	14,2	5,3	48,3	11,4
<b>Tristeza</b>	2,7	7,9	10,5	2,8	14,9	61,2

Tabla 69: Clasificación de emociones en RekEmozio, usando modelo lineal.

Emoción/ Clase	Enojo (%)	Asco (%)	Miedo (%)	Felicidad (%)	Neutral (%)	Tristeza (%)	Sorpresa (%)
<b>Enojo</b>	18,0	12,2	17,5	21,4	5,5	6,7	18,8
<b>Asco</b>	11,3	20,7	21,0	14,6	10,1	11,0	11,3
<b>Miedo</b>	9,5	10,0	31,2	13,5	14,8	10,7	10,3
<b>Felicidad</b>	21,3	8,0	16,4	28,9	11,7	4,2	9,5
<b>Neutral</b>	7,2	8,2	31,9	11,1	27,5	6,8	7,2
<b>Tristeza</b>	5,7	10,0	28,8	6,7	26,4	15,3	7,0
<b>Sorpresa</b>	27,9	21,9	10,8	18,0	3,1	4,9	13,5

Como es posible apreciar, usando el kernel lineal, la clasificación de emociones es bastante pobre, aunque para Emo-DB los valores son un poco más altos, por ser menos emociones, lo que facilita un poco la clasificación (ver Tabla 68). En general se puede ver que el bajo porcentaje de exactitud de clasificación se debe a la confusión con otras emociones. Por ejemplo, en SAVEE el Enojo y la Felicidad tienden a confundirse, como también el Asco con Neutral (aunque no al revés). Tristeza y Neutral también tienen valores similares de clasificación. Esta situación de confusión se repite en las otras bases de datos. Un caso particular es el Asco, que se confunde con dos o más emociones, y el Miedo le sigue con un comportamiento similar. Entonces, ¿qué se puede hacer para evitar estas confusiones, que sólo generan más falsos positivos?

Se probó con el modelo de kernel polinomial esperando mejores resultados, es decir, una mejor separación de las emociones. Se probó aun sabiendo que el kernel lineal es más útil para grandes conjuntos de características como *emobase2010*, además que es el que utilizó en la investigación la base de datos SAVEE. Para evitar probar tres veces cada nuevo modelo, se decidió sólo probar en una de las bases de datos, en este caso se probó en SAVEE, la cual ya mostró tener el problema de la confusión mencionado anteriormente.

Tabla 70: Clasificación de emociones en SAVEE, usando modelo polinomial grado 2.

<b>Emoción/ Clase</b>	<b>Enojo (%)</b>	<b>Asco (%)</b>	<b>Miedo (%)</b>	<b>Felicidad (%)</b>	<b>Neutral (%)</b>	<b>Tristeza (%)</b>	<b>Sorpresa (%)</b>
<b>Enojo</b>	26,4	11,3	10,4	21,3	9,1	12,4	8,9
<b>Asco</b>	12,8	19,6	7,9	9,9	8,4	16,6	8,4
<b>Miedo</b>	9,3	8,7	22,1	17,5	6,3	17,1	19,1
<b>Felicidad</b>	17,9	8,3	10,1	36,0	4,6	11,1	12,1
<b>Neutral</b>	7,9	13,4	5,0	3,4	51,5	14,5	4,4
<b>Tristeza</b>	9,1	15,9	11,4	6,2	23,4	25,9	8,1
<b>Sorpresa</b>	6,2	5,8	25,9	17,2	2,5	10,1	32,4

Como se mencionó en la descripción de los modelos, se entrenaron modelos de kernel polinomial de grado 2, variando el valor de gamma y C. Al hacer el contraste de los modelos con las muestras de prueba, los resultados fueron iguales para todos estos modelos, sin importar el valor de gamma y C. En la Tabla 70 se puede notar que los valores de clasificación de cada emoción consigo misma disminuyeron en uno o más puntos porcentuales, exceptuando Felicidad. En general los valores de las emociones distintas a la contrastada aumentaron y la propia disminuyó. Esto descarta el uso de kernel polinomial de grado 2 para la clasificación de emociones.

Una posterior prueba del kernel polinomial de grado 3 se hizo, obteniendo resultados más bajos, y a la vez más parejos, los cuales se mantuvieron sin importar la variación de gamma y C. La siguiente tabla refleja estos resultados:

Tabla 71: Clasificación de emociones en SAVEE, usando modelo polinomial grado 3.

Emoción/ Clase	Enojo (%)	Asco (%)	Miedo (%)	Felicidad (%)	Neutral (%)	Tristeza (%)	Sorpresa (%)
<b>Enojo</b>	19,4	14,0	12,4	19,0	11,2	13,9	10,1
<b>Asco</b>	13,0	16,2	10,1	9,8	25,9	15,3	9,7
<b>Miedo</b>	11,2	11,0	18,8	14,7	7,7	17,6	18,9
<b>Felicidad</b>	19,5	12,8	11,5	26,3	5,5	12,7	11,6
<b>Neutral</b>	10,3	13,5	7,7	5,4	43,5	13,1	6,6
<b>Tristeza</b>	11,4	15,5	12,2	7,6	24,1	20,6	8,6
<b>Sorpresa</b>	9,4	9,3	22,3	14,5	3,5	12,6	28,5

Como consecuencia de estas pruebas, convence casi completamente el uso de modelo de kernel lineal para la clasificación de muestras, en especial de emociones.

#### **A.4.2. Prueba 2-A: Clasificación de emociones por grupo - Grupo1**

En la prueba anterior se identificó una dificultad al momento de clasificar las emociones, esa dificultad fue que se confunden unas emociones con otras, sin dar una clara separación. Revisando la clasificación de cada una de las emociones en su respectiva base de datos, fue posible identificar dos grupos mayores en que las emociones tienden a confundirse más: Enojo-Alegría-Sorpresa-Miedo (Grupo1) y Neutral-Tristeza-Asco (Grupo2). Estos grupos parecieran dividir las emociones en agitadas y calmadas, respectivamente.

El dividir las emociones en grupo, permitiría primero seleccionarlas según su categoría, y luego que cae en su categoría correspondiente, se identifica la emoción correcta. Al tener menos emociones por grupo, aun confundiéndose entre ellas, debería ser más fácil separar una de otra, al ser menor la cantidad de factores que puedan producir ruido al momento de clasificar, como lo es la gran cantidad de clases. Esto también ayudaría a reducir los falsos positivos.

Entonces, el foco cambió para intentar separar de la mejor manera las emociones en sus respectivos grupos, para luego clasificarlas individualmente dentro de sus grupos.

¿Qué otra importancia extra tiene esta separación grupal? No hay que olvidar que uno de los propósitos principales de este trabajo es detectar emociones que indiquen un estado vulnerable. Al separar en grupos, se

pueden separar de mejor manera las emociones vulnerables de las no vulnerables. La tristeza caería en el grupo de las vulnerables.

En esta prueba se quiso conocer cómo afecta esta división en grupos a cada base de datos, no sólo contrastando el modelo creado contra su conjunto de prueba, sino también contrastando con los conjuntos de prueba de las otras bases de datos. ¿Y por qué esto? Porque lo ideal de este trabajo, es que se pueda clasificar una emoción sin importar el género, idioma o identidad de la persona que la expresó. Por lo tanto, si logramos obtener buenos resultados en los contrastes cruzados, podríamos tener un modelo de clasificación que sea transversal a las características recién mencionadas, especialmente el idioma. Claramente esto podría tener problemas con ciertos idiomas, pero se espera que con los idiomas más comúnmente hablados no lo sea tanto, por tener ellos algunas raíces u otros rasgos en común.

### **Muestreo**

En cada una de las bases de datos, tanto el conjunto de entrenamiento como el de prueba fueron divididos en los dos grupos recién mencionados. Como no hay muestras de sorpresa en Emo-DB, el Grupo1 sólo cuenta con Enojo, Alegría y Miedo.

### **Extracción de Características**

Se eligió probar con el vector *emobase2010* primeramente en esta división de grupos, y ver cuál era su comportamiento.

### **Modelos y Entrenamiento**

Para cada base de datos se creó un modelo, de dos clases, Grupo1 y Grupo2. El kernel elegido fue el lineal.

### **Resultados y Análisis**

Con el fin de reflejar el contraste de cada grupo (1 y 2) de cada base de datos, con todas las base de datos, se separaron los datos obtenidos de la siguiente manera: de una base de datos, se contrastan sus dos grupos del conjunto de prueba, contra cada una de la bases de datos. Se muestra la exactitud del modelo de clasificación para esos grupos, y la cantidad de verdaderos positivos para cada grupo, descrito de la forma VP/Total de muestras a clasificar.

Tabla 72: Clasificación de grupos emociones en SAVEE, contrastado con todas las bases de datos emocionales.

	SAVEE			Emo-DB			RekEmozio		
Grupo-SAVEE/ Clase	Grupo 1 (%)	Grupo 2 (%)	VP	Grupo 1 (%)	Grupo 2 (%)	VP	Grupo 1 (%)	Grupo 2 (%)	VP
<b>Grupo 1</b>	89,3	10,7	106/112	27,6	72,4	32/112	90,5	9,5	104/112
<b>Grupo 2</b>	14,2	85,8	105/116	4,1	95,9	113/116	69,7	30,3	26/116

Tabla 73: Clasificación de grupos emociones en Emo-DB, contrastado con todas las bases de datos emocionales.

	SAVEE			Emo-DB			RekEmozio		
Grupo-Emo-DB / Clase	Grupo 1 (%)	Grupo 2 (%)	VP	Grupo 1 (%)	Grupo 2 (%)	VP	Grupo 1 (%)	Grupo 2 (%)	VP
<b>Grupo1</b>	80,4	19,6	109/130	70,1	29,9	96/130	61,0	39,0	87/130
<b>Grupo2</b>	26,4	73,6	69/87	3,9	96,1	86/87	56,2	43,8	39/87

Tabla 74: Clasificación de grupos emociones en RekEmozio, contrastado con todas las bases de datos emocionales.

	SAVEE			Emo-DB			RekEmozio		
Grupo-Rek/ Clase	Grupo 1 (%)	Grupo 2 (%)	VP	Grupo 1 (%)	Grupo 2 (%)	VP	Grupo 1 (%)	Grupo 2 (%)	VP
<b>Grupo 1</b>	47,3	52,7	8/16	91,8	8,2	15/16	83,1	16,9	15/16
<b>Grupo 2</b>	9,3	90,7	12/12	49,2	50,8	7/12	39,1	60,9	7/12

Los valores que muestran estas tablas indican en general una buena clasificación de los grupos de una base de datos consigo misma, siendo un

poco más baja en el Grupo2 de RekEmozio. En el contraste de bases de datos distintas los resultados fueron menos satisfactorios. Al revisar la lista de los falsos positivos, se detectaron específicamente dos emociones que generaron estos resultados, miedo y asco, especialmente el asco.

Para ver con exactitud un caso donde los resultados no son tan buenos, basta con mirar la Tabla 72, correspondiente a los resultados de la base de datos SAVEE. Al contrastar con Emo-DB, el Grupo1, tiende a identificarse más como Grupo2 de Emo-DB. En Emo-DB el asco se confunde en mayor medida con el enojo, pero en SAVEE se confunde más con neutral. Entonces, las muestras que se parecen al enojo en SAVEE, serán identificadas como similares al asco según Emo-DB, y por lo tanto caerán en el Grupo2.

Por el análisis hecho, se concluye que la división de las emociones en estos dos grupos puede no ser la mejor, dado que hay emociones que para una base de datos caen mejor en un grupo, pero para otra base de datos, caen mejor en el otro. Esto disminuye las probabilidades de lograr una buena clasificación transversal.

#### **A.4.3. Prueba 2-B: Clasificación de emociones por grupo - Grupo2**

Se observó en la prueba anterior que una de las emociones que más genera complicaciones al momento de clasificar es el Asco. Analizando la situación, esta emoción no es muy común durante nuestra vida diaria, asociándose principalmente a comida o a algo que vemos en televisión o escuchamos de otra persona, pero no es una emoción que se exprese todos los días. Además no es una emoción que podamos definir realmente como una emoción vulnerable, que son las emociones que se quieren encontrar específicamente en este trabajo. Por estos motivos es que se decidió excluir de la clasificación, y así obtener una mejor separación de los grupos, siendo esto lo que se quiso verificar en esta prueba.

Al eliminar el asco de la clasificación grupal, los dos grupos que quedan son Enojo-Alegría-Sorpresa-Miedo (Grupo1) y Neutral-Tristeza (Grupo2).

#### **Muestreo**

Las muestras para entrenamiento y prueba siguen siendo las mismas que las de la prueba anterior, pero esta vez sin incluir el asco.

#### **Extracción de Características**

Se mantuvo el uso del vector *emobase2010*, al obtener una buena clasificación de una base de datos contra si misma.

#### **Modelos y Entrenamiento**

También se mantuvo el uso del kernel lineal.

## Resultados y Análisis

La estructura de las tablas es la misma que la de las tablas de la prueba anterior. Para cada base de datos, se contrastan sus grupos contra el modelo de todas las bases de datos, mostrando la exactitud de clasificación y la cantidad de verdaderos positivos detectados.

Tabla 75: Clasificación de grupos emociones en SAVEE, contrastado con todas las bases de datos emocionales.

Grupo-SAVEE/ Clase	SAVEE			Emo-DB			RekEmozio		
	Grupo 1 (%)	Grupo 2 (%)	VP	Grupo 1 (%)	Grupo 2 (%)	VP	Grupo 1 (%)	Grupo 2 (%)	VP
<b>Grupo 1</b>	89,4	10,6	104/112	39,6	60,4	45/112	94,2	5,8	112/112
<b>Grupo 2</b>	12,6	87,4	83/88	1,1	98,9	88/88	83,7	16,3	5/88

Tabla 76: Clasificación de grupos emociones en Emo-DB, contrastado con todas las bases de datos emocionales.

Grupo-Emo-DB/ Clase	SAVEE			Emo-DB			RekEmozio		
	Grupo 1 (%)	Grupo 2 (%)	VP	Grupo 1 (%)	Grupo 2 (%)	VP	Grupo 1 (%)	Grupo 2 (%)	VP
<b>Grupo1</b>	96,8	3,2	127/130	92,7	7,3	123/130	88,7	11,3	125/130
<b>Grupo2</b>	60,4	39,6	25/64	11,5	88,5	61/64	90,1	9,9	0/64

Tabla 77: Clasificación de grupos emociones en RekEmozio, contrastado con todas las bases de datos emocionales.

Grupo-Rek/ Clase	SAVEE			Emo-DB			RekEmozio		
	Grupo 1 (%)	Grupo 2 (%)	VP	Grupo 1 (%)	Grupo 2 (%)	VP	Grupo 1 (%)	Grupo 2 (%)	VP
Grupo 1	86,4	13,6	14/16	89,2	10,8	14/16	81,8	18,2	16/16
Grupo 2	68,8	31,2	1/8	49,8	50,2	5/8	49,0	51,0	4/8

La exclusión del asco de los conjuntos de emociones logró acrecentar la exactitud de clasificación en el Grupo1, y a su vez, aumentó la cantidad de verdaderos positivos en este grupo. Pero la clasificación del Grupo2 empeoró sus resultados en algunos casos, en especial para la base de datos RekEmozio, al contrastar ésta con las otras bases de datos, y viceversa.

La posible causa de la mala clasificación en el Grupo2, puede deberse a la confusión que existe entre la emoción Miedo y Neutral, y Miedo y Tristeza en las distintas bases de datos, ya identificada en la prueba anterior. Esta confusión genera que muestras de Neutral y Tristeza tiendan a ser clasificadas como del Grupo1, porque se parecen a las muestras de Miedo. Esto ocurre principalmente al contrastar bases de datos distintas.

#### A.4.4. Prueba 2-C: Clasificación de emociones por grupo - Grupo3

Con la confusión de emociones descrita en la prueba previa, se tomó una nueva decisión. Esta vez se separaron las emociones en tres grupos: Enojo-Alegría-Sorpresa (Grupo1), Neutral-Tristeza (Grupo2), Miedo (Grupo3). Esta nueva separación grupal podría solucionar el problema de las confusiones.

#### Muestreo

Las muestras de entrenamiento separaron en los tres grupos recién descritos, es decir, se extrajeron las muestras de Miedo que antes estaban en el Grupo1 y ahora formaron su propio grupo. Para tener una visión más amplia y real de lo que ocurre durante la clasificación, las muestras de prueba se desagruparon, quedando todas juntas.

#### Extracción de Características

El vector de características fue *emobase2010*.

## Modelos y Entrenamiento

El nuevo modelo de tres grupos se entrenó con el kernel lineal.

### Resultados y Análisis

Para visualizar mejor los resultados, y por la variación de los grupos en las muestras de prueba, la composición de las tablas cambió un poco. Para cada base de datos, se contrastan todas las muestras de prueba (se le llamó *Todo*), con los modelos de todas las bases de datos. Como resultado se muestran las cantidades de verdaderos positivos al contrastar con cada base de datos y la relación con la cantidad de muestras que deberían haberse clasificado correctamente. Esta información se expresa de como: *VP/Total clasificados en grupo/Total de muestras que deberían caer en ese grupo*.

En esta ocasión no se incluyó la exactitud de clasificación. Se revisaron los resultados anteriores, y la exactitud de clasificación coincide cercanamente al porcentaje de VP, por lo que para tener una rápida visión de los resultados sólo se dejó VP como información en las tablas.

Tabla 78: Clasificación en tres grupos de emociones para SAVEE, contrastado con todas las bases de datos emocionales.

Grupo-SAVEE/ Clase	SAVEE			Emo-DB			RekEmozio		
	G1 VP	G2 VP	G3 VP	G1 VP	G2 VP	G3 VP	G1 VP	G2 VP	G3 VP
<b>Todo</b>	78/ 94/ 84	82/ 91/ 88	7/ 15/ 28	65/ 99/ 84	71/ 101/ 88	0/ 0/ 28	84/ 200/ 84	0/ 0/ 88	0/ 0/ 28

Tabla 79: Clasificación en tres grupos de emociones para Emo-DB, contrastado con todas las bases de datos emocionales.

Grupo-Emo-DB/ Clase	SAVEE			Emo-DB			RekEmozio		
	G1 VP	G2 VP	G3 VP	G1 VP	G2 VP	G3 VP	G1 VP	G2 VP	G3 VP
<b>Todo</b>	58/ 119/ 96	32/ 39/ 64	7/ 36/ 34	92/ 97/ 96	61/ 69/ 64	22/ 28/ 34	96/ 189/ 96	1/ 4/ 64	0/ 0/ 34

Tabla 80: Clasificación en tres grupos de emociones para RekEmozio, contrastado con todas las bases de datos emocionales.

Grupo- Rek/ Clase	SAVEE			Emo-DB			RekEmozio		
	G1 VP	G2 VP	G3 VP	G1 VP	G2 VP	G3 VP	G1 VP	G2 VP	G3 VP
<b>Todo</b>	10/ 14/ 12	7/ 10/ 8	0/ 0/ 4	2/ 2/ 12	4/ 4/ 8	4/ 18/ 4	9/ 10/ 12	6/ 8/ 8	2/ 6/ 4

Realmente es complicado clasificar una emoción como el miedo y evitar su influencia sobre las otras emociones. Al mirar los resultados, se puede notar que la cantidad de verdaderos positivos para el Grupo1 y para el Grupo2 son casi siempre cercanas al total de muestras correctas de esos grupos. Pero para el Grupo3, muestras de la emoción miedo, son casi nulas en algunos casos, lo que significa una mayor cantidad de falsos positivos en los otros dos grupos.

Analizando lo que es la clasificación de las emociones en estos tres grupos, cuando se utiliza el modelo de RekEmozio y se contrasta con las otras bases de datos, los resultados son casi nulos para el Grupo2 y Grupo3. Por otro lado, al contrastar los modelos de las otras bases de datos con las muestras de prueba de RekEmozio, principalmente el modelo de Emo-DB, las muestras del actor de Actor009 son las que fueron peor clasificadas. A pesar de que los resultados se puedan deber a la poca cantidad de muestras con las que cuenta la sub-base de datos de RekEmozio, quizás haya alguna diferencia entre las muestras de Actor006 y Actor009, además del idioma y género, que generan esa mala clasificación.

#### **A.4.5. Prueba 2-D: Clasificación de emociones por grupo - Grupo3 - RekEmozio modificada**

Se revisaron las muestras de RekEmozio, en búsqueda de alguna diferencia sustancial entre las pistas de audio de cada actor. Se encontró que las pistas de audio de Actor009 incluían uno o dos segundos silencio al principio ante de que el actor dijera la frase correspondiente. Para las emociones, la duración es una característica importante, la unidad de análisis que se elija debería ser igual para todas las muestras. Esto se explica mejor en 3.4. Por lo tanto, estos segundos de silencio podrían afectar el correcto reconocimiento de la emoción.

Al igual que cuando se extrajeron las palabras de las pistas de audio para reconocer género según el idioma (sección A.3.10), se hizo uso de la misma técnica descrita en 3.4 para extraer las frases de los archivos de

audio de Actor009, quitándole los segundos iniciales (y finales si existen) de silencio.

Las nuevas pruebas que se hicieron sólo afectaron fueron el contraste del modelo RekEmozio modificada, y el contraste de todas las bases de datos con RekEmozio modificada. De ahora en adelante sólo se dirá RekEmozio para referirse a la base de datos con los cambios recién descritos, dado que más adelante estos cambios se mantienen.

### **Muestreo**

Todas las muestras de prueba se mantuvieron, salvo las de RekEmozio, las cuales fueron procesadas para extraer sólo la frase que se expresa en ellas. Esto último también fue hecho para las muestras de entrenamiento de RekEmozio.

### **Extracción de Características**

Se hizo uso del vector *emobase2010*.

### **Modelos y Entrenamiento**

El nuevo modelo de RekEmozio, conteniendo las muestras modificadas, fue entrenado con el kernel lineal.

### **Resultados y Análisis**

En las tablas siguientes sólo se incluyen los resultados que se vieron afectados por los cambios en las muestras de RekEmozio.

*Tabla 81: Contraste de modelo RekEmozio en muestras de SAVEE.*

	<b>RekEmozio</b>		
<b>Grupo-SAVEE/Clase</b>	<b>G1 VP</b>	<b>G2 VP</b>	<b>G3 VP</b>
<b>Todo</b>	84/200/84	0/0/28	0/0/88

*Tabla 82: Contraste de modelo RekEmozio en muestras de Emo-DB.*

	<b>RekEmozio</b>		
<b>Grupo-Emo-DB/Clase</b>	<b>G1 VP</b>	<b>G2 VP</b>	<b>G3 VP</b>
<b>Todo</b>	96/191/96	0/1/34	0/2/64

Tabla 83: Clasificación de emociones en RekEmozio, contrastado con todas las bases de datos emocionales.

Grupo- Rek/ Clase	SAVEE			Emo-DB			RekEmozio		
	G1 VP	G2 VP	G3 VP	G1 VP	G2 VP	G3 VP	G1 VP	G2 VP	G3 VP
<b>Todo</b>	10/ 13/ 12	7/ 10/ 8	0/ 0/ 4	2/ 2/ 12	4/ 4/ 8	4/ 18/ 4	11/ 13/ 12	8/ 11/ 8	0/ 0/ 4

La extracción del silencio inicial no logró muchos cambios en la clasificación grupal. El cambio importante es que todas las emociones del Grupo2 de RekEmozio se clasificaron bien. Aunque sea un cambio pequeño, es importante. Este grupo, como se mencionó antes, es uno de los más importantes a clasificar. En él se encuentra la tristeza, una emoción que denota un estado vulnerable, y si se puede detectar correctamente, se estaría cumpliendo parte del propósito de este trabajo.

#### **A.4.6. Prueba 2-E: Clasificación de emociones por grupo - Grupo3 - Unificación de bases de datos**

En las pruebas anteriores se vio que la clasificación de emociones en los tres grupos elegidos, tenía buenos, aunque no perfectos, resultados al momento de clasificar las muestras de una base de datos con el modelo de la misma. Para las clasificaciones cruzadas, una base de datos con otra, los resultados no siempre fueron tan buenos. Entonces, ¿por qué no juntar lo bueno de cada una y de esa manera obtener una mejor clasificación para todas?

Y así fue, como todas las muestras de entrenamiento de cada base de datos se unificaron en un gran conjunto de muestras. Éstas fueron divididas en los tres últimos grupos definidos, y el modelo creado a partir de ellas se contrastó con las muestras de prueba de todas las bases de datos.

#### **Muestreo**

Las muestras de prueba fueron las mismas de la prueba previa, es decir, ya contando con los cambios en RekEmozio. Tal como se describió recién, se juntaron todas las muestras de entrenamiento, y se separaron en los tres grupos definidos. Para darle un nombre a esta unificación, se le llamó Todo-BD, representando a todas las bases de datos.

#### **Extracción de Características**

Aquí también se extrajeron las características de *emobase2010*.

## Modelos y Entrenamiento

El modelo de Todo-BD fue entrenado usando el kernel lineal.

## Resultados y Análisis

Usando la misma simbología para expresar la información obtenida en la prueba pasada, se muestran a continuación los resultados de esta prueba:

Tabla 84: Clasificación de grupos emociones en RekEmozio, contrastado con Todo-BD.

	Todo-BD		
Grupo-SAVEE/Clase	G1 VP	G2 VP	G3 VP
<b>Todo</b>	74/89/84	85/94/88	10/17/28

Tabla 85: Clasificación de grupos emociones en Emo-DB, contrastado con Todo-BD.

	Todo-BD		
Grupo-Emo-DB/Clase	G1 VP	G2 VP	G3 VP
<b>Todo</b>	91/103/96	55/63/64	19/28/34

Tabla 86: Clasificación de grupos emociones en RekEmozio, contrastado con Todo-BD.

	Todo-BD		
Grupo-Rek/Clase	G1 VP	G2 VP	G3 VP
<b>Todo</b>	10/11/12	8/9/8	3/4/4

Los resultados obtenidos superaron en creces a las expectativas. La cantidad de verdaderos positivos son casi la totalidad de las muestras clasificadas, y la cantidad de falsos positivos es mínima, comparada con los primeros. Aún el Miedo sigue siendo una piedra en el camino, generando unos cuantos falsos positivos, pero esta vez se logró clasificar una mayor cantidad de ellos. Todo esto demuestra que la unificación de las bases de datos, dotó al modelo de las características propias de cada una, y que además se complementaron entre ellas. Esta clasificación será la clasificación de grupos que se mantendrá.

Esta clasificación basada en distintos tipos de muestras, evidencia que es posible clasificar emociones independientemente del género, identidad e idioma de la voz de la persona en la pista de audio.

#### **A.4.7. Prueba 3-A: Reconocimiento de tristeza - Unificación de Grupo2**

En la prueba pasada, gracias a la unificación de muestras, se logró obtener una buena separación de los grupos, disminuyendo drásticamente los falsos positivos. Ya habiendo separado cada grupo, en especial el Grupo2, ahora el interés cae en ser capaces de clasificar con la mayor exactitud un audio triste dentro del Grupo2.

Siguiendo con la unificación de muestras de entrenamiento, se tomó el Grupo2 de éstas, y se dividió en Neutral y Tristeza. El modelo generado con estas muestras se contrastó con las muestras que cayeron en el Grupo2 en la prueba pasada.

##### **Muestreo**

Para cada base de datos, se tomaron como muestras de prueba sólo aquellas que fueron clasificadas como Grupo2 en la prueba pasada, incluyendo falsos positivos. El grupo de entrenamiento quedó como se describió recién, y se le llamará Todo-NS.

##### **Extracción de Características**

Al ser sólo dos clases las que se quieren clasificar, la cantidad de muestras es menor, por lo que además de usar el vector *emobase2010*, también se probó con *emo\_large*, el cual podría funcionar mejor en un caso como este, con menos clases. Este último tiene cerca de 6500 características [5]. Se sabe que mientras más características, la clasificación podría ser mejor, pero podría ser más ineficiente. Los resultados lo decidirán.

##### **Modelos y Entrenamiento**

Con ambos vectores se generó un modelo con el kernel lineal.

##### **Resultados y Análisis**

Esta vez, sólo se incluyen los datos de la clasificación dentro del Grupo2. Los datos se encuentran en un formato similar al anteriormente utilizado. X/Y/Z representa la cantidad de verdaderos positivos (X), en comparación con el total de clasificados como Tristeza del Grupo2 (Y), y la cantidad real de muestras que debieron haberse clasificado correctamente dentro del Grupo2 (Z). En la siguiente tabla se pueden ver inmediatamente los resultados para cada modelo:

Tabla 87: Reconocimiento de muestras de Tristeza en el Grupo Neutral-Tristeza.

Modelo Todo-NS/BD	SAVEE	Emo-DB	RekEmozio
<b>Todo-NS/ emobase2010-lineal</b>	16/25/27	19/23/28	2/3/4
<b>Todo-NS/ emo_large-lineal</b>	0/0/27	6/6/28	0/0/4

Claramente, el modelo lineal en base a *emobase2010* sigue siendo la mejor opción de clasificación de emociones. Los datos obtenidos muestran que la división entre Neutral y Tristeza presenta algunas dificultades, pero por lo menos la cantidad de verdaderos positivos supera el 50%.

Tal como se separaron seis emociones en tres grupos, puede existir una forma de desasociar un poco las emociones tristes de las neutrales. En las próximas pruebas se intentará una mejor separación de estas dos emociones.

#### **A.4.8. Prueba 3-B: Reconocimiento de tristeza - Arousal y Valence - Arousal +/- y Valence +/-**

Revisando la literatura sobre el reconocimiento de emociones en voz, dos conceptos aparecieron frecuentemente asociados a la clasificación de emociones en diferentes grupos, estos conceptos son *Arousal* y *Valence*. Estos conceptos fueron propuestos por primera vez por Valstar et al. [17] en el "The Continuous Audio/Visual Emotion Challenge - AVEC". AVEC 2013 es el tercer desafío que se hace para reconocer emociones mediante video y audio. El primero se realizó el año 2011. Para este año (2013) el nombre del desafío cambió a "The Continuous Audio/Visual Emotion and Depression Recognition Challenge", porque ahora consta de dos sub-desafíos; el primero es reconocimiento de emociones (allí se usan Arousal y Valence) y el segundo es para reconocer depresión.

Volviendo a los conceptos Arousal y Valence, así se les define en el paper mencionado:

- *Arousal*: Sentimiento global de un individuo de dinamismo o letargia. Involucra actividad mental y preparación física para actuar ante un evento.
- *Valence*: sentimiento general de un individuo de bienestar o aflicción.

Estas definiciones no son muy claras, pero dan una idea de cómo separar las distintas emociones. Dentro de la literatura revisada, se encontró un proyecto llamado AffectAura [18]. En éste, se reconocen distintas emociones mediante distintos medios de extracción de datos. Entre ellos

están la grabación de voz y video con una WebCam, reconocimiento de postura usando Kinect, registro de las tareas que se hacen durante el día, entre otros. Todos estos aparatos registraban cada movimiento, sonido y actividad del usuario del sistema, que se encontraba sentado frente a su escritorio trabajando.

En AffectAura, para el reconocimiento de emociones también hacen uso de las dimensiones afectivas Arousal y Valence, y en su paper se encuentra una definición de ellas un poco más clara:

- *Arousal*: Indica si la expresión (o emoción) es calmada o agitada.
- *Valence*: Indica si la emoción expresada es negativa o positiva.

Con estas definiciones queda más claro que la dimensión Arousal representa cuán agitada o calmada es la expresión de la emoción. Pero aún quedan dudas sobre la dimensión Valence, porque no queda claro en qué sentido la expresión es positiva o negativa.

Entonces, con el propósito de entender totalmente estos conceptos, se hizo un último intento, haciendo hallazgo de una tesis sobre reconocimiento de emociones del habla, "Emotion Recognition From Speech Signals" [19]. Ahí se definen Arousal y Valence de la siguiente manera:

- *Arousal*: Releja el grado de excitación que produce en nosotros cada emoción, va desde la soñolencia/aburrimiento a la excitación total.
- *Valence*: Se refiere a una emoción como positiva o negativa, su rango va desde un sentimiento no placentero hasta uno placentero.

Al revisar esta tesis, además de encontrar definiciones mucho más claras y precisas de Arousal y Valence, permitió conocer de dónde vienen estos conceptos. En el año 1980, James A. Russell introdujo una configuración circular llamada *Circumflex of Affect*, en la cual cada emoción representa una entidad de dos polos, Arousal y Valence. Esta configuración circular se puede separar en cuatro cuadrantes, lo que se puede ver en la Ilustración 25.

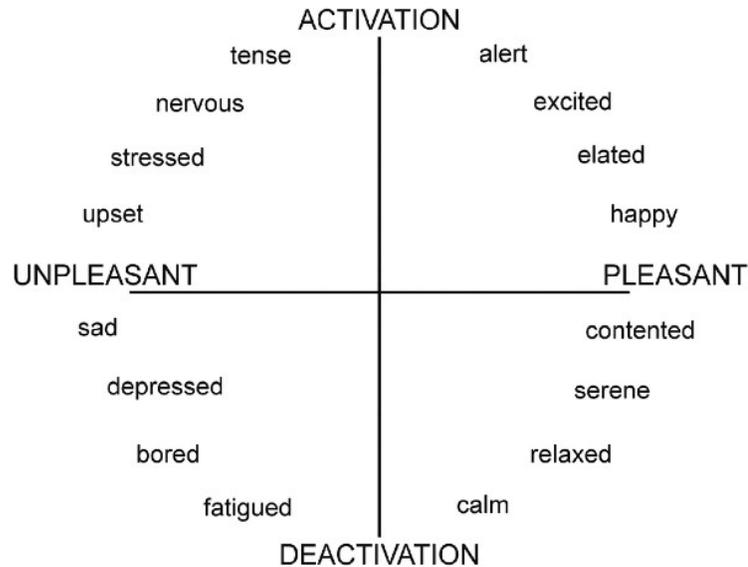


Ilustración 25: Una de las versiones del Circumflex of Affect ideado por Russell [20].

Ya entendiendo mejor qué significan Arousal y Valence, y habiendo conocido el *Circumflex of Affect*, el siguiente paso era ver cómo hacer uso de estos conocimientos. Tal como en casi todas las pruebas de emociones, en que se separaron las emociones en grupos, esta vez se podría hacer lo mismo, pero mirado desde el punto de vista de las dimensiones afectivas Arousal y Valence.

En la ilustración del *Circumflex of Affect*, se puede observar que la Tristeza se encuentra en el cuadrante de “Emoción Calmada y No placentera”. Por otra parte, el Enojo está en el cuadrante de “Emoción Agitada y No placentera”, y la Felicidad en “Emoción Agitada y Placentera”. ¿Pero dónde caen la Sorpresa, el Miedo y la emoción Neutral?

Para el mismo año 2013 se realizó el desafío InterSpeech 2013, constando de varios sub-desafíos, siendo uno de ellos la detección y clasificación de emociones basándose en las dimensiones afectivas Arousal y Valence [21]. En su paper es posible encontrar una tabla con cada una de las emociones que son parte del conjunto de prueba con el que cuenta, clasificadas en Arousal positivo o negativo y Valence positivo o negativo. Arousal positivo sería una emoción agitada, y negativo sería emoción calmada. Para Valence, positivo es una emoción placentera, y negativo es una emoción no placentera. Cabe destacar, que en la tabla mencionada aparece el Asco como una emoción no clasificable, es decir, no tiene un valor determinado de Arousal y Valence. Esto reafirma la decisión tomada anteriormente al excluirla. A continuación, una tabla resumen con las seis

emociones que son parte del conjunto de emociones de este trabajo y sus valores respectivos de Arousal y Valence según InterSpeech 2013:

*Tabla 88: Arousal y Valence de cada emoción, según los valores definidos para la base de datos de prueba de InterSpeech 2013. ? : Clasificado como otro y excluido de la evaluación. \*: no aparece en la clasificación.*

<b>Emoción/Dimensión Afectiva</b>	<b>Arousal</b>	<b>Valence</b>
<b>Enojo</b>	Positivo	Negativo
<b>Alegría</b>	Positivo	Positivo
<b>Miedo (Pánico)</b>	Positivo	Negativo
<b>Neutral*</b>	No determinado	No determinado
<b>Tristeza</b>	Negativo	Negativo
<b>Sorpresa?</b>	No determinado	No determinado

Se halló cómo clasificar el Miedo según Arousal y Valence, pero aún faltaba por clasificar Neutral y Sorpresa. A pesar de que este último quedó excluido de la clasificación del sub-desafío de clasificación de emociones de InterSpeech 2013, para este trabajo se siguió considerando, ya que ha sido posible clasificarle correctamente en su grupo.

Se revisaron otros estudios buscando conocer en qué grupo se clasificaría Neutral y Sorpresa, y también para saber qué métodos y conjuntos de características se usarían para poder clasificar estas emociones. Se encontró uno en que clasificaban emociones de distintas bases de datos según estas dimensiones afectivas [22]. En este estudio se clasificaron emociones de nueve bases de datos, entre ellas Emo-DB. En la Ilustración 26 se puede ver cómo quedan clasificadas las emociones, en especial Neutral y Sorpresa, que son las que faltaban.

En el estudio se hicieron dos clasificaciones binarias, Arousal positivo y negativo, y Valence positivo y negativo. Se compararon dos métodos de clasificación, HMM y SVM, siendo este último el que dio mejores resultados. La clasificación con SVM se hizo utilizando el conjunto de características *emo\_large*. Hay que tener en consideración, que las clasificaciones hechas fueron por cada base de datos individual, y no todas en conjunto.



Ilustración 26: Arousal y Valence para emociones de bases de datos emocionales.

Este estudio fue la pieza que faltaba para poder empezar con las pruebas finales, en búsqueda de clasificar y reconocer mejor la Tristeza.

Después de haber realizado la prueba de esta sección y las posteriores secciones, se notó que se había cometido un error, colocando el Miedo en el mismo cuadrante que Neutral, y no junto al Enojo. Es por ello, que las pruebas que lo ameritaban, se repitieron, incluyendo los resultados de la agrupación incorrecta y los resultados de la agrupación correcta. Tal como se realizó en el estudio mencionado, se llevaron a cabo dos clasificaciones binarias, una para Arousal y otra para Valence.

### Muestreo

Para cada una de las dimensiones afectivas en estudio, se crearon dos grupos, Neg y Pos, en el conjunto de entrenamiento. En el primero van aquellas emociones que caen en el rango negativo de la dimensión afectiva, y en el segundo, las correspondientes al rango positivo.

Hubo dos versiones de las muestras de entrenamiento. En la primera el Miedo está en el grupo Pos de Arousal y Neg de Valence. En la segunda está en el grupo Neg de Arousal y Pos de Valence. Las muestras de emociones fueron las mismas que se usaron para la prueba anterior.

### Extracción de Características

Se extrajeron características de dos vectores, *emo\_large*, que fue utilizado en el estudio, y *emobase2010*, que es el que mejores resultados ha dado.

## **Modelos y Entrenamiento**

Se intentaron entrenar dos modelos para cada dimensión, usando kernel polinomial (usado en estudio) y lineal, siendo cada modelo de dos clases, Neg y Pos. De los dos modelos uno fue entrenado en base a *emo\_large*, y otro en base a *emobase2010*.

Los modelos en base a *emo\_large* no lograron ser generados para cualquiera de los kernel utilizados, que fueron el polinomial de grado 2, polinomial de grado 3, y el lineal. Los de kernel polinomial lanzaron un error, y para el lineal, la generación del modelo no terminaba, por lo que el proceso fue detenido.

En el estudio funcionó utilizando *emo\_large*, pero en él el modelo creado fue para cada base de datos por separado, y en esta prueba se mezclaron todas las bases de datos con las que se contaba. Por los resultados vistos en experimentos pasados, *emo\_large* no serviría para muestras tan diversas como las muestras de emociones escogidas.

## **Resultados y Análisis**

Sólo se muestran las tablas de los resultados que interesan para reconocer correctamente la Tristeza, por lo que sólo están los resultados de la cantidad de muestras bien clasificadas en Arousal negativo y en Valence negativo. El formato es similar al previamente usado, X/Y/Z, donde X es la cantidad de verdaderos positivos que son de Tristeza, Y es la cantidad de clasificados en ese grupo, y Z la cantidad de muestras de Tristeza que realmente deberían haber sido clasificados correctamente en esa clasificación. Además, se muestran los resultados contrastados contra todas las muestras de prueba (Todo) y contra las muestras que cayeron anteriormente en el Grupo 2.

Las tablas siguientes muestran los resultados de clasificación del modelo generado en que las muestras de Miedo estaban incorrectamente agrupadas.

Tabla 89: Clasificación de emociones en Arousal negativo, con Miedo en Arousal negativo.

Modelo/BD	SAVEE	Emo-DB	RekEmozio
<b>Arousal-poli2/Todo</b>	28/132/28	28/114/28	4/14/4
<b>Arousal-poli2/Grupo2</b>	27/93/27	28/63/28	4/8/4
<b>Arousal-poli3/Todo</b>	28/151/28	28/128/28	4/17/4
<b>Arousal-poli3/Grupo2</b>	27/94/27	28/63/28	4/8/4
<b>Arousal-lineal/Todo</b>	28/138/28	28/103/28	4/15/4
<b>Arousal-lineal/Grupo2</b>	27/91/27	28/62/28	4/9/4

Tabla 90: Clasificación de emociones en Valence negativo, con Miedo en Valence positivo.

Modelo/BD	SAVEE	Emo-DB	RekEmozio
<b>Valence-poli2/Todo</b>	10/38/28	13/67/28	1/3/4
<b>Valence-poli2/Grupo2</b>	9/12/27	13/14/28	1/1/4
<b>Valence-poli3/Todo</b>	0/11/28	0/28/28	0/1/4
<b>Valence-poli3/Grupo2</b>	0/0/27	0/0/28	0/0/4
<b>Valence-lineal/Todo</b>	27/131/28	25/140/28	4/13/4
<b>Valence-lineal/Grupo2</b>	26/71/27	25/42/28	4/6/4

Tabla 91: Clasificación de emociones en Arousal negativo y Valence negativo, con Miedo en Arousal negativo y Valence positivo.

Modelo/BD	SAVEE	Emo-DB	RekEmozio
<b>ArousalNeg-ValenceNeg-poli2/Grupo2</b>	9/12/27	13/14/28	1/1/4
<b>ArousalNeg-ValenceNeg-poli3/Grupo2</b>	0/0/27	0/0/28	0/0/4
<b>ArousalNeg-ValenceNeg-lineal/Grupo2</b>	26/70/27	25/42/28	4/6/4

En estas tablas se muestran los resultados de clasificación del modelo generado en que las muestras de Miedo sí fueron correctamente agrupadas.

Tabla 92: Clasificación de emociones en Arousal negativo, con Miedo en Arousal positivo.

Modelo/BD	SAVEE	Emo-DB	RekEmozio
<b>Arousal-poli2/Todo</b>	27/102/28	28/70/28	2/6/4
<b>Arousal-poli2/Grupo2</b>	26/89/27	28/59/28	2/5/4
<b>Arousal-poli3/Todo</b>	28/112/28	28/87/28	4/9/4
<b>Arousal-poli3/Grupo2</b>	27/94/27	28/63/28	4/8/4
<b>Arousal-lineal/Todo</b>	28/100/28	28/62/28	4/10/4
<b>Arousal-lineal/Grupo2</b>	27/91/27	28/60/28	4/9/4

Tabla 93: Clasificación de emociones en Valence negativo, con Miedo en Valence negativo.

Modelo/BD	SAVEE	Emo-DB	RekEmozio
<b>Valence-poli2/Todo</b>	21/111/28	27/148/28	4/14/4
<b>Valence-poli2/Grupo2</b>	20/43/27	27/45/28	4/5/4
<b>Valence-poli3/Todo</b>	27/158/28	28/170/28	4/17/4
<b>Valence-poli3/Grupo2</b>	26/84/27	28/60/28	4/8/4
<b>Valence-lineal/Todo</b>	28/183/28	28/180/28	4/21/4
<b>Valence-lineal/Grupo2</b>	27/87/27	28/59/28	4/8/4

Tabla 94: Clasificación de emociones en Arousal negativo y Valence negativo, con Miedo en Arousal positivo y Valence negativo.

Modelo/BD	SAVEE	Emo-DB	RekEmozio
<b>ArousalNeg-ValenceNeg-poli2/Grupo2</b>	20/43/27	27/42/28	2/5/4
<b>ArousalNeg-ValenceNeg-poli3/Grupo2</b>	26/82/27	28/59/28	4/8/4
<b>ArousalNeg-ValenceNeg-lineal/Grupo2</b>	26/87/27	25/42/28	4/6/4

La Tabla 91 y Tabla 94 reflejan la clasificación final de las muestras que coincidieron para Arousal negativo y Valence negativo, que es el cuadrante

en que cae la Tristeza. Sólo se incluyen los resultados en la clasificación dentro del Grupo 2, porque en esta clasificación se obtuvieron mejores resultados a nivel de clasificación de Arousal y Valence por separado. A pesar de esto, y la gran cantidad de verdaderos positivos, la cantidad de falsos positivos es mucho mayor en la mayoría de los casos. Gran parte de los falsos positivos fueron muestras de emoción Neutral, y algunos de la emoción Miedo.

De los kernel utilizados, el que proveyó mejores resultados fue el polinomial de grado 2, seguido por el lineal. El polinomial de grado 3 logró clasificar correctamente muy pocas o ninguna muestra de Tristeza.

Aunque haya resultado en el estudio investigado, no es la mejor técnica para el tipo de muestras que se ha manejado en este trabajo, por lo que se puede intentar con otras divisiones para lograr reconocer de forma más precisa la Tristeza.

#### **A.4.9. Prueba 3-C: Reconocimiento de tristeza - Arousal y Valence - 4 grupos**

Se vio que la clasificación binaria para dimensión afectiva no logró tan buenos resultados como se podría haber esperado. En esta prueba la perspectiva se cambió un poco, en vez de separar por cada dimensión, mejor se aprovechó la división en cuadrantes que el *Circumflex of Affect* proporciona. Y eso fue lo que se hizo. Se separaron las muestras de emociones en cuatro grupos, cada uno correspondiendo a cada cuadrante.

##### **Muestreo**

Dado que la nueva división sería según el *Circumflex of Affect*, se formaron cuatro grupos en las muestras de entrenamiento. Como se contó anteriormente, se cometió un error al agrupar las muestras de Miedo. En esta prueba se contabilizan la incorrecta y correcta agrupación de estas muestras.

##### **Extracción de Características**

Los conjuntos de características elegidos para esta prueba fueron *emo\_large* y *emobase2010*.

##### **Modelos y Entrenamiento**

Nuevamente los modelos entrenados en base a *emo\_large* no pudieron ser generados, a causa de los mismos problemas de la prueba pasada. Así que, sólo se pudieron generar modelos en base a *emobase2010*. Se probó con el kernel polinomial de grado 2, con  $\gamma=1$  y  $C=1$ , y el kernel lineal.

##### **Resultados y Análisis**

La presentación de los datos obtenidos es bastante similar a la anterior prueba. En cada tabla se encuentran los datos para cada uno de los modelos

contrastados contra Todo y Grupo 2. AV4 nn se refiere a esta prueba, Arousal y Valence en 4 grupos y nn el cuadrante donde está la Tristeza, Arousal negativo y Valence Negativo, por lo que se muestran sólo los resultados para ese cuadrante.

Tabla 95: Clasificación de emociones en cuadrante Arousal negativo y Valence negativo, con Miedo en Arousal negativo y Valence positivo.

Modelo/BD	SAVEE	Emo-DB	RekEmozio
<b>AV4 nn-poli2/Todo</b>	11/15/28	16/16/28	0/0/4
<b>AV4 nn-poli2/Grupo2</b>	11/14/28	16/16/28	0/0/4
<b>AV4 nn-lineal/Todo</b>	18/29/27	22/28/28	4/5/4
<b>AV4 nn-lineal/Grupo2</b>	17/25/27	22/27/28	4/5/4

Tabla 96: Clasificación de emociones en cuadrante Arousal negativo y Valence negativo, con Miedo en Arousal positivo y Valence negativo.

Modelo/BD	SAVEE	Emo-DB	RekEmozio
<b>AV4 nn-poli2/Todo</b>	17/27/28	22/27/28	2/2/4
<b>AV4 nn-poli2/Grupo2</b>	16/26/27	22/25/28	2/2/4
<b>AV4 nn-lineal/Todo</b>	15/23/28	22/23/28	2/2/4
<b>AV4 nn-lineal/Grupo2</b>	15/23/27	22/23/28	2/2/4

Esta clasificación generó resultados mucho mejores que en el experimento pasado. Aunque la cantidad de verdaderos positivos es menor, la cantidad de falsos positivos también lo es. Es posible observar que en general los resultados son mejores cuando el modelo contrastado es en aquel que el Miedo fue correctamente agrupado (Tabla 96). Además, la clasificación mejora al clasificar dentro del Grupo 2, al igual que la prueba pasada. Esta especie de doble filtro demuestra tener mejores resultados.

El uso de kernel lineal sigue mostrando ser la mejor opción. La disyuntiva es cuál de estas dos clasificaciones es la mejor. En la siguiente tabla se comparan los promedios de los porcentajes de exactitud con que se reconoció correctamente muestras de Tristeza, para cada uno de las dos pruebas hechas en esta sección. Miedo I se refiere a la agrupación incorrecta de las muestras de Miedo, y Miedo C es la agrupación correcta.

Tabla 97: Exactitud de clasificación de muestras de Tristeza en cada prueba y base de datos.

Modelo/BD	SAVEE Tristeza (%)	Emo-DB Tristeza (%)	RekEmozio Tristeza (%)
AV4 nn-lineal-Miedo I/Grupo2	0,66	0,75	0,63
AV4 nn-lineal-Miedo C/Grupo2	0,60	0,69	0,52

La mejor exactitud de clasificación la provee el modelo en que las muestras de Miedo están mal agrupadas, pero esto también se observó en las muestras que fueron falsos positivos, también tenían una mayor exactitud de clasificación en este modelo.

Entre elegir una mejor clasificación y un mayor número de muestras bien clasificadas, se prefirió la segunda opción. Si es necesario, después se pueden filtrar según su valor de exactitud, estableciendo un umbral.

#### **A.4.10. Prueba 3-D: Reconocimiento de tristeza - Arousal y Valence - 5 grupos**

Como se había agrupado mal inicialmente la emoción Miedo, y los resultados anteriores en base a esa agrupación tenían una cantidad no despreciable de falsos positivos, se revisó lo que podría estar generando eso. Se debía principalmente a que muestras de emoción Neutral caían en el grupo de la Tristeza. Entonces, se probó sacando las muestras de Neutral y colocándolas en un grupo aparte. Las muestras quedaron divididas en 5 grupos.

##### **Muestreo**

Se tomaron las mismas muestras de entrenamiento en que el Miedo estaba mal agrupado, y se extrajeron las muestras de Neutral a un grupo aparte.

##### **Extracción de Características**

El vector de características usado fue *emobase2010*.

##### **Modelos y Entrenamiento**

Se usó el kernel lineal para entrenar el modelo.

##### **Resultados y Análisis**

El formato de los datos presentados es similar al anterior y también se muestran sólo los resultados correspondientes al cuadrante Arousal negativo y Valence negativo.

Tabla 98: Clasificación de emociones en cuadrante Arousal negativo y Valence negativo, con Miedo en Arousal negativo y Valence positivo, y Neutral aparte.

Modelo/BD	SAVEE	Emo-DB	RekEmozio
AV5-lineal/Todo-BD	19/35/28	21/25/28	3/3/4
AV5-lineal/Grupo2	18/32/27	21/23/28	3/3/4

Los cambios fueron un poco buenos y un poco malos, porque aumentaron los falsos positivos en la clasificación para SAVEE, pero mejoró un poco para las muestras Emo-DB, disminuyendo los falsos positivos. Para RekEmozio, mejoró en el sentido en que todos los clasificados fueron correctos.

El efecto en SAVEE, se debió a lo mismo que se quería evitar. Al separar Neutral, aumentó las probabilidades de que muestras de Neutral fueran detectadas como Tristeza, porque ya no estaba el efecto que la emoción Miedo generaba en la emoción Neutral en el modelo.

#### **A.4.11. Prueba 3-E: Reconocimiento de tristeza - Arousal y Valence - 4 grupos sin Neutral**

Se hizo un último intento, se extrajo la emoción Neutral de las muestras de entrenamiento, pensando que esto podría alejar un poco las muestras de prueba Neutral de ser clasificadas como Tristeza (cuadrante Arousal negativo y Valence negativo).

##### **Muestreo**

Las muestras de entrenamiento sólo contaron con 4 grupos, y en el grupo de Arousal negativo y Valence positivo sólo estaban las muestras de Miedo. Las muestras de Neutral se sacaron por completo del conjunto de entrenamiento.

##### **Extracción de Características**

Se usó el vector de características *emobase2010*.

##### **Modelos y Entrenamiento**

El modelo fue entrenado con el kernel lineal.

##### **Resultados y Análisis**

Las siguientes tablas muestran los resultados de esta clasificación en que Neutral no es parte del modelo contrastado.

Tabla 99: Clasificación de emociones en cuadrante Arousal negativo y Valence negativo, con Miedo en Arousal negativo y Valence positivo, y sin Neutral.

Modelo/BD	SAVEE	Emo-DB	RekEmozio
AV4SN-lineal/Todo-BD	25/85/28	26/41/28	4/8/4
AV4SN-lineal/Grupo2	21/81/27	26/43/28	4/8/4

Al sacar las muestras de emoción Neutral, la clasificación empeoró mucho más, generando muchos falsos positivos. La clasificación en el Grupo 2 hizo que casi todas las muestras cayeran en Arousal negativo y Valence negativo, sin importar la emoción, siendo casi todas Neutral y Tristeza.

Esta es la última prueba que se hizo en clasificación de emociones específicamente. La mejor clasificación de Tristeza se logró al contrastar el modelo basado en la división de las emociones en los cuatro cuadrantes del *Circumflex of Affect*, sobre el Grupo 2. Para clasificar otras emociones habría que comparar cada uno de los modelos y ver cuál se ajusta mejor, o probar con otro modelo.

#### A.5. Pruebas de Conversaciones

Cuando llegue el momento de utilizar el sistema, las grabaciones que se obtengan no serán tan similares a las ya utilizadas en las pruebas anteriores, específicamente debido a la naturalidad o mayor realidad que éstas tendrán, ya no serán grabaciones actuadas como las que se encuentran en las bases de datos antes mencionadas. Por ello, se grabaron algunas conversaciones, en busca de demostrar el buen funcionamiento de los modelos creados, o para encontrar falencias, tanto para la identificación de la emoción, el género y la identidad de la persona.

##### A.5.1. Prueba 1: Conversación entre 2 hombres

Se grabó una conversación entre dos personas de género masculino. La conversación se basó en la lectura de dos páginas de una obra de teatro. Las personas involucradas en esta conversación, fueron dos personas que ya habían participado en el proceso inicial de recolección de datos, cuando se grabaron voces para el reconocimiento de género e identidad, Hombre2 y Hombre5.

El audio fue procesado para extraer las frases de la conversación. Tal como se hizo antes, esta vez también se hizo como se describe en la sección 3.4. Con esta técnica se obtuvieron 28 muestras.

## Proceso de Reconocimiento

Cuando se hizo esta prueba por primera vez, el enfoque que se utilizó era el inicialmente señalado (ver Ilustración 19 en sección A.3), en el que primero se reconocía género, luego identidad, y al final la emoción. Entonces se comenzó por intentar reconocer el género de los hablantes.

Primero se probó con el modelo que mejor había funcionado hasta ese momento, este fue el que contenía las muestras iniciales de Hombre y Mujer, además de algunas muestras de emociones, todas con el volumen normalizado (sección A.3.7). Para ello se normalizaron las 28 muestras de la conversación. Los resultados de la clasificación fueron los siguientes:

Tabla 100: Clasificación según género en muestras de primera conversación, primer intento.

Conversación1/Clase	Mujer (%)	Hombre (%)	VP
<b>Todo</b>	43,7	56,3	17

Siendo que todas las muestras eran de voces masculinas, sólo un poco más de la mitad fueron bien clasificadas, por lo que se probó con el modelo anterior que había tenido mejores resultados. Aquél en que las muestras de emoción aún no eran incluidas en el entrenamiento. Esta vez se probó en las muestras no normalizadas.

Tabla 101: Clasificación según género en muestras de primera conversación, segundo intento.

Conversación1/Clase	Mujer (%)	Hombre (%)	VP
<b>Todo</b>	9,3	90,7	28

Aunque esta última clasificación resultó bastante mejor, logrando reconocer todas las muestras como masculinas, se hicieron más pruebas en reconocimiento de género. Esto debido a que en las pruebas posteriores este reconocimiento siguió teniendo problemas. Principalmente estos problemas fueron los que generaron la creación de más pruebas en la sección de clasificación según género, pruebas que comienzan en A.3.8 y terminan en A.3.11, siendo esta última que se eligió como la mejor opción.

Todo cambió al hacer las nuevas pruebas de clasificación según género. Con ello cambió el enfoque de reconocimiento general, por lo que lo primero que había que hacer era reconocer la emoción, y después el género. La mejor clasificación de emociones para encontrar si un audio corresponde a Tristeza fue una combinación de lo visto en A.4.6 y en A.4.9, una clasificación con "doble" filtro, clasificando primero en 3 grupos que se

definieron según pruebas iniciales, y luego en los 4 grupos o cuadrantes formado por las dimensiones Arousal y Valence.

Se clasificaron las muestras según emoción, esperando que todas las muestras cayeran en el Grupo 2 según el primer filtro, y luego en el Grupo 4 según el segundo filtro, dado que al ser una lectura, y no verdadera actuación, se espera que las muestras extraídas se clasifiquen como Neutral.

Tabla 102: Clasificación según emoción en muestras de primera conversación, primer filtro (Todo-BD).

Conversación1/Clase	Grupo 1 (%)	Grupo 2 (%)	Grupo 3 (%)	VP
<b>Todo</b>	5,9	87,9	6,3	28/28

Tabla 103: Clasificación según emoción en muestras de primera conversación, segundo filtro (Arousal negativo y Valence negativo, AV4 nn).

Conversación1/Clase	Grupo 1 (%)	Grupo 2 (%)	Grupo 3 (%)	Grupo 4 (%)	VP
<b>Grupo 2 - Todo-BD</b>	8,6	1,9	4,2	85,3	27/28

Se puede ver que todas las muestras, menos una, cayeron finalmente en el Grupo 4 del segundo filtro, indicando que son muestras de emoción Neutral, tal como se esperaba. La muestra de audio que faltó para ser clasificada correctamente cayó en el Grupo 1, confirmando la confusión que se genera entre Miedo y Neutral.

Ya habiendo reconocido la emoción, el siguiente paso es la clasificación según género. Como es una emoción más calmada, se vio en A.3.11 que esta clasificación funciona mejor cuando se normaliza el volumen de las muestras. Por lo tanto, se normalizó el volumen de las muestras que fueron bien clasificadas.

Tabla 104: Clasificación según género en muestras de primera conversación, que pasaron el segundo filtro de emoción. Se clasifica de acuerdo a emoción detectada.

Conversación1/Clase	Mujer (%)	Hombre (%)	VP
<b>Grupo4 - AV4 nn - normalizado</b>	26,3	73,7	27/27

Con estos resultados se confirma lo que es la clasificación según género. Ahora sólo queda el reconocimiento de identidad. Se probó con el reconocimiento de identidad que mejores resultados arrojó (A.3.4), donde el modelo está basado en *emobase2010*. La prueba se hizo sobre las mismas muestras, pero sin el volumen normalizado.

*Tabla 105: Reconocimiento de identidad en muestras de primera conversación, usando modelo basado en emobase2010.*

<b>Hombre</b>	<b>VP</b>	<b>FP</b>
<b>Hombre1</b>	0	27
<b>Hombre2</b>	0	0
<b>Hombre3</b>	0	0
<b>Hombre4</b>	0	0
<b>Hombre5</b>	0	0

Este reconocimiento colocó todas las muestras en Hombre1, siendo que ninguno de los hablantes es Hombre1. Esto recordó el problema de clasificación según género, y el efecto de las emociones en esto. Entonces, se probó el modelo en que se usó el vector de característica *IS10\_paraling*, que aunque arrojaba una menor exactitud, puede ser una clasificación más real.

*Tabla 106: Reconocimiento de identidad en muestras de primera conversación, usando modelo basado en IS10\_paraling.*

<b>Hombre</b>	<b>VP</b>	<b>FP</b>
<b>Hombre1</b>	0	10
<b>Hombre2</b>	0	0
<b>Hombre3</b>	0	2
<b>Hombre4</b>	0	0
<b>Hombre5</b>	10	5

Se obtuvieron mejores resultados, pero no perfectos. Esto indica que es necesario analizar con mayor profundidad esta etapa de reconocimiento, para que sea más efectiva.

### **A.5.2. Prueba 2: Conversación entre una mujer y un hombre**

Esta segunda prueba, consistió en una conversación entre un hombre y una mujer conversando, algo más natural que la anterior, sin estar leyendo. La conversación fue entre Hombre5 y Mujer2, lo que permite probar desde el reconocimiento de emociones hasta reconocimiento de identidad.

Con la extracción de frases se obtuvieron 72 muestras. A causa del solapamiento de voces durante la conversación, 11 muestras contienen voces de ambos hablantes. De las 61 restantes, de forma clara se puede decir que 48 corresponden a Mujer2 y 13 a Hombre5.

#### **Proceso de Reconocimiento**

Como se contó en la prueba pasada, primero se probó con el modelo para reconocer género que mejor había funcionado para las muestras iniciales y las muestras de emociones, hasta que se realizó la grabación de esta conversación. Los resultados fueron buenos en general, si no contamos las muestras de audio con solapamiento:

*Tabla 107: Clasificación según género en muestras de segunda conversación, primer intento.*

<b>Conversación1/Clase</b>	<b>VP</b>	<b>FP</b>
<b>Mujer</b>	12	6
<b>Hombre</b>	45	5

Como comentario, la mayoría de las muestras con solapamiento fueron bien clasificadas si tomamos en cuenta la voz que más predomina en ellas. Se decidió probar nuevamente con el mejor modelo que había clasificado según género al inicio de todo, el que estaba basado en *emobase2010*, y que se había probado en muestras de personas como Mujer2 y Hombre5.

*Tabla 108: Clasificación según género en muestras de segunda conversación, segundo intento.*

<b>Conversación1/Clase</b>	<b>VP</b>	<b>FP</b>
<b>Mujer</b>	41	0
<b>Hombre</b>	12	10

Aquí decayeron los buenos resultados. Se encontró una gran cantidad de falsos positivos en el grupo de los hombres. Las muestras con solapamiento de voces, en general fueron bien clasificadas, con algunas excepciones.

Como se había cambiado el enfoque, se probó partir esta vez clasificando según la emoción, y luego según género. Todas las muestras también deberían ser reconocidas como Neutral, al igual que la prueba pasada.

Tabla 109: Clasificación según emoción en muestras de segunda conversación, primer filtro (Todo-BD).

Conversación2/Clase	Grupo 1 (%)	Grupo 2 (%)	Grupo 3 (%)	VP
<b>Todo</b>	18,6	62,9	18,5	52/72

Tabla 110: Clasificación según emoción en muestras de segunda conversación, segundo filtro (Arousal negativo y Valence negativo, AV4 nn).

Conversación2/Clase	Grupo 1 (%)	Grupo 2 (%)	Grupo 3 (%)	Grupo 4 (%)	VP
<b>Grupo 2 - Todo-BD</b>	14,5	10,3	18,0	57,2	39/52

Al aplicar el primer filtro, muchas muestras no cayeron en el Grupo 2. Ocho muestras cayeron en el Grupo 1 y doce en el Grupo 2. Lo interesante de esto es que dentro de las que cayeron en los grupos 1 y 2, nueve fueron muestras con solapamiento de voces. De alguna manera, se puede decir que el filtro también evita que muestras no muy claras caigan en el grupo correspondiente, es una especie de doble filtro para estas muestras. Por un lado es bueno, para detectar mejor una emoción, en este caso Neutral, y en el propósito principal de este trabajo, Tristeza. Pero por otro lado, si se quisiera reconocer otras emociones, esto sólo ocasionaría ruido. Por lo tanto, lo ideal sería poder detectar de otra manera estas muestras con solapamiento de voces, y descartarlas antes de cualquier clasificación, o si es posible, extraer con mayor precisión las frases que pueda contener, evitando el solapamiento.

Después de aplicar el segundo filtro, una de las dos muestras que cayeron en el Grupo 2 según Todo-BD, cayó en el Grupo 1 según AV4 nn. Esto es otro ejemplo de lo que se dijo recién sobre el doble filtro, pero que no es la mejor opción para descartar este tipo de muestras.

De las 72 muestras, si no se cuentan las muestras con solapamiento, quedan 61. De esas 61, 39 fueron totalmente bien clasificadas. El número no es tan malo, pensando que es casi dos tercios del total, además de que esta prueba representa un poco más a la realidad.

Algo curioso de esta prueba, fueron las 9 muestras que después del segundo filtro cayeron en el cuadrante Arousal negativo y Valence negativo, es decir, Tristeza. Habrá que revisar con calma este tipo de casos, para ver si el umbral, que se mencionó como opción para descartar por completo muestras que no sean Tristeza, es la solución.

Después de la clasificación de emociones, vino la clasificación de género. Al pasar por los filtros de emociones, varias muestras de mujer se perdieron. La siguiente tabla detalla cómo resultó esta siguiente clasificación:

*Tabla 111: Clasificación según género en muestras de segunda conversación, que pasaron el segundo filtro de emoción. Se clasifica de acuerdo a emoción detectada.*

<b>Conversación2</b>	<b>VP</b>	<b>FP</b>
<b>Grupo4 - AV4 nn - normalizado - Mujer</b>	19	1
<b>Grupo4 - AV4 nn - normalizado - Hombre</b>	10	4

De lo que se observa, se puede notar que casi todas las muestras fueron bien clasificadas. El falso positivo en el grupo de mujeres, es una muestra considerada con solapamiento, y en ella predomina la voz femenina. En ese caso, se podría decir que sí fue bien clasificada. Para el caso del grupo de hombres, o voces masculinas, los 4 falsos positivos corresponden a muestras de voz femenina.

Por último queda el reconocimiento de identidad. Primero con el modelo basado en *emobase2010*.

Tabla 112: Reconocimiento de identidad en muestras de segunda conversación, usando modelo basado en emobase2010.

<b>Mujer</b>	<b>VP</b>	<b>FP</b>
<b>Mujer1</b>	0	0
<b>Mujer2</b>	3	0
<b>Mujer3</b>	0	0
<b>Mujer4</b>	0	17
<b>Mujer5</b>	0	0

Tabla 113: Reconocimiento de identidad en muestras de segunda conversación, usando modelo basado en emobase2010.

<b>Hombre</b>	<b>VP</b>	<b>FP</b>
<b>Hombre1</b>	0	6
<b>Hombre2</b>	0	5
<b>Hombre3</b>	0	0
<b>Hombre4</b>	0	3
<b>Hombre5</b>	0	0

Como se puede ver, casi todas las muestras fueron mal clasificadas. De alguna manera también se repite un fenómeno mencionado en A.3.1, sobre la confusión debido al parentesco, en este caso entre Mujer2 y Mujer4, y Hombre4 y Hombre5.

Nuevamente se probó con el modelo basado en *IS10\_paraling*, en busca de mejores resultados.

Tabla 114: Reconocimiento de identidad en muestras de segunda conversación, usando modelo basado en IS10\_paraling.

Mujer	VP	FP
Mujer1	0	0
Mujer2	5	0
Mujer3	0	1
Mujer4	0	12
Mujer5	0	2

Tabla 115: Reconocimiento de identidad en muestras de segunda conversación, usando modelo basado en IS10\_paraling.

Hombre	VP	FP
Hombre1	0	4
Hombre2	0	0
Hombre3	0	0
Hombre4	0	7
Hombre5	1	2

Aunque mejoraron casi nada los resultados, se puede ver que la distribución cambió, acentuándose la confusión de muestras que se debería principalmente al parentesco. El reconocimiento de identidad queda pendiente como algo a mejorar. No se puede decir aún que funciona correctamente.

### **A.5.3. Prueba 3: Conversación entre una mujer y un hombre, con solapamiento de voces**

En esta última prueba de conversación participaron el mismo Hombre5 y la Mujer2, ambos conversando al unísono, para que se superpusieran las voces. En la conversación anterior se encontraron muestras con solapamiento y muchas de ellas fueron clasificadas en otros grupos no correspondientes a Neutral y Tristeza, lo que se mencionó que podría ser como un filtro. En esta prueba no fue tan así. Gran parte de las muestras cayó en el Grupo 2 según Todo-BD, y esas mismas, cayeron en el Grupo 4 según AV4 nn.

Tabla 116: Clasificación según emoción en muestras de tercera conversación, primer filtro (Todo-BD).

Conversación3/Clase	Grupo 1 #	Grupo 2 #	Grupo 3 #
<b>Todo</b>	0	106	36

Tabla 117: Clasificación según emoción en muestras de tercera conversación, segundo filtro (Arousal negativo y Valence negativo, AV4 nn).

Conversación3/Clase	Grupo 1 #	Grupo 2 #	Grupo 3 #	Grupo 4 #
<b>Grupo 2 - Todo-BD</b>	0	0	0	106

No se alcanzaron a hacer más pruebas con estas muestras. Al menos lo probado indica claramente que se debe hacer un esfuerzo por evitar que muestras no bien definidas sean clasificadas como correctas.

## Anexo B: Scripts y Resultados

Para tener una idea más clara de cómo funcionó el proceso de testeo de modelos para una clasificación, se presentan algunos ejemplos del código y de los archivos de resultados.

### B.1. Código del script `test_model_folders.sh`

```
#!/bin/bash
AUDIO_DIR=$1
MODEL=$2
CONF_NAME=$3

model_base=`basename $MODEL`
model_base=${model_base%.*.*}

SMILE_DIR="/.../opensmile-2.0-rc1/opensmile"
folders=`ls -d $AUDIO_DIR*/`
conf_name=$CONF_NAME
num_class=1
for folder in $folders; do
    folder_name=`basename $folder`
    lsvm_files=`find $folder -name \*$conf_name.lsvm`
    if [ "$lsvm_files" == "" ]
    then
        audio_files=`ls $folder/*.wav`
        for audio_file in $audio_files; do
            SMILEextract -C $SMILE_DIR/config/$conf_name.conf -I
            $audio_file -O $folder/
            $folder_name-$conf_name.arff
        done
        # delete string field in arff file
        sed "s/'noname',//g" -i $folder/
        $folder_name-$conf_name.arff
        sed "s/@attribute name string//g" -i $folder/
        $folder_name-$conf_name.arff
        # transform arff to libsvm
    fi
done
```

```
perl $SMILE_DIR/scripts/modeltrain/arffToLsvm.pl $folder/  
$folder_name-$conf_name.arff  
  
fi  
  
python read_predict.py --pf $folder/  
$folder_name-$model_base-predict  
  
done
```

### ***B.2. Ejemplo de resultados en archivo predict***

```
labels 1 2 3 4 5 6 7  
1 0.519612 0.00439782 0.0151183 0.0524622 0.350404 0.049766 0.00824022  
0.000647876  
1 0.80628 0.000156071 0.0120753 0.0029579 0.169077 0.00839862 0.00105441  
6 0.172805 0.060123 0.138828 0.230354 0.0745095 0.276272 0.0471088  
1 0.730643 0.00285521 0.0200031 0.020421 0.166393 0.0428293 0.016855  
1 0.754601 1.80754e-05 0.0035422 0.000589675 0.237643 0.00332973 0.000276183  
3 0.212787 0.0522294 0.292114 0.150225 0.0955337 0.170504 0.0266075
```

### ***B.3. Ejemplo de resultados en archivo predict-results***

```
1 0.58846583871 58.8  
2 0.00892254712784 0.9  
3 0.0467847268097 4.7  
4 0.0531971665161 5.3  
5 0.246448175806 24.6  
6 0.0471017359516 4.7  
7 0.0090798184 0.9  
labels 1.17741935484 117.7
```

## **Anexo C: Aplicación Móvil y su Estudio de Usabilidad**

Se incluyen en esta sección los instrumentos de usabilidad realizados para cada método de evaluación. Además se agregaron algunas de las imágenes de la interfaz de la aplicación que no fueron incluidas en los capítulos anteriores.

### ***C.1. Pauta de Usuario Final***

## Pauta de Evaluación Usuario Final

### Introducción

La presente pauta tiene por objetivo evaluar la usabilidad de la aplicación móvil de EmoStatus. Es importante destacar que la aplicación se encuentra en una fase intermedia, solo permitiendo la navegación y generación de acciones de forma básica, y sin datos reales. Si se desea realizar un cambio en la aplicación, por ejemplo, modificar un dato, esto será posible, pero los cambios no se guardarán.

Esta Pauta debe ser aplicada luego que UD. haya explorado y navegado detenidamente por la aplicación, con uno o más objetivos en mente.

### Antecedentes

Nombre del evaluador

Edad

Género

Experiencia en uso de aplicaciones móviles

### Evaluación

	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
La aplicación es fácil de utilizar.					
Es fácil navegar entre una sección y otra.					
Es fácil encontrar la información u opción de configuración deseada.					
Los botones o elementos de acción funcionan correctamente.					
El uso del color es apropiado.					
El diseño general de la aplicación es apropiado.					
Los iconos de la aplicación son claramente identificables.					
Me gustan los iconos de la aplicación.					
Los iconos de la aplicación me transmiten información.					
La información entregada es clara.					
El contenido que se entrega es relevante.					
La interfaz de la aplicación es placentera.					

### Cuestionario

1.- ¿Qué te gustó de la aplicación?

2.- ¿Qué no te gustó de la aplicación?

3.- ¿Qué agregarías a la aplicación?

**Observaciones o comentarios**

## ***C.2. Pauta de Evaluación Heurística respondida por Experto1***

## Pauta de Evaluación Heurística

### Introducción:

La presente pauta tiene por objetivo evaluar la usabilidad de la aplicación móvil de EmoStatus. Es importante destacar que la aplicación se encuentra en una fase intermedia, solo permitiendo la navegación y generación de acciones de forma básica, y sin datos reales. Si se desea realizar un cambio en la aplicación, por ejemplo, modificar un dato, esto será posible, pero los cambios no se guardarán.

Esta Pauta debe ser aplicada luego que UD. haya explorado y navegado detenidamente por la aplicación, con uno o más objetivos en mente.

### Antecedentes

Nombre del evaluador

Edad	
<input checked="" type="checkbox"/>	Entre 20 y 25
<input type="checkbox"/>	Entre 26 y 35
<input type="checkbox"/>	Entre 36 y 45
<input type="checkbox"/>	Mayor de 45

Estudios	
<input checked="" type="checkbox"/>	Título Profesional
<input type="checkbox"/>	Postítulo
<input type="checkbox"/>	Magister
<input type="checkbox"/>	Doctorado

Experiencia en uso de aplicaciones móviles	
<input type="checkbox"/>	Ocasionalmente
<input type="checkbox"/>	Una vez a la semana
<input checked="" type="checkbox"/>	Varios días a la semana
<input type="checkbox"/>	Todos los días de la semana

### Evaluación

I. Visibilidad del estado del sistema	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
1.1. La aplicación muestra claramente dónde se encuentra el usuario	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.2. Las opciones posibles de explorar están claramente señaladas	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

II. Relación entre sistema y mundo real	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
2.1. El lenguaje es claro	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.2. Los conceptos utilizados son entendibles	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.3. Las palabras son de significado conocido	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.4. Los iconos generan significado	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

III. Control del usuario y libertad	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
3.1. Es fácil regresar al punto inmediatamente anterior	<del>X</del>				
3.2. Es fácil volver a la sección principal (lista de usuarios monitoreados) desde cualquier otra sección	<del>X</del>				
3.3. Provee botones o gestos para volver o dar paso a otra sección	<del>X</del>				
3.4. La aplicación muestra buena visibilidad al cambiar de orientación	<del>X</del>				

IV. Consistencia y estándares	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
4.1. La aplicación muestra claramente dónde se encuentra el usuario	<del>X</del>				
4.2. Las opciones posibles de explorar están claramente señaladas	<del>X</del>				
4.3. Existen coherencias entre el título de una sección y su contenido	<del>X</del>				
4.4. Sólo existe un botón que lleve a un mismo sitio	<del>X</del>				

V. Prevención de errores	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
5.1. Existen mensajes que prevengan posibles errores				<del>X</del>	
5.2. Es posible prever posibles errores		<del>X</del>			
5.3. La aplicación no induce a cometer errores		<del>X</del>			

VI. Reconocer en lugar de recordar	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
6.1. Los iconos son fácilmente reconocibles	<del>X</del>				
6.2. Los elementos de acción puede identificarse claramente	<del>X</del>				
6.3. Es posible reconocer dónde se encuentra el usuario	<del>X</del>				

VII. Flexibilidad y eficiencia de uso	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
7.1. Los lugares son de fácil acceso	<del>X</del>				
7.2. Es fácil modificar las opciones de grabación en la sección de monitoreo	<del>X</del>				
7.3. Es fácil encontrar información en el historial de un usuario monitoreado				<del>X</del>	

VIII. Estética y diseño minimalista	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
8.1. La información es relevante		<del>X</del>			
8.2. La cantidad de información entregada es suficiente	<del>X</del>				
8.3. Las secciones están correctamente organizadas	<del>X</del>				

IX. Reconocimiento, diagnóstico y recuperación de errores	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
9.1. Es fácil reconocer cuando ocurre un error				<del>X</del>	
9.2. Después de que ocurre un error es fácil volver a la sección de origen			<del>X</del>		
9.3. Cuando ocurre un error existen mecanismos para solucionarlos			<del>X</del>		

X. Ayuda y documentación	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
10.1. Existe algún tipo de ayuda o indicación en la aplicación					<del>X</del>
10.2. Cuando existe ayuda, ésta es específica			<del>X</del>		
10.3. La ayuda está asequible			<del>X</del>		

¿Cómo clasifica globalmente la aplicación?	Excelente	Bueno	Neutro	Regular	Deficiente
I. Visibilidad del estado del sistema	X				
II. Relación entre sistema y mundo real		X			
III. Control del usuario y libertad	X				
IV. Consistencia y estándares	X				
V. Prevención de errores		X			
VI. Reconocer en lugar de recordar	X				
VII. Flexibilidad y eficiencia de uso		X			
VIII. Estética y diseño minimalista	X				
IX. Reconocimiento, diagnóstico y recuperación de errores					X
X. Ayuda y documentación					X

### Justificación y Comentarios libres

- Como la app no esta funcional no es posible generar errores, pero en el caso de imprimir un mailo Telegram no realiza ninguna validación. En la punta 1.1=4.1 y 1.2=4.2
- La app es intuitiva, pero no hay ningún tipo de ayuda o explicación de lo que consultan cada monitoreo o las alertas, y al activar la alerta SMS se cobra, monto y a quien, etc... o si los monitoreos solo se diferencian en el medio.

### ***C.3. Pauta de Evaluación Heurística respondida por Experto2***

## Pauta de Evaluación Heurística

### Introducción:

La presente pauta tiene por objetivo evaluar la usabilidad de la aplicación móvil de EmoStatus. Es importante destacar que la aplicación se encuentra en una fase intermedia, solo permitiendo la navegación y generación de acciones de forma básica, y sin datos reales. Si se desea realizar un cambio en la aplicación, por ejemplo, modificar un dato, esto será posible, pero los cambios no se guardarán.

Esta Pauta debe ser aplicada luego que UD. haya explorado y navegado detenidamente por la aplicación, con uno o más objetivos en mente.

### Antecedentes

Nombre del evaluador

\_\_\_\_\_

Edad	
<input checked="" type="checkbox"/>	Entre 20 y 25
<input type="checkbox"/>	Entre 26 y 35
<input type="checkbox"/>	Entre 36 y 45
<input type="checkbox"/>	Mayor de 45

Estudios	
<input checked="" type="checkbox"/>	Título Profesional
<input type="checkbox"/>	Postítulo
<input type="checkbox"/>	Magíster
<input type="checkbox"/>	Doctorado

Experiencia en uso de aplicaciones móviles	
<input type="checkbox"/>	Ocasionalmente
<input type="checkbox"/>	Una vez a la semana
<input type="checkbox"/>	Varios días a la semana
<input checked="" type="checkbox"/>	Todos los días de la semana

### Evaluación

I. Visibilidad del estado del sistema	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
1.1. La aplicación muestra claramente dónde se encuentra el usuario	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1.2. Las opciones posibles de explorar están claramente señaladas	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

II. Relación entre sistema y mundo real	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
2.1. El lenguaje es claro	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.2. Los conceptos utilizados son entendibles	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.3. Las palabras son de significado conocido	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.4. Los iconos generan significado	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

III. Control del usuario y libertad	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
3.1. Es fácil regresar al punto inmediatamente anterior	X				
3.2. Es fácil volver a la sección principal (lista de usuarios monitoreados) desde cualquier otra sección	X				
3.3. Provee botones o gestos para volver o dar paso a otra sección		X			
3.4. La aplicación muestra buena visibilidad al cambiar de orientación	X				

IV. Consistencia y estándares	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
4.1. La aplicación muestra claramente dónde se encuentra el usuario	X				
4.2. Las opciones posibles de explorar están claramente señaladas	X				
4.3. Existen coherencias entre el título de una sección y su contenido	X				
4.4. Sólo existe un botón que lleve a un mismo sitio	X				

V. Prevención de errores	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
5.1. Existen mensajes que prevengan posibles errores				X	
5.2. Es posible prever posibles errores				X	
5.3. La aplicación no induce a cometer errores		X			

VI. Reconocer en lugar de recordar	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
6.1. Los iconos son fácilmente reconocibles	X				
6.2. Los elementos de acción puede identificarse claramente	X				
6.3. Es posible reconocer dónde se encuentra el usuario	X				

VII. Flexibilidad y eficiencia de uso	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
7.1. Los lugares son de fácil acceso	X				
7.2. Es fácil modificar las opciones de grabación en la sección de monitoreo	X				
7.3. Es fácil encontrar información en el historial de un usuario monitoreado		X			

VIII. Estética y diseño minimalista	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
8.1. La información es relevante	X				
8.2. La cantidad de información entregada es suficiente		X			
8.3. Las secciones están correctamente organizadas		X			

IX. Reconocimiento, diagnóstico y recuperación de errores	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
9.1. Es fácil reconocer cuando ocurre un error				X	
9.2. Después de que ocurre un error es fácil volver a la sección de origen			X		
9.3. Cuando ocurre un error existen mecanismos para solucionarlos	X				

X. Ayuda y documentación	Muy de acuerdo	De acuerdo	Neutro	En desacuerdo	Muy en desacuerdo
10.1. Existe algún tipo de ayuda o indicación en la aplicación					X
10.2. Cuando existe ayuda, ésta es específica			—		
10.3 La ayuda está asequible					X

¿Cómo clasifica globalmente la aplicación?	Excelente	Bueno	Neutro	Regular	Deficiente
I. Visibilidad del estado del sistema	X				
II. Relación entre sistema y mundo real	X				
III. Control del usuario y libertad		X			
IV. Consistencia y estándares	X				
V. Prevención de errores					X
VI. Reconocer en lugar de recordar	X				
VII. Flexibilidad y eficiencia de uso		X			
VIII. Estética y diseño minimalista	X				
IX. Reconocimiento, diagnóstico y recuperación de errores				X	
X. Ayuda y documentación					X

### Justificación y Comentarios libres

- Al ingresar correo no hay validación (x@y.z)
- Pueden agregarse gestos 'swipe' para pasar de una pantalla a otra
- no hay ayuda ni documentación
  - \* Popup inicial con 'no mostrar nuevamente'
  - \* ? Botón ayuda
- En el historial agregar algo para facilitar navegación por ejemplo si quiero ver un status de hace un mes
- Atención con las opciones de acceso infrecuente en cuanto a tamaños, colores, etc.

#### ***C.4. Pauta de tareas de Paper Prototyping***

## Tarea 1: Revisar Historial

Meta/Salida	Navegar hasta el historial del Tío Pepe.
Entrada	Tío Pepe
Suposiciones	Se debería presionar Tío Pepe directamente para llevarte a sus opciones.
Pasos	Debería presionar sobre Tío Pepe y esto debería llevarlo al menú de opciones, luego de eso presiona historial y llega a la sección.
Instrucciones para el usuario	Entre a la aplicación, busque al Tío Pepe en la lista de usuarios monitoreados y revise su historial.
Preguntas para el usuario	¿Es clara la información que ve en el historial? En caso de no serla, ¿qué mejoraría? ¿Le parece suficiente la información entregada? En caso de no serla, ¿qué otra información agregaría? ¿Le agrada la iconografía utilizada para definir los estados anímicos? Si no le agrada, qué otro tipo de iconografía utilizaría para representar la misma información.
Notas	

## Tarea 2: Configurar grabación

Meta/Salida	Configure la opción de grabación de monitoreo.
Entrada	Domingo seleccionado, y martes deseleccionado en días de la semana. 20: 30 hora fin.
Suposiciones	Debería activar configuración de monitoreo de grabación antes de ingresar datos.
Pasos	Volver a opciones de Tío Pepe, seleccionar monitoreo, activar grabación y luego configurar días de la semana y hora fin.
Instrucciones para el usuario	Vaya a la sección de monitoreo del Tío Pepe, configure sus días de la semana, añadiendo el Domingo, quitando el Martes, y modifique la hora fin a las 20:30.
Preguntas para el usuario	¿Se nota que las opciones de grabación están inhabilitadas por estar la opción de monitoreo por grabación apagada? ¿Es apropiada la manera de seleccionar los días de la semana? En caso de no ser apropiada, ¿qué otra forma de selección utilizaría?
Notas	

### Tarea 3: Modificar correo para alerta

Meta/Salida	Ingresar una nueva dirección de correo para alerta por correo
Entrada	Nuevo correo electrónico
Suposiciones	Debería ingresar nuevo correo en textfield y guardarlo antes de volver atrás.
Pasos	Ir a la sección de alertas y allí seleccionar correo electrónico. Luego modificar textfield y guardar.
Instrucciones para el usuario	Vaya a la sección de alertas y seleccione la opción correo electrónico, una vez allí, ingrese un nuevo correo.
Preguntas para el usuario	
Notas	

#### Tarea 4: Ingresar repetición de notificación

Meta/Salida	Modificar frecuencia de repetición de notificación de EmoStatus
Entrada	30 min de repetición
Suposiciones	Debería seleccionar el dropdown, y luego allí seleccionar 30 min.
Pasos	Volver a sección alertas y seleccionar EmoStatus notificación. Una vez adentro seleccionar 30 min en el dropdwon.
Instrucciones para el usuario	Ingrese 30 min en la frecuencia de repetición de notificación de EmoStatus.
Preguntas para el usuario	¿Es apropiada la forma de elegir los minutos de la frecuencia de repetición? ¿Usaría otro elemento para seleccionarla? ¿cuál?
Notas	

## Tarea 5: Elegir diseño de opciones de usuario monitoreado

Meta/Salida	Comparar opciones de diseño para usuario monitoreado
Entrada	Interfaz con tabs e interfaz con lista
Suposiciones	
Pasos	
Instrucciones para el usuario	Analice ambas interfaces y elija la que le parezca más apropiada
Preguntas para el usuario	¿Cuál de las dos interfaces le parece más apropiada? ¿por qué?
Notas	

### ***C.5. Pauta de Observación***

## Pauta de Observación Usuario Final

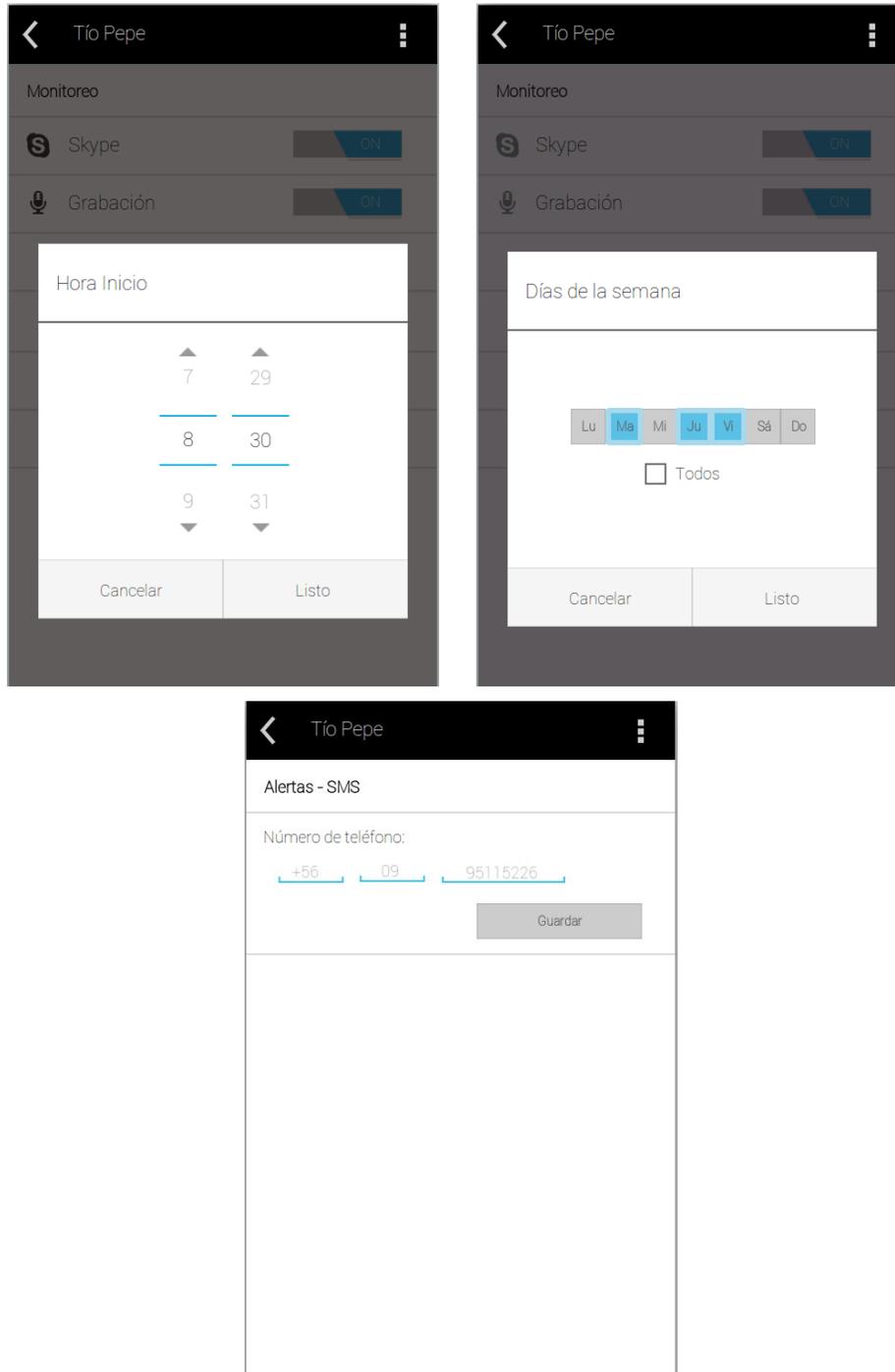
Información			
Nombre			
Sexo	F	M	
Edad			
Experiencia con aplicaciones móviles	Alta	Media	Baja

Parámetro a observar	Sí	No	Regular	Intentos	Comentarios
<b>Tarea 1: Leer historial de estados de hoy del Tío Pepe</b>					
Selecciona correctamente al Tío Pepe					
Realiza correctamente la tarea					
<i>Hora de inicio</i>					
<i>Hora de término</i>					
<i>Tiempo</i>					
<b>Tarea 2: Revisar si el Tío Pepe estuvo triste la semana pasada</b>					
Se mantiene en la misma sección					
Usa el scroll para buscar más abajo en el historial					
Realiza correctamente la tarea					
<i>Hora de inicio</i>					
<i>Hora de término</i>					
<i>Tiempo</i>					
<b>Tarea 3: Desactivar monitoreo por Skype para la Abuela Lorena</b>					
Vuelve a la lista de usuarios monitoreados y selecciona a la Abuela Lorena					
Entra a la sección de monitoreo inmediatamente					
Desactiva la opción correctamente					
<i>Hora de inicio</i>					
<i>Hora de término</i>					
<i>Tiempo</i>					
Parámetro a observar	Sí	No	Regular	Intentos	Comentarios
<b>Tarea 4: Activar opción Grabación en monitoreo para la Abuela Lorena, y cambiar los días de la semana seleccionados</b>					
Se mantiene en la misma sección					

Parámetro a observar	Sí	No	Regular	Intentos	Comentarios
Activa la opción de Grabación antes de modificar los días					
Selecciona y de-selecciona días de la semana correctamente					
<i>Hora de inicio</i>					
<i>Hora de término</i>					
<i>Tiempo</i>					
<b>Tarea 5: Cambiar el número de teléfono para alertar por SMS, para el usuario Tío Pepe</b>					
Vuelve a la lista de usuarios monitoreados y selecciona Tío Pepe					
Entra a la sección de alertas inmediatamente					
Cambia correctamente el número de teléfono					
<i>Hora de inicio</i>					
<i>Hora de término</i>					
<i>Tiempo</i>					
<b>Tarea 6: Activar las notificaciones de la aplicación como alerta</b>					
Se mantiene en la misma sección					
Activa la opción correctamente					
<i>Hora de inicio</i>					
<i>Hora de término</i>					
<i>Tiempo</i>					
<b>Tarea 7: Revisar el historial de algún día que tenga datos, en el calendario</b>					
Presiona el icono de calendario					
Selecciona opción de ver en calendario					
Selecciona un día habilitado					
<i>Hora de inicio</i>					
<i>Hora de término</i>					
<i>Tiempo</i>					

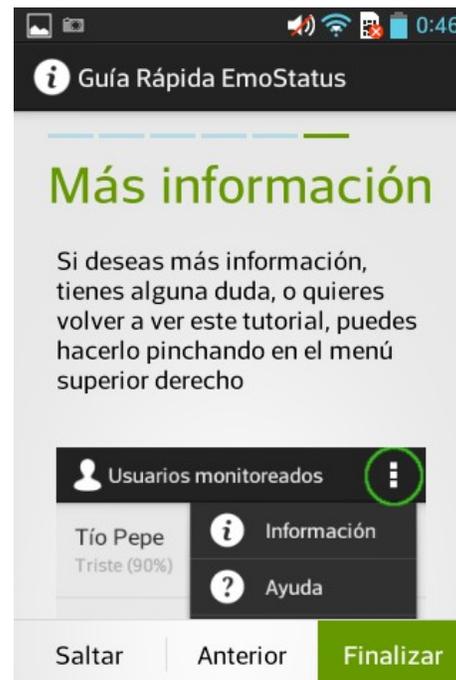
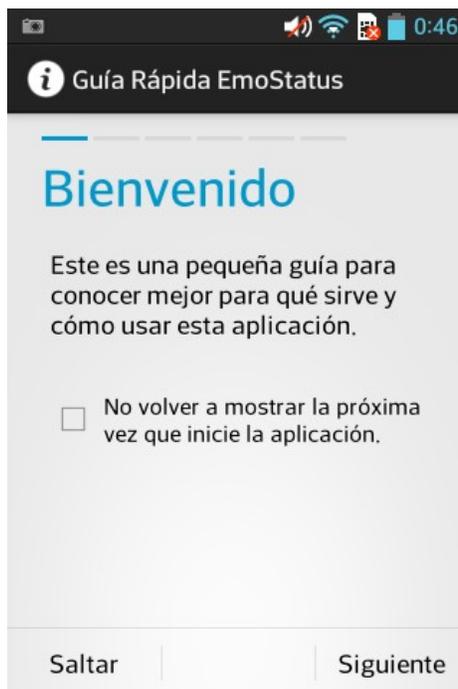
## Anexo D: Mockups diseñados

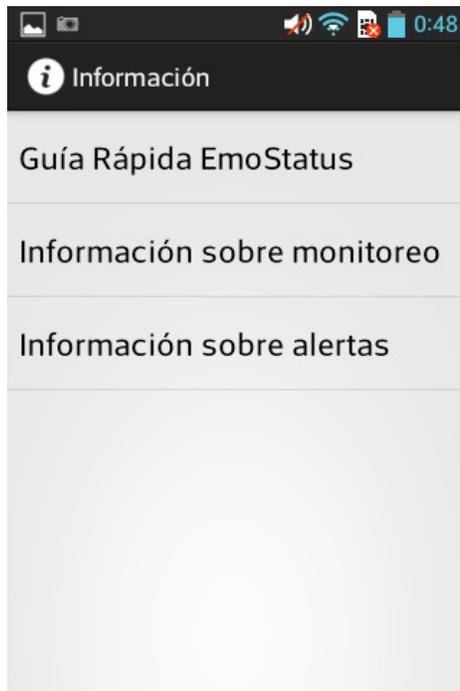
Se ven a continuación las interfaces de elección de hora y elección de días en sección Monitoreo, y la interfaz de ingreso de número telefónico en sección Alertas.



## Anexo E: Pantallas de Segundo Prototipo de la Aplicación Móvil

Las siguientes imágenes muestran pantallas no incluidas en el trabajo. Éstas son las interfaces de la Guía Rápida, sección Información e ingreso de correo en sección Alertas, en ese orden.





## Anexo F: Diagramas de Clases de Aplicación Móvil

En el trabajo se mostró el diagrama de clases del paquete *com.thesis.emostatus*. Las siguientes imágenes son los diagramas de clases de los paquetes *adapters* y *persistance*.

adapters

