



Universidad de Valladolid

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS INDUSTRIALES
DEPARTAMENTO DE INGENIERÍA DE SISTEMAS Y AUTOMÁTICA

TRABAJO FIN DE MÁSTER:

DISEÑO Y CONSTRUCCIÓN DE UNA CABEZA MECATRÓNICA DE ASPECTO REALISTA

Trabajo fin de máster presentado por:
David César Loza Matovelle
para el grado de Máster en Procesos y Sistemas
entregado por la Universidad de Valladolid

Dirigida por:
Dr. D. Eduardo Zalama Casanova
Dr. D. Jaime Gómez García-Bermejo

Valladolid, 2013

Dedicado a mi familia

Agradecimientos

Hay muchas personas e instituciones que se deben agradecer en un trabajo. Primero a Dios porque sin él nada fuera posible. A mi país, la República del Ecuador, que a través de la Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación (Senescyt) me ha dado la oportunidad de realizar mis estudios en el exterior.

También deseo agradecer al Ministerio de Ciencia e Innovación y la Junta de Castilla y León por soportar el trabajo realizado.

Deseo agradecer a mis directores del proyecto fin de máster D. Eduardo Zalama Casanova y D. Jaime Gómez García-Bermejo por permitirme ser parte del proyecto, su constante seguimiento y ayuda.

A D. José Luis González por el excelente trabajo en la construcción de la cabeza mecatrónica.

Agradecer a los compañeros del área de robótica: Samuel Marcos, Salvador Domínguez, Jaime Pulido, Roberto Pinillos, Sergio Herguedas y Raúl Feliz por todo el tiempo que se tomaron en ayudarme y sus importantes contribuciones en el desarrollo del proyecto.

A Dña. María Jesús de la Fuente coordinadora del máster, que tanto en conocimientos y otros menesteres me ha ayudado, creo que sin su invaluable ayuda no se podría haber realizado mis estudios.

A toda mi familia que a pesar de la distancia siempre me han sabido alentar y saber que se preocupan de mí. En especial a mi madre que sin su ayuda creo que sería muy difícil alcanzar esta meta.

A todos mis amigos y personas que no he nombrado pero que me han ayudado en el día a día, es muy difícil adaptarse y creo que sin su ayuda hubiera sido imposible llevar a feliz término este proyecto.

Muchas gracias a todos.

*David Loza
Valladolid, Julio 2013*

Resumen

En el trabajo actual se incorporan elementos conocidos en la robótica social y nuevas ideas para contribuir en el desarrollo de una mejor comunicación. El objetivo del trabajo es el desarrollo de una cabeza mecatrónica realista con inspiración en el sistema neuromuscular facial humano y altas capacidades sociales. Es decir, que proporcionen al usuario que interactúa con el robot la misma cantidad de información visual que la que provoca el rostro de un interlocutor humano.

La primera parte del trabajo habla acerca de la fisiología de la cabeza humana, tanto huesos como músculos principales. También se habla acerca del Sistema de Codificación de Acciones Faciales (FACS) como herramienta para analizar los gestos. Se detallan, las simplificaciones hechas al FACS para el diseño propuesto. El número de unidades de acción implementadas son 32.

También, en esta parte se detalla los aspectos mecánicos, la estructura de la cabeza mecatrónica y como se ha basado el desarrollo en el FACS, se detalla la posición de los servos, así como los mecanismos de los ojos, cuello y boca. Se han obtenido 23 grados de libertad a partir del accionamiento de un conjunto de servomotores.

La segunda parte trata acerca del control de expresiones emocionales así como los módulos desarrollados. El control de expresiones emocionales está integrado por tres módulos: módulo de control, módulo de unidades de acción y el módulo de expresiones. También, se describe el software desarrollado como la interface de usuario que permite interactuar con los módulos de control.

Además, se explica cómo se han implementado los diferentes módulos de interacción social. Se describen los módulos de visión, conocimiento, habla, joystick, seguimiento y operación remota. Se detalla la estructura de los mismos, sus características principales, el modo de funcionamiento, así como la arquitectura implementada en el robot.

Por último, se expone la evaluación de la cabeza mecatrónica por parte de diferentes usuarios en relación a la percepción de las diferentes expresiones emocionales y su desempeño, evaluando las posibilidades de estos tipos de

dispositivos como elementos de interfaz hombre máquina y que pueden ser de utilidad en: puntos de información, sistemas de atención al usuario, recepcionistas, etc. y los medios que se utilizaron en la misma, se analiza desde diferentes puntos de vista el desempeño de la cabeza mecatrónica. El trabajo finaliza con las conclusiones así como las líneas futuras de investigación del proyecto.

Índice general

Agradecimientos	III
Resumen	IV
Lista de figuras	IX
Lista de tablas	XI
1. Introducción	12
1.1. Motivación y problemática	12
1.2. Objetivos	14
1.3. Robots sociales	14
1.3.1. Concepto	14
1.3.2. Características	15
1.3.3. Antecedentes	16
1.3.4. Clasificación de los Robots Sociales	17
1.4. Estructura del Documento	18
2. La expresión humana y la robótica	20
2.1. Anatomía de la cara y la cabeza, detrás de la expresión humana	20
2.1.1. Huesos de la cabeza	20
2.1.2. Musculatura del rostro	22
2.2. FACS - Facial Action Code System	23
2.2.1. Concepto	23
2.2.2. Características	23
2.3. Antropomorfismo deseado en un robot.	26
2.3.1. Opiniones acerca del grado de antropomorfismo de un robot	26
2.3.2. El valle inexplicable	27
2.4. Del rostro humano a un modelo mecánico.	29
2.5. Conclusiones	31

3. Diseño mecánico y control de expresiones	32
3.1. Diseño mecánico de la cabeza mecatrónica siguiendo el FACS .	32
3.1.1. Ubicación de servomotores.	33
3.1.2. Modelo Mecánico	35
3.1.3. Mecanismos Especiales	36
3.1.4. Modelo Terminado	39
3.2. Programas utilizados en el proyecto	39
3.2.1. ROS - Robotic operative system	39
3.2.2. Python	40
3.2.3. OpenCV	41
3.3. Control de expresiones	41
3.3.1. Módulo controlador de servos	42
3.3.2. Módulo de unidades de acción	43
3.3.3. Modulo gestor de emociones	45
3.4. Operación remota	46
3.5. Animación de expresiones	47
3.5.1. Estructura de fotogramas	47
3.5.2. Integración de unidades de acción y expresiones básicas	48
3.6. Conclusiones	51
4. Módulos agregados	52
4.1. Visión	52
4.1.1. Detección de movimiento	53
4.1.2. Detección de rostros	54
4.1.3. Detección de objetos	56
4.1.4. Flujo óptico	58
4.2. Seguimiento de rostros y objetos	60
4.3. Habla	62
4.3.1. Generación automática de visemas	63
4.3.2. Integración de visemas	64
4.4. Conocimiento	65
4.5. Arquitectura desarrollada	66
4.6. Conclusiones	68
5. Resultados y evaluación del proyecto.	70
5.1. Resultados de los diferentes elementos en el robot	70
5.1.1. Unidades de acción	70
5.1.2. Expresiones básicas	71
5.1.3. Animación de expresiones	72
5.2. Evaluación de la cabeza mecatrónica	73
5.2.1. Participantes	74

5.2.2.	Descripción de la prueba	74
5.2.3.	Resultados	75
5.2.4.	Discusión	78
6.	Conclusiones y futuros desarrollos	80
6.1.	Conclusiones	80
6.2.	Trabajos futuros	82
6.3.	Publicaciones	83
A.	Técnica de modulación por ancho de pulso (MAP)	84
B.	Estructura en ROS del robot	86
B.1.	Estructura en ROS	86
B.2.	Estructura del robot en ROS	87
C.	Formato de la encuesta realizada	90
	Bibliografía	93

Índice de figuras

1.1. Robots de CARTIF	13
1.2. Robots Humanoides	17
2.1. Estructura ósea de la cabeza	21
2.2. Musculatura de la cabeza humana	22
2.3. Ejemplos de Unidades de Acción	24
2.4. Grupo de expresiones básicas universales. De izquierda a derecha y de arriba a abajo: miedo, tristeza, sorpresa, disgusto, alegría y enfado	25
2.5. Intensidad de acuerdo al FACS	25
2.6. Ejemplos de Intensidad	26
2.7. Ejemplo de asimetría	26
2.8. Valle inexplicable propuesto por Mori [1]	28
3.1. Modelo CAD de la cabeza mecatrónica	33
3.2. Modelo Mecánico: (a). Frontal (b). Mandíbula superior (c). Mandíbula inferior	35
3.3. (a) Músculos principales (b) Mecanismo del ojo.	36
3.4. Mecanismo del cuello.	38
3.5. Estructura de la boca.	38
3.6. Vista Frontal y Lateral de la cabeza terminada.	39
3.7. Sistema de control de expresiones.	41
3.8. Tarjeta SSC32.	42
3.9. a) Factores de intensidad implementadas en el FACS (b) Factor de intensidad.	44
3.10. Ejemplo de ejecución durante el tiempo de un servo	45
3.11. Estructura de fotogramas	47
3.12. Programa para generar expresiones - Unidades de acción.	48
3.13. Ejemplos de expresiones de la base.	49
3.14. Programa para generar expresiones - Expresiones básicas.	49
4.1. Esquema de la detección de movimiento	53

4.2.	Detección de movimiento	54
4.3.	Flujograma de la detección de rostros	55
4.4.	Detección de rostros	56
4.5.	Flujograma de la detección de objetos	57
4.6.	Detección de objetos	58
4.7.	Flujo óptico	60
4.8.	Flujograma del tracking	61
4.9.	Visemas en español	62
4.10.	Esquema del proceso de habla en el robot	63
4.11.	Programa para generar expresiones - Integración de visemas	65
4.12.	Arquitectura modular	66
4.13.	Máquina de estados	67
4.14.	Interface de usuario para operación remota	68
5.1.	Diferentes unidades de acción	71
5.2.	Diferentes expresiones obtenidas con la cabeza mecatrónica. (a) Neutral (b) Disgusto (c) Enfado (d) Felicidad	72
5.3.	Secuencia animada	73
5.4.	Expresiones básicas al cincuenta por ciento de intensidad frente al porcentaje de expresiones reconocidas por los usuarios.	76
5.5.	Expresiones básicas al cien por ciento de intensidad frente al porcentaje de expresiones reconocidas por los usuarios.	77
A.1.	Técnica de modulación por ancho de pulso	85
B.1.	Estructura en ROS del robot	89

Índice de tablas

2.1. Huesos relevantes del rostro humano.	21
2.2. Músculos relevantes del rostro humano.	22
2.3. AU en las expresiones universales.	24
2.4. Músculos a representar con sus correspondientes expresiones emocionales.	30
2.5. Unidades de acción agregadas.	31
3.1. Músculos asociados a los diferentes servomotores.	34
3.2. Características de los servomotores.	34
3.3. Diferentes funciones del cuello y sus músculos.	37
3.4. Unidades de Acción asociadas a cada servo	43
5.1. Análisis de la muestra	74
5.2. Frecuencia y porcentaje de expresiones reconocidas al cincuenta por ciento de intensidad	75
5.3. Frecuencia de reconocimiento de expresiones al cien por ciento de intensidad	76
5.4. Analisis anova entre grupos de expresiones	77
5.5. Promedio y desviaciones estándar para cada variable analizada	78

Capítulo 1

Introducción

“Hay que manifestar los sueños para que puedan comenzar a realizarse.”

– *Paulo Coelho*

El primer capítulo describe los robots sociales. Se habla acerca de las ventajas que tiene el desarrollo de robots con apariencia humana. También se definen el objetivo general, como los objetivos secundarios del trabajo. Se resaltan las principales ideas del escrito y algunos de los temas tratados aquí serán desarrollados en los próximos capítulos del trabajo fin de máster. Por último, se describe la estructura de la presente memoria.

1.1. Motivación y problemática

Una de las principales formas de comunicación del ser humano son sus expresiones. En [2] se sostiene que el rostro de una persona crea las señales más obvias e inmediatas a la formación de impresiones. Por lo tanto, si se desea mejorar el desempeño de un robot que interactúe con los humanos, debería ser capaz de dar la información visual necesaria para que el mensaje sea perfectamente entendido por el usuario. El desarrollo de robots con alto grado de realismo tiene su origen en varios aspectos:

- Posibilitar una interacción más simple y natural, de forma que se mejore la adaptación del robot a las necesidades del usuario.
- Mejorar las capacidades y prestaciones de los robots dentro del hogar. Es cada vez más frecuente que los robots se involucren en tareas cotidianas las cuales demandan un mayor número de prestaciones por parte de los robots hacia sus usuarios.



Figura 1.1: Robots de CARTIF

- Mejorar el uso de nuevas tecnologías a personas con dificultades físicas.
- Permitir el desarrollo de nuevas herramientas para evaluar la interacción humana.

Este proyecto se ha desarrollado en la Fundación Cartif, dentro de la división de Robótica, en la que se cuenta con una amplia experiencia en el desarrollo de robots sociales y de servicio. Trabajos previos realizados en Cartif, como son el robot Arisco [3] y el Avatar Pedro [4] han motivado a profundizar la investigación de la interacción hombre-máquina. En la figura 1.1, se muestran una serie de robots desarrollados en Cartif.

En este trabajo se han realizado una serie de aportaciones en diferentes áreas para lo que se han estudiado diversas estrategias de socialización, implementando y mejorando algunas ya desarrolladas

1.2. Objetivos

El objetivo principal es **incorporar capacidades sociales necesarias para una correcta interacción con las personas en una cabeza mecatrónica de apariencia realista.**

Teniendo en cuenta un amplio objetivo principal sea subdividido en diferentes objetivos secundarios:

- Analizar las aproximaciones mecánicas que se pueden hacer de un rostro humano, cuales son las más adecuadas y representativas.
- Diseñar una arquitectura modular basada en el FACS (*Facial Action Code System*) la cual permita dar una flexibilidad en el desarrollo de expresiones y potenciar las características gestuales del robot.
- Basados en la arquitectura anterior, implementar diferentes módulos, añadiendo características necesarias, para una correcta interacción. Como son visión, voz, reconocimiento de objetos y rostros, etc.
- Obtener expresiones complejas y a partir de ellas generar comportamientos, que el robot pueda procesar o resolver.
- Evaluar la interacción del robot con los seres humanos a través de entrevistas y fotografías de control al interaccionar.

Como complemento a los objetivos anteriores, se han propuestos objetivos secundarios adicionales:

- Revisar el estado del arte en la robótica social, investigar las aportaciones recientes en este campo.
- Analizar las diferentes investigaciones realizadas en CARTIF con respecto a la robótica social, así como los resultados obtenidos.

1.3. Robots sociales

1.3.1. Concepto

Existen diferentes conceptos acerca del término robot social. Duffy [5] define a un robot social, como un robot capaz de interaccionar con los seres humanos y otros robots. Todo robot social es constituido por 4 elementos

esenciales. La parte física que es la estructura del robot, sobre la cual se encontrarán los diferentes elementos como son bastidores y mecanismos necesarios. La parte reactiva conformada por los sensores y actuadores del robot. La parte deliberativa enmarca el grado de decisión del robot. Finalmente, la parte social se comunica con el alto nivel en su entorno.

De acuerdo a [6] los robots sociales son agentes que forman parte de un grupo heterogéneo: un grupo de robots o seres humanos. Son capaces de reconocerse entre sí y participar en la interacción social. Poseen un historial (perciben e interpretan el mundo en términos de su propia experiencia), se comunican explícitamente y aprenden unos de otros.

Breazeal [7] en su libro define a un robot social, como un robot capaz de comunicarse con alto nivel. Entiende y se comunica adecuadamente con los humanos, se adapta y aprende para mejorar sus estrategias sociales.

De acuerdo a Bartneck y Forlizzi [8] un robot social funciona de manera autónoma e interacciona con humanos siguiendo un patrón de comportamiento esperado.

1.3.2. Características

En el campo de la robótica existe actualmente un creciente interés, por el desarrollo de robots con la naturalidad requerida para interaccionar con los humanos, este nivel de interacción sería útil para realizar diferentes tareas, utilizando estrategias similares a las usadas por los humanos.

Al fin de interaccionar de manera adecuada, los robots sociales poseen algunas características particulares. En [9] se especifican algunas características:

- Expresar y percibir emociones.
- Entender y reconocer los modelos de otros agentes.
- Ser capaces de comunicarse con las personas manteniendo diálogos de alto nivel.
- Establecer y mantener relaciones sociales.
- Usar un lenguaje natural (gestos, expresiones, etc.).
- Tener una personalidad y características propias.
- Aprender y desarrollar competencias sociales.

Se debe tener en cuenta que la interacción con las personas, es un tema difícil, que depende de muchos factores sociales, culturales, psicológicos inclusive regionales y que tanto el desarrollo como sus características particulares estarán subscritas al área en donde se lo va emplear.

1.3.3. Antecedentes

Aunque los comienzos de la robótica datan de los años 50 con un robot utilizado en General Motor llamado *Unimate* y que servía en sus líneas de ensamblaje. No es hasta la década de los 80 donde Rodney Brooks con una serie de artículos [10], [11] sobre inteligencia artificial, da origen a una nueva filosofía denominada “*subsumption architecture*”, la cual tiene como inspiración la biología. En esta filosofía los robots tienen funciones muy sencillas con un objetivo particular. Estos principios influenciaron a muchos investigadores para profundizar en diferentes ámbitos, entre ellos la robótica social.

En la década de los 90 se presentaron varias contribuciones a la robótica social. En [12] se analizan robots biológicamente inspirados, estos emulan comportamientos de animales como abejas y hormigas, con funciones muy sencillas individualmente pero logran un objetivo común.

Otra contribución importante fue el robot Kismet [13] realizado en el MIT por la Dr. Cynthia Breazeal. Este robot poseía altas capacidades gestuales además de poder ver y oír a las personas con las que interacciona. Los primeros proyectos tenían aproximaciones a rostros bastante caricaturescas, con orejas, boca y ojos grandes.

Otro ejemplo relevante es el robot WE-4R Waseda [14] con un sistema mecánico sencillo en la cabeza, pero con la ayuda del cuerpo, manos y torso es capaz de realizar expresiones complejas. Estas primeras propuestas a pesar de tener aceptación, gran expresividad y un sistema mecánico sencillo están limitadas por una reducida capacidad gestual. Además que su apariencia no es natural.

Uno de los primeros proyectos con apariencia humana es Roman [15] desarrollado en el RRLab aunque sus expresiones adolecen de un realismo significativo, posee un sistema emocional central interesante.

Para el año 2000 Honda introduce un robot de forma humanoide Asimo, fruto de dos décadas de investigación (en el año 1986 se creó un prototipo bípedo denominado E0). Asimo es un robot humanoide de 130 cm de alto, pesa unos 54 Kg y posee 34 grados de libertad, es capaz de subir escaleras, correr, reconoce caras de personas y realizar pequeñas tareas.

Actualmente existen muchas propuestas, por ejemplo, el humanoide HRP-4C [16] tiene una altura de 1.58 m, pesa 43 Kg, y tiene 30 grados de libertad.



Figura 1.2: Robots Humanoides

Posee la apariencia de una joven japonesa, camina de manera natural y ha realizado coreografías, su costo aproximado es de 200.000 dólares.

Kokoro de la Universidad de Osaka, relacionada con los proyectos DER-2 y DER-3. De apariencia humana puede hablar, mover su cuerpo de manera natural. El robot Cub (en este proyecto están involucradas 11 naciones europeas y es de open hardware) tiene la apariencia de un niño y posee 53 grados de libertad, un alto de 104 cm, y un peso de 22 Kg. En la figura 1.2 se puede ver un ejemplo de los robots humanoides desarrollados actualmente.

Todas estas propuestas tienen una estética similar a un ser humano y tienen un alto realismo. Además, de integrar altas capacidades sociales en su desarrollo.

1.3.4. Clasificación de los Robots Sociales

De acuerdo a [9] los robots sociales se clasifican en: antropomórficos, zoomórficos, caricaturesco y funcionales.

- Los robots antropomórficos tienen una apariencia similar a un ser humano, en este paradigma el robot se crea buscando un equilibrio entre capacidades sociales y aspecto. Algunos ejemplos son Kokoro y HRP4C.
- Los robots Zoomórficos imitan las criaturas vivas, como ejemplo tenemos al AIBOT desarrollado por Sony, un perro robot que interacciona

con sus dueños. De la misma tendencia el iCat, un robot con apariencia de gato que juega con sus dueños.

- Lo robots con apariencia caricaturesca poseen rasgos diferenciados. A pesar de tener un bajo realismo estos robots puede obtener la atención por su peculiar aspecto por ejemplo Kismet y Sacarino.
- Los robots funcionales son desarrollados con un fin específico. En donde sus capacidades sociales le servirán para poder interaccionar de mejor forma y dar un mejor servicio.

En el mismo artículo hace referencia a otro criterio de clasificación por la personalidad exhibida de los robots.

- Juguetes, estos robots únicamente presentan un servicio específico, además de realizar ciertas rutinas asociadas a su construcción.
- Mascotas, exhiben comportamientos asociados a animales reales como perros y gatos.
- Caricaturesca, su conducta es exagerada la cual facilita la interacción con las personas.
- Artificial son robots con rasgo propios en su conducta y estarán íntimamente relacionada con la función que realicen.
- Humana, son los robots que presentaran una conducta y comportamiento similar a los humanos.

1.4. Estructura del Documento

El documento se estructura en seis capítulos. El primer capítulo nos da un visión general de la robótica social, define los objetivos planteados en el proyecto. El segundo capítulo detalla los principales huesos y músculos de la cara así como la caracterización por medio del FACS, se analiza el valle inexplicable formulado por Mori así como sus implicaciones, finalmente se analizan las simplificaciones realizadas en el trabajo propuesto.

El tercer capítulo aborda el diseño mecánico del robot así como las aproximaciones de la fisiología humana que se hicieron para representarlo. También en este capítulo se explica la arquitectura utilizada y se describe cada uno de los diferentes módulos del sistema gestual, se describe cada uno de ellos, su fundamento, así como sus principales características.

El capítulo cuatro describe los módulos agregados para dar naturalidad a la cabeza como son la visión, reconocimiento de objetos, habla, detección de movimiento, operación remota y la máquina de estados; como en el caso anterior se describe su fundamento y sus características principales.

En el capítulo cinco habla de la evaluación realizada al modelo, las pruebas, niveles de desempeño alcanzado y los resultados obtenidos.

Por último, en el capítulo seis, se presentan las conclusiones obtenidas del trabajo global, se indican las futuras líneas de investigación y desarrollo.

Capítulo 2

La expresión humana y la robótica

“La raza humana tiene un arma verdaderamente eficaz: la risa.”

– *Mark Twain*

El capítulo analiza la fisiología del rostro humano, se describe el FACS (*Facial Action Code System* o Sistema de Codificación de Acciones Faciales). Se analiza el antropomorfismo en los robots así como el valle inexplicable. Por último, las simplificaciones hechas a la expresión humana en el diseño propuesto.

2.1. Anatomía de la cara y la cabeza, detrás de la expresión humana

El estudio de la anatomía de la cara y la cabeza es importante para la comprensión de aspectos y funciones de la misma. Este estudio debe servir como fuente de inspiración del diseño de electromecánico de la cabeza, pues se trata de emular el comportamiento neuromuscular facial humano.

2.1.1. Huesos de la cabeza

El cráneo humano y los huesos de la cara sirven de base para la musculatura facial además de proteger órganos importantes como son el cerebro, ojos, lengua, etc. Gran parte de la apariencia de la cara humana depende de la forma de estos huesos. Además, conforman las fosas nasales y permite la apertura y cierre de la cavidad bucal.

El más importante de estos huesos craneales para la apariencia del rostro es el hueso frontal. El hueso frontal es la base de la parte superior de la cara por encima de los globos oculares. Los huesos de la cara se pueden dividir en dos porciones, llamadas mandíbulas. La superior, formada por un solo hueso, el maxilar superior. La inferior, más complicada, formada por 6 huesos pares, maxilar inferior, malar, unguis, cornete inferior, hueso propio de la nariz, y palatino, y uno impar, el vómer.

El cráneo también incluye 14 huesos faciales. Estos huesos conforman la parte delantera inferior del cráneo y constituyen el marco para la mayor parte de la cara (Ver tabla 2.1).

Tabla 2.1: Huesos relevantes del rostro humano.

Hueso	Cantidad
Frontal	1
Palatino	2
Vómer	1
Lagrimal	2
Cornete interior	2
Cigomático	2
Maxilar superior	2
Maxilar inferior	1
Hueso nasal	2

La figura 2.1 muestra la vista lateral y frontal del cráneo humano.

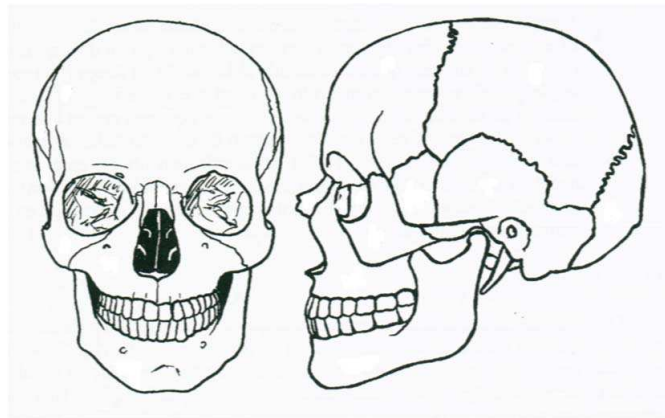


Figura 2.1: Estructura ósea de la cabeza

2.1.2. Musculatura del rostro

La estructura muscular del rostro se encuentra directamente relacionada con las expresiones. Principalmente nos centraremos en los músculos que dan a entender cómo la cara se mueve. A continuación se enumeran los músculos principales del rostro humano y el área a la cual pertenecen (Ver tabla 2.2).

Tabla 2.2: Músculos relevantes del rostro humano.

Área	Músculos más significativos
Frente y cejas	Occipital y frontal
Ojos y párpados	Elevador palpebral superior, orbicular del ojo, corrugador superciliar
Labios y boca	Cuadrado del labio superior e inferior, canino, triangular, cigomáticos mayor y menor, bucinador, orbicular de la boca, risorio, mirtiforme
Nariz	Piramidal, transverso, dilatador de las fosas nasales, elevadores común y propio, transverso
Orejas	Auriculares anterior superior e inferior

La musculatura facial se puede subdividir en dos grandes grupos, los de la expresión y los de la masticación, la figura 2.2 representa la estructura muscular facial.



Figura 2.2: Musculatura de la cabeza humana

2.2. FACS - Facial Action Code System

2.2.1. Concepto

La anatomía de la cabeza analiza elementos importantes aunque únicamente puede obtenerse información de la contracción producida. Por lo tanto, es necesario determinar la inferencia existente entre las expresiones faciales, lo que dichas expresiones pretenden representar y lo que otros son capaces de interpretar a partir de las expresiones generadas.

Se han desarrollado estándares que permitan codificar y parametrizar de forma objetiva los movimientos faciales. Con el fin de relacionar las emociones y las expresiones faciales a partir de patrones específicos de movimientos en el rostro. Entre los estándares más conocidos están: EMFACS [17](Ekman y Friesen, 1982), *Monadic Phases* [18](Tronick, Als, y Brazelton, 1980), *the Maximally Discriminative Facial Movement Coding System* [19] (MAX: Izard, 1979), y *Facial Action Coding System* [20] (FACS: Ekman y Friesen, 1978).

Una de las técnicas más extendidas es el Sistema de Codificación de Acciones Faciales (*Facial Action Coding System*, FACS) [20], es el método más ampliamente utilizado y versátil para medir y describir comportamientos faciales. Ekman y Friesen desarrollaron el FACS originalmente en la década de los setenta en base de los estudios de Carl-Herman Hjortsjö [21]. Determinaron la contracción de cada músculo facial (individualmente y en combinación con otros músculos) y el cambio en la apariencia del rostro. También, examinaron los videos de conducta facial para identificar los cambios específicos y la mejor forma de diferenciar una de otra. Su objetivo era crear un medio confiable para categorizar el comportamiento facial.

2.2.2. Características

El FACS se estructura por unidades de acción (*Action Units* AU), una AU es una contracción o relajación de determinados músculos que producen un cambio en la apariencia del rostro. El FACS no se basa en el movimiento de cada músculo, esto se debe a que en algunos gestos un determinado músculo aparece más de una vez y también debido a que los cambios en la apariencia producidos son relativamente independientes al movimiento de un solo músculo. En la figura 2.3 se puede ver algunos ejemplos de unidades de acción.

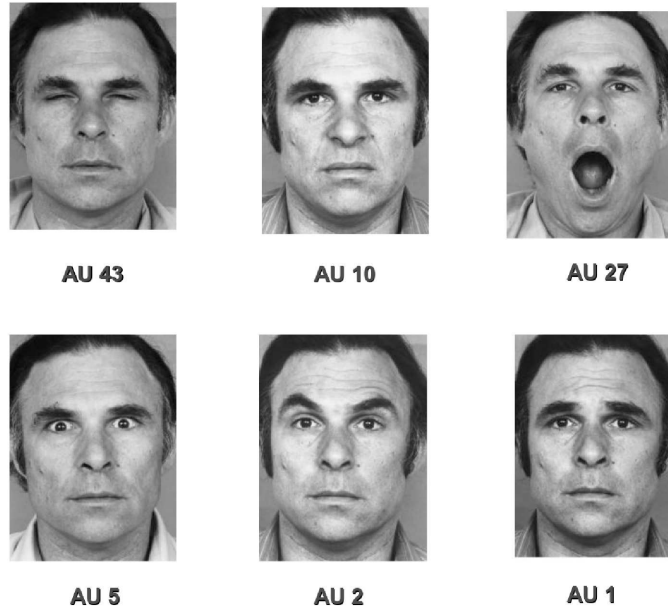


Figura 2.3: Ejemplos de Unidades de Acción

De acuerdo a los trabajos de Ekman existen 6 expresiones universales: miedo, ira, tristeza, asco, sorpresa, alegría y desprecio. Cada una quedaría definida (no completamente como veremos más adelante) por un número determinado de unidades de acción. En la tabla 2.3 se listan las expresiones universales con sus respectivas unidades de acción.

Tabla 2.3: AU en las expresiones universales.

Expresión Universal	Unidades de acción
Felicidad	6+12
Tristeza	1+4+15
Sorpresa	1+2+5+26
Miedo	1+2+4+5+20+26
Ira	4+5+7+23
Disgusto	9+15+16

En la figura 2.4 se aprecian las expresiones universales.



Figura 2.4: Grupo de expresiones básicas universales. De izquierda a derecha y de arriba a abajo: miedo, tristeza, sorpresa, disgusto, alegría y enfado

Una expresión no consiste únicamente en un número determinado de AUs. El FACS también analiza elementos como duración, intensidad, y la asimetría. El FACS cuantifica cinco niveles de intensidad A, B, C, D y E, siendo A el nivel más bajo y E el nivel más alto de intensidad de la unidad de acción. Los intervalos de intensidad no son iguales, la figura 2.5 muestra la representación gráfica de cada uno de los niveles de intensidad en el FACS.

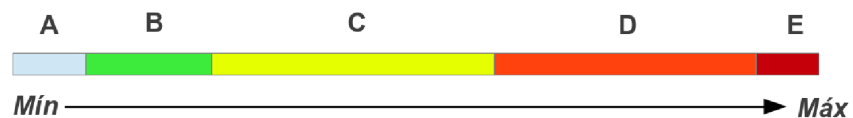


Figura 2.5: Intensidad de acuerdo al FACS

En la figura 2.6 tenemos un ejemplo de diferentes intensidades, como podemos ver en las fotos se cuenta con las mismas AUs, la diferencia entre ellas únicamente es la intensidad.



Figura 2.6: Ejemplos de Intensidad

La asimetría se produce cuando las AUs no son las mismas tomando diferentes ejes de simetría, el FACS considera con las letras L (*Left*) , R (*Right*) , T (*Top*) y B (*Botton*) las diferentes posibilidades, la figura 2.7 por ejemplo sería representada en el FACS con AU2-L.



Figura 2.7: Ejemplo de asimetría

En cuanto a la duración de las AUs esto dependerá de la expresión y el sentir, existen emociones cortas como la sorpresa, el asco, etc. y otras más duraderas como la sonrisa, etc.

El FACS ha sido ampliamente utilizado en diferentes campos, por ejemplo, en la animación [22], reconocimiento de mentiras [23], diagnóstico de algunas enfermedades [24] entre otros.

2.3. Antropomorfismo deseado en un robot.

2.3.1. Opiniones acerca del grado de antropomorfismo de un robot

En [25] se define al antropomorfismo como: “*la tendencia a atribuir características humanas a objetos inanimados, animales y otros con el fin de*

ayudarnos a racionalizar una situación”, también se habla de antropomorfismo cuando se le atribuye características de un animal tales como perros y gatos a los robots.

Por el año 2000 Brian R. Duffy analizó a través de una serie de artículos [25], [26] el grado necesario de antropomorfismo en los robots. Para la fecha asumía una perspectiva pragmática, la tendencia era ver a los robots como una máquina utilizando cualidades humanas como personalidad, gestos e incluso expresiones faciales.

Posteriormente algunos estudios acerca del grado de antropomorfismo daban resultados contrapuestos. En [27] se dice que los agentes con un mayor grado de antropomorfismo en su apariencia resultan más creíbles y son preferidos por los usuarios. En cambio en, [28] y [29] se muestra, los agentes menos antropomórficos originan percepciones más positivas y son preferidos para la interacción.

Estudios más recientes realizados por la Universidad de Cambridge [30] sobre cuatro robots con diferentes grados de antropomorfismo (2 con aspecto mecánico y 2 con un aspecto humano). Se concluyó que no se encontró un impacto significativo de acuerdo a la edad o el género. Otra conclusión fue que a un mayor grado de antropomorfismo se creaba más empatía en el usuario.

Parece que la pregunta en cuanto al grado necesario de antropomorfismo se ha convertido en analizar los casos en donde un robot es necesario que tenga un alto grado del mismo. Por ejemplo en [31] se analiza si un paciente se siente incómodo ante la presencia de un robot en un examen médico.

2.3.2. El valle inexplicable

En la década de los 70 Masahiro Mori postuló el denominado valle inexplicable o *“uncanny valley”* [1]. En él se analiza el nivel de aprobación que tienen las personas a los robots, Mori explica que mientras permanezca con bajo antropomorfismo provocara la aceptación por parte del usuario, mientras más se acerque a la apariencia humana incrementara su aceptación hasta un punto en donde el nivel de aceptación baje drásticamente para luego volver a subir pasando determinado rango. El concepto se representa en la figura 2.8.

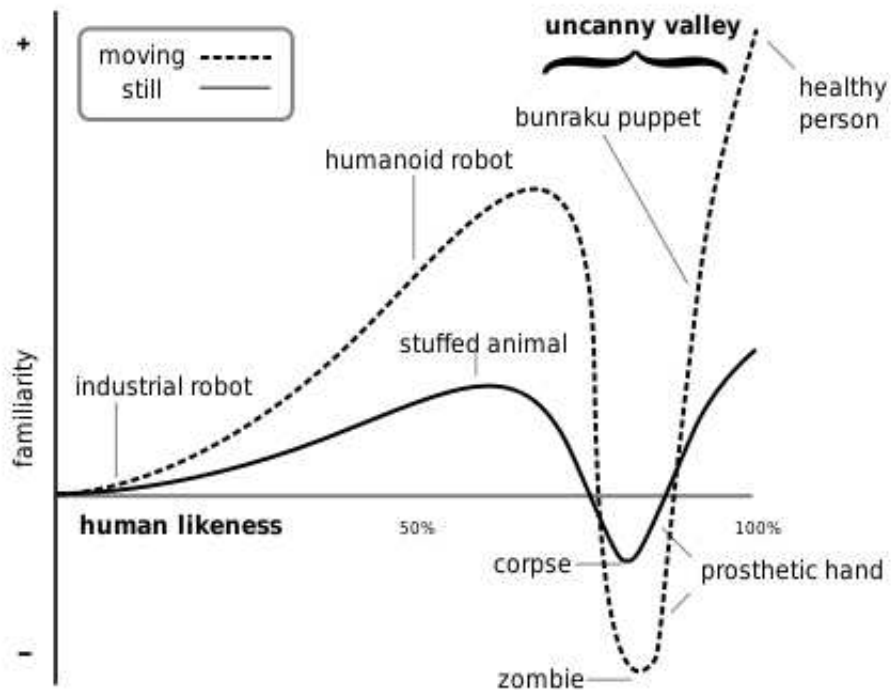


Figura 2.8: Valle inexplicable propuesto por Mori [1]

Existen trabajos como son los de MacDorman y Ishiguro [32] y David Hanson [33] que al tratar de demostrar la existencia del valle inexplicable dieron resultados opuestos. En [34] se intenta afirmar que la repulsión se produce por la alusión a la muerte o enfermedad, en las investigaciones se han tomado imágenes estáticas así como encuestas. Aunque lo analizado es interesante no es concluyente, es necesario identificar qué características sociales del robot permiten integrarse al entorno, sus aspectos dinámicos.

En [35] se analiza la dinámica del robot relacionada con su aspecto, sus resultados indican que es más difícil lograr un nivel de aceptación por las personas cuando el robot interactúa con ellas, que si estuviera inanimado o en una fotografía.

Es difícil definir cuando un robot pueda caer en el valle inexplicable, dependerá de muchos factores, sociales, culturales, psicológicos inclusive regionales, aunque es posible conseguir resultados positivos. Un caso en donde el androide ha pasado los efectos del valle inexplicable es el HRP4C.

2.4. Del rostro humano a un modelo mecánico.

Actualmente en todas las propuestas robóticas se realizan simplificaciones que permitan reducir tanto el número de componentes faciales emulados como su comportamiento. El FACS describe más de 60 AUs diferentes que puede realizar el rostro, e implementar todas y cada una de ellas en un modelo mecánico supondría construir un sistema extremadamente complejo, difícil de parametrizar, y poco práctico. Es necesario, por tanto, realizar una simplificación, seleccionando aquellos músculos y AUs más significativos en la expresión humana para ser implementados en el robot.

Siguiendo los trabajos de [36] y trabajos previos realizados en CARTIF con avatares animados realistas [4] se ha presentado una reducción del número de AUs faciales que se deben emular para conseguir mostrar de forma convincente las seis expresiones emocionales universales descritas en el FACS: disgusto, tristeza, enfado, alegría, miedo y sorpresa. Teniendo únicamente en cuenta dichas unidades de acción, el total de AUs significativas queda reducido a 17 (Ver tabla 2.4). De esta forma se pretende emular la acción muscular facial tanto desde el punto de vista del resultado visual externo, como desde el punto de vista anatómico y fisonómico. Como se verá en capítulo 3, cada AU ha sido emulada por uno o varios servos colocados en función de la posición anatómica del músculo o grupos de músculos que emulan.

Se ha desarrollado al robot con un sistema de visión que permite el seguimiento del usuario (capítulo 4). El sistema de seguimiento hace que sea necesario incluir ocho unidades de acción adicionales a las consideradas para la generación de expresiones, relacionadas con el movimiento de los ojos y del cuello. Tal y como demuestran diferentes estudios [37], [38], [39], resultan determinantes para una correcta comunicación y percepción de las distintas expresiones emocionales (por ejemplo, una expresión de enfado es mucho más reconocible si los movimientos de la superficie del rostro se complementan con un giro del cuello que baje la cabeza y se levanta la mirada). Por último, se ha incluido la unidad de acción relacionada con el parpadeo, ya que es un movimiento que los humanos realizan de forma continua e inconsciente, y otorga al robot una sensación de naturalidad aun cuando se encuentra en posición de reposo. Por tanto, al conjunto de unidades consideradas en la tabla 2.4 como esenciales, se han añadido las 9 unidades de acción de la tabla 2.5.

Tabla 2.4: Músculos a representar con sus correspondientes expresiones emocionales.

AU	Acción	Músculos involucrados	Expresión emocional
1	Levantar el interior de las cejas	Frontal (par lateral)	Tristeza, Sorpresa, Miedo
2	Levantar el exterior de las cejas	Frontal (par medial)	Sorpresa, Miedo
4	Bajar las cejas	Corrugador superciliar, depresor superciliar, prócero	Tristeza, Enfado, Miedo, Disgusto
5	Levantar el párpado superior	Elevador del párpado superior	Enfado, Sorpresa, Miedo, Disgusto
6	Levantar las mejillas	Orbicular del ojo (par orbital)	Alegría
7	Estrechar los párpados	Orbicular del ojo (par palpebral)	Enfado, Disgusto
9	Arrugar la nariz	Elevador del labio superior	Disgusto
10	Elevar el labio superior	Elevador del labio superior y alar de la nariz	Miedo, Disgusto
11	Elevar el labio superior	Elevador del labio superior	Tristeza
12	Separar las comisuras de los labios	Cigomático mayor	Alegría
15	Bajar las comisuras de los labios	Triangular	Enfado
17	Levantar la barbilla	Borla	Enfado, Disgusto
25	Separar los labios	Orbicular de la boca, borla	Alegría, Tristeza, Miedo
26	Bajar la mandíbula	Masetero, temporal, Pterigoideo interno	Alegría, Tristeza, Sorpresa, Miedo
29	Apretar los labios	Masetero, temporal	Enfado

Tabla 2.5: Unidades de acción agregadas.

AU	Acción	Músculos involucrados
51	Girar el cuello a la izquierda	Conjunto de músculos del cuello
52	Girar el cuello a la derecha	Conjunto de músculos del cuello
53	Girar el cuello hacia arriba	Conjunto de músculos del cuello
54	Girar el cuello hacia abajo	Conjunto de músculos del cuello
61	Mirar hacia la izquierda	Músculos extraoculares
62	Mirar hacia la derecha	Músculos extraoculares
63	Mirar hacia arriba	Músculos extraoculares
64	Mirar hacia abajo	Músculos extraoculares
65	Parpadeo	Orbicular del ojo (par palpebral)

2.5. Conclusiones

En este capítulo se ha concluido que:

- El grado de antropomorfismo necesario en los robots está en debate, hay diferentes opiniones acerca del grado requerido, algunos robots han superado exitosamente el valle inexplicable como son Kokoro y HRP4C.
- En el diseño se ha tenido en cuenta las nociones fisiológicas de una cabeza humana real, sus huesos y músculos principales, a partir de lo cual se desarrolló un modelo mecánico.
- Empleando el Sistema de Codificación de Acciones Faciales (FACS) y basados en trabajos anteriores sea reducido el número de unidades de acción necesarias. El total de unidades de acción implementadas en el robot es de 32.

Capítulo 3

Diseño mecánico y control de expresiones

“El diseño debe seducir, educar y, quizás lo más importante, provocar una respuesta emocional.”

– *April Greiman*

La integración de detalles y características anatómicas del rostro humano a un modelo mecánico no es una tarea trivial. A medida, que el modelo aumenta su complejidad, más difícil resulta la integración y ajuste de cada uno de los elementos introducidos para conseguir un adecuado realismo.

En este capítulo se describe las aproximaciones mecánicas hechas de una cabeza real. Se explican las herramientas de software utilizadas. Así como, se detallan las diferentes unidades de software añadidas para generar expresiones con el robot, por último se presenta el software desarrollado para implementarlas.

3.1. Diseño mecánico de la cabeza mecatrónica siguiendo el FACS

El diseño de un sistema mecánico que emule un rostro humano supone una selección de los músculos a implementar, así como de los posibles movimientos que dichos músculos serán capaces de realizar. Esta sección explica el diseño mecánico del robot. Se especifican los músculos asociados a cada servomotor así como su selección y sus respectivas características, se detallan algunos mecanismos de los ojos y el cuello.

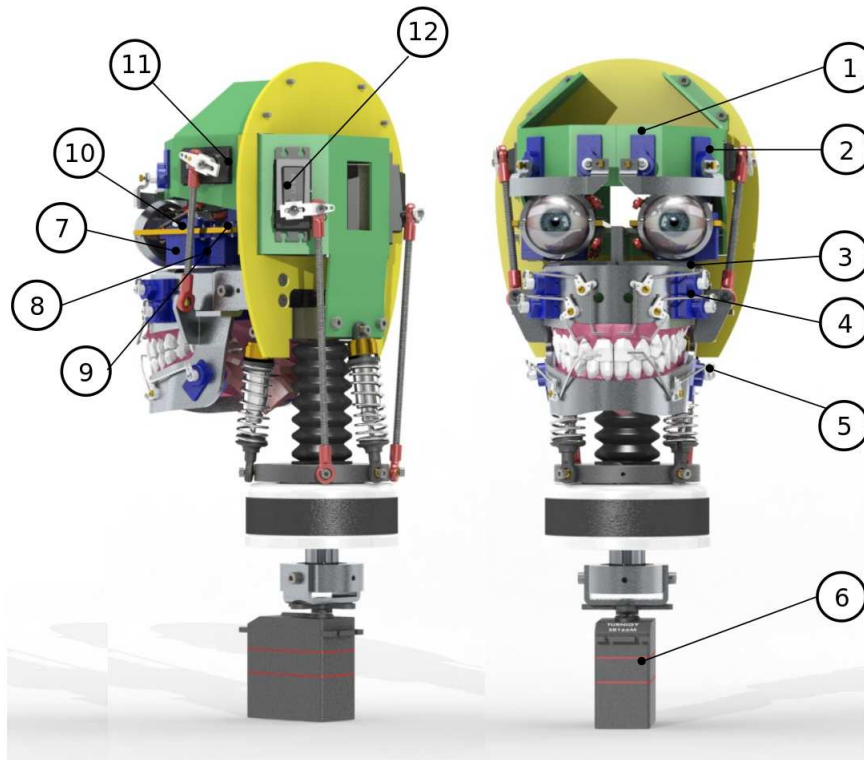


Figura 3.1: Modelo CAD de la cabeza mecatrónica

3.1.1. Ubicación de servomotores.

Los servomotores se han ubicado para representar uno o varios músculos del rostro humano. Diferentes unidades de acción fueron asociadas a determinados músculos. En la figura 3.1 se puede observar la posición de los diferentes servomotores.

La tabla 3.1 identifica el número de servomotor de acuerdo a la figura 3.1, asocia al músculo más significativo y especifica el área en donde se ve su efecto. Con esta distribución de servomotores se han obtenido 23 grados de libertad.

Las características principales de los servomotores se listan en la tabla 3.2. Para controlar la posición de cada servomotor se utiliza la técnica de modulación por anchura de pulso (PWM), la técnica se detalla en el apéndice 1.

El control del robot es en bucle abierto. El controlador emite una señal y supone que el servo alcanza la posición deseada en un lapso de tiempo. No se recibe ninguna notificación llegado a la posición deseada, aunque interna-

Tabla 3.1: Músculos asociados a los diferentes servomotores.

N.	Músculos más significativos	Área
1	Frontal (Par medial)	Frente y Cejas
2	Frontal (Par lateral)	Frente y Cejas
3	Levantar el labio superior	Boca
4	Zigomático mayor	Boca
5	Triangular, borla	Boca
6,12	Músculos del cuello	Cuello
7	Levanta parpados	Ojos
8	Recto superior e inferior	Ojos
9	Medial y lateral	Ojos
10	Par orbital and palpebral	Ojos
11	Masetero, temporal, pterigoideo	Quijada

Tabla 3.2: Características de los servomotores.

N	Tipo de servo	Características
1	Turnigy TG9e	Dimensiones: 23 x 12.2 x 29mm Peso: 9gr Torque: 1.5kg.cm (4.8V) Velocidad de Operación: 0.10seg/60 Voltaje de Operación: 4.8V Rango de Temperatura: 0 a 55 C
2	Turnigy MG958	Dimensiones: 40.2 x 20.1 x 36.8mm Peso: 65gr Torque: 12.0kg.cm (4.8V) Velocidad de operación: 0.2seg/60 Voltaje de operación: 4.8V Rango de temperatura: -20 a 60 C
3	Turnigy S8166M	Dimensiones: 75 x 59 x 27mm Peso: 154gr Torque: 33kg-cm Velocidad de operación: 0.21 deg/sec Voltaje de operación: 4.8V Rango de temperatura: -20 a 60 C

mente se tiene un potenciómetro que cierra el circuito.

3.1.2. Modelo Mecánico

En la figura 3.2 se muestra los tres elementos principales del modelo mecánico: frontal, mandíbula superior y mandíbula inferior.

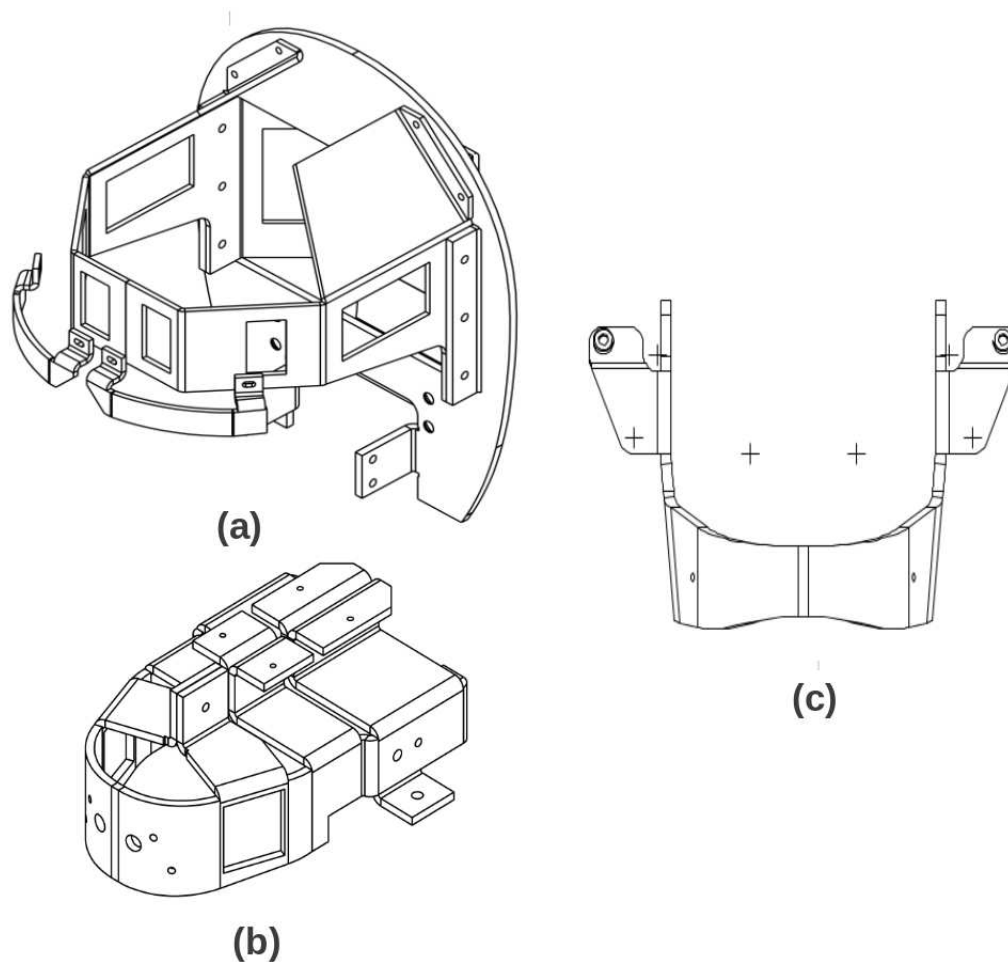


Figura 3.2: Modelo Mecánico: (a). Frontal (b). Mandíbula superior (c). Mandíbula inferior

La parte frontal tendrán los servomotores que ejecutaran el movimiento de las cejas, cuello y mandíbula. La mandíbula superior contendrá los servomotores del mecanismo de los ojos. Por último la mandíbula inferior contendrá 2 servomotores que representara el labio inferior.

3.1.3. Mecanismos Especiales

Esta sección describe algunos de los mecanismos especiales que posee el robot, como son el cuello, ojos, boca, etc. Se especifican los diferentes músculos involucrados y su equivalente mecánico.

Mecanismo de los ojos

El mecanismo del ojo ha sido diseñado con base en la anatomía humana (Fig. 3.3). El esquema de la Figura 3.3.a muestra los diferentes músculos responsables del movimiento del ojo: (a) recto superior, (b) oblicuo superior, (c) recto inferior, (d) recto lateral y (e) recto medial.

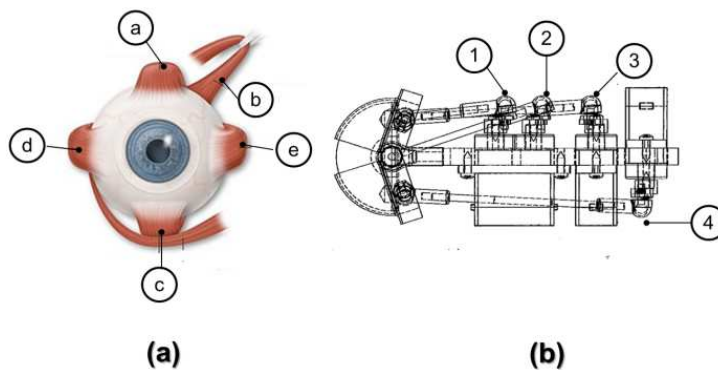


Figura 3.3: (a) Músculos principales (b) Mecanismo del ojo.

La aproximación mecánica desarrollada se observa en la figura 3.3.b. El servomotor 1 emula el músculo recto superior que permite abrir y cerrar el párpado superior. El músculo recto inferior está representado por el servomotor 4, que permite abrir y cerrar el párpado inferior. Los seis músculos que permiten el movimiento del ojo (recto superior, recto inferior, recto externo, recto interno, oblicuo superior del ojo y oblicuo inferior del ojo) se emulan con los servomotores 2 y 3.

El globo ocular tiene una dimensión de unos 25 mm de diámetro, cámaras de ese tamaño fueron integradas al diseño para posibilitar captar el entorno. En el capítulo 4 se detallara el proceso de visión del robot.

Tabla 3.3: Diferentes funciones del cuello y sus músculos.

Función	Músculo
Extensión	Esternocleidomastoideos
Flexión	Recto anterior, recto lateral, recto largo y recto anterior
Inclinación	Recto lateral y dorsal largo y los músculos que realizan la extensión y la flexión
Rotación	Músculos oblicuos externo e interno, fascículo oblicuo superior, esplenio y esternocleidomastoideo

Mecanismo del cuello

El mecanismo del cuello debe cumplir cuatro funciones importantes: extensión, flexión, inclinación y rotación. En la tabla 3.3 se indica la función y los músculos que la producen.

Las acciones de los músculos están representadas por tres servomotores, uno sirve para la rotación, mientras que la tensión, flexión y extensión es producida por dos servomotores con la unión de dos muelles y un freno de disco. La aproximación mecánica se puede ver en la figura 3.4

Otros Mecanismos

Una parte compleja de implementar se encuentra en la zona de la boca. No es posible simular la acción del músculo orbicular por un solo servo, ya que se trata de un músculo de tipo elipsoidal. Se ha realizado una aproximación por medio de 6 servomotores (4 en la parte superior y 2 en la parte inferior), que simulan las inserciones más significativas de otros músculos de expresión en el orbicular de la boca. Además se reforzó la forma elipsoidal del orbicular con una máscara de silicona. El resultado final de la situación de las diferentes inserciones emuladas se puede observar de forma esquemática en la figura 3.5.

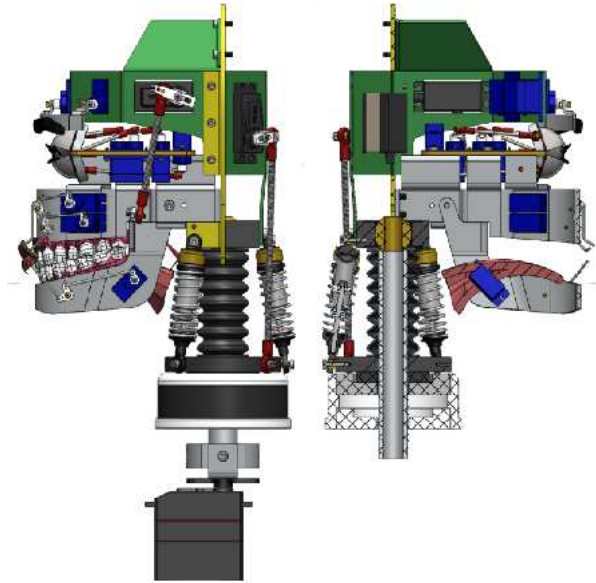


Figura 3.4: Mecanismo del cuello.

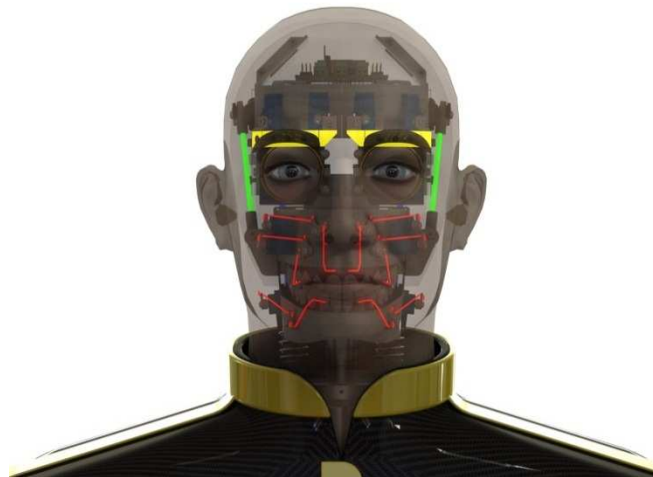


Figura 3.5: Estructura de la boca.

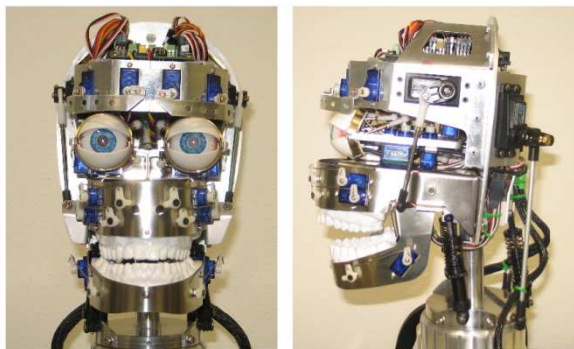


Figura 3.6: Vista Frontal y Lateral de la cabeza terminada.

3.1.4. Modelo Terminado

En la figura 3.6 se pueden apreciar las vistas lateral y frontal del modelo construido. Sobre ésta base se agregará una piel de silicona, que completará la sensación de realismo en la cabeza mecatrónica.

3.2. Programas utilizados en el proyecto

Antes de analizar el control de expresiones se hace una pequeña reseña del software empleado en la programación de la cabeza mecatrónica. Se revisan brevemente ROS, python y openCV que han sido las herramientas utilizadas.

3.2.1. ROS - Robotic operative system

De acuerdo a [40] ROS es un sistema operativo (bajo licencia BSD) que se plantea como un desarrollo a larga escala del proyecto STAIR de la Universidad de Standford y del Personal Robots Program de Willow Garage. Sus autores definen la filosofía de ROS en los siguientes puntos:

- Red punto a punto.
- Herramientas base.
- Sencillo.
- Multilenguaje.
- Libre y de código abierto.

La versión actual de ROS es la Groovy Galápagos (Nov 2012). ROS además de ser un sistema operativo, nos ofrece un conjunto de paquetes que tienen diferentes funcionalidades por ejemplo: drivers, simuladores, plugins, visualizadores, etc. También ofrece paquetes que sirven de conexiones con diferentes aplicaciones como son matlab, octave, arduino, opencv y varios.

ROS tiene soporte directo en Ubuntu Linux, en forma experimental en otros sistemas operativos como: Windows, Fedora, Raspbian, Debian, etc. Soporte total en dos lenguajes C++ y Python, aunque también se puede programar en otros lenguajes como Java.

Por último existen proyectos similares como: Player, YARP, Orocos, CAR-MEN, Orca, MOOS, y Microsoft Robotics Studio entre otros. Aunque al parecer ROS posee mayor apoyo que sus competidores en este sentido, de acuerdo a su página web (<http://www.ros.org>) existen 175 organizaciones o individuos que han publicado repositorios en ROS, más de un ciento de robots trabajan en ROS algunos de ellos tienen instrucciones detalladas en la página. Al momento existen 3699 (Nov 2012) paquetes disponibles en su página.

Se ha seleccionado ROS debido a que es un software libre. Además, en el proyecto aporta muchas ventajas como por ejemplo los drivers del controlador, drivers de diferentes cámaras, compatibilidad con la librería openCV, etc. También es compatible con la filosofía de varios robots desarrollados en CARTIF.

3.2.2. Python

Python es un lenguaje de programación escrito por Guido van Rossum, fue desarrollado en los años ochenta. La última versión es la 3.3.1 (en prueba), aunque actualmente se soportan dos oficialmente 2.7.3 y 3.3.0. Python es un lenguaje de alto nivel que hace énfasis en una clara sintaxis (no utiliza las llaves, en cambio, utiliza los espacios para delimitar subrutinas entre otros detalles como la eliminación de los paréntesis en algunos comandos).

Se trata de un lenguaje de programación multiparadigma, soporta la programación orientada a objetos (POO), programación imperativa y programación funcional. Además usa tipado dinámico y funciona en diferentes plataformas (Linux, Mac OS X y Windows). Es desarrollado bajo una licencia de código abierto a partir de la versión 2.1.1, que es compatible con la licencia pública general (GNU).

Se ha seleccionado Python para la programación del robot. Esto se debe a que se tiene experiencia previa en el lenguaje, es sencillo de implementar, es soportado por ROS y tiene una comunidad activa en la Web.

3.2.3. OpenCV

OpenCV (Open Source Computer Vision) es una librería de visión desarrollada bajo licencia BSD, soportada en un inicio por Intel y actualmente por Willow Garage e Itseez.

Su primera versión data de enero 1999, y tiene un sin fin de funciones (más de 500) como son: captura, análisis, manipulación de imágenes, video, etc. Tiene diferentes usos como son: detección de movimiento, control de procesos, reconocimiento de objetos, analizar formas y figuras, detectar rostros entre otros. OpenCV es multiplataforma funciona en Linux, Mac OS X y Windows.

OpenCV se ha utilizado en el proyecto para el análisis de las imágenes, seguimiento y la detección tanto de objetos como de rostros.

3.3. Control de expresiones

El control de expresiones tiene la función de integrar los servomotores, para ello dispone de tres módulos: el módulo controlador de servos, el módulo de unidades de acción y el módulo gestor de emociones. La figura 3.7 muestra el esquema empleado para el control de expresiones.

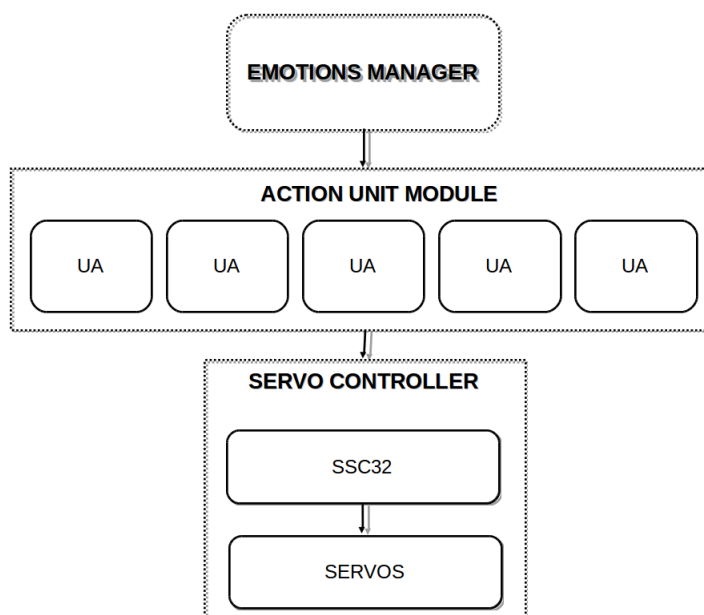


Figura 3.7: Sistema de control de expresiones.

3.3.1. Módulo controlador de servos

La unidad de control está formada por dos partes: hardware y software. La componente hardware es la tarjeta SSC32 que permite controlar hasta 32 servos y tiene una resolución de 1 μ S, con un rango de 0.50 mS a 2.50 mS. Además, la tarjeta permite la sincronización de varios servos al mismo tiempo y soporta 4 entradas digitales o analógicas para otros sensores. En la figura 3.8 se puede apreciar una foto del modelo de la tarjeta ssc32.



Figura 3.8: Tarjeta SSC32.

El segundo elemento es el software necesario para gestionar el movimiento de los servos en función de las órdenes recibidas por el nivel superior de Unidades de Acción. Además es el encargado de verificar que los servos se encuentren dentro de los límites preestablecidos de funcionamiento adecuado, teniendo en cuenta valores como las velocidades máximas de movimiento, límites máximos y mínimos de desplazamiento.

El controlador realiza la función de modular la anchura de pulso con los límites y parámetros que pueden ser modificados en un fichero configurable de extensión yaml. Esto permite que las modificaciones sobre las características del programa sean más fáciles de realizar y que se pueda reconfigurar sin alterar el software.

Listing 3.1: Archivo de configuración de servos

```
servo_0 :
  id: 0
  init_ang: -0.16 # 1400
  min_ang: -0.94 # 900
  max_ang: 0.63 # 1900
```

vel_max: 4.3

...

Como se muestra en la lista 3.1 la configuración es sencilla. Existen los siguientes elementos: una etiqueta de identificación (servo 0), el canal (id), los límites superiores e inferiores (minang, maxang) y la velocidad máxima (velmax).

3.3.2. Módulo de unidades de acción

Este módulo contiene todas las AUs codificadas como diferentes movimientos de los servos. Dicha codificación se basa en el análisis realizado en el capítulo 2. En la tabla 3.4 se listan las unidades de acción asociadas a cada uno de los servo motores.

Tabla 3.4: Unidades de Acción asociadas a cada servo

Servos	Unidades de Acción
1	AU1
2	AU2
1 - 2	AU4
3	AU11 - AU10
4	AU6
1 - 2 - 3	AU9
5	AU15 - AU17
6	AU51 - AU52
7	AU5 - AU61 - AU62 - AU63 - AU64
8	AU45
9	AU6
7 - 9	AU7
4	AU12
10	AU26
4 - 5	AU25
11 - 12	AU53 - AU54

Este módulo también permite seleccionar el nivel de intensidad de cada AU, lateralidad y tiempo de ejecución. En el capítulo 2 se mencionó la intensidad de acuerdo al FACS. Para modificar la posición del servomotor de

acuerdo a la intensidad elegida se multiplica por un factor el cual denominamos factor de intensidad K_i . En la ecuación 1 se establece la relación entre el K_i e i . Donde i es la intensidad codificada con números del 1 al 5, siendo 1 el nivel A y 5 el nivel E de intensidad, respectivamente.

$$K_i = 0,5[\tanh(i - 3) + 1] \quad (3.1)$$

La figura 3.9 muestra la curva que describe el conjunto de valores que puede adoptar el factor de intensidad K_i así como la codificación de intensidades empleada en el FACS.

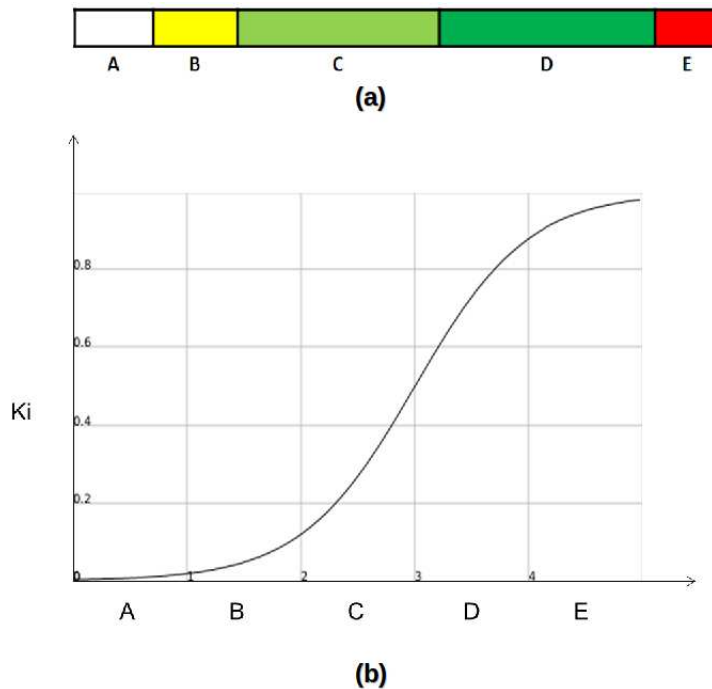


Figura 3.9: a) Factores de intensidad implementadas en el FACS (b) Factor de intensidad.

Para la lateralidad de expresiones se programaron unidades de acción independientes.

Con respecto al tiempo de ejecución la figura 3.10 muestra la posición del servomotor en función del tiempo. El tiempo de ejecución t_0 se ha establecido como el tiempo que tarda la unidad de acción en realizarse y volver a su posición inicial (parte de la expresión neutra). Este tiempo dependerá de la emoción que se esté ejecutando. Expresiones tales como ira y sorpresa tendrán por lo general un menor tiempo de activación t_1 (mayor velocidad)

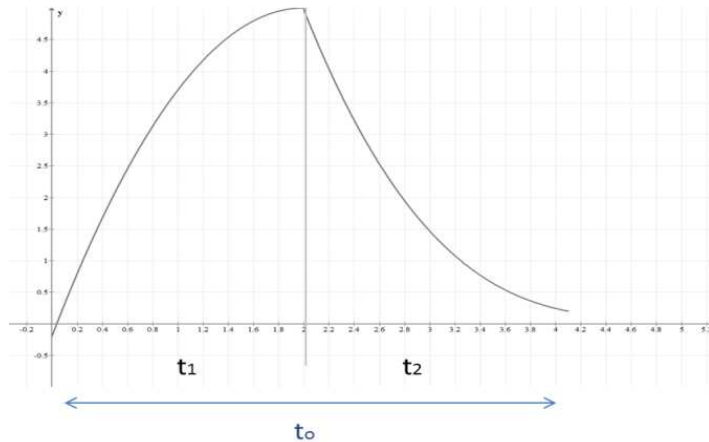


Figura 3.10: Ejemplo de ejecución durante el tiempo de un servo

que expresiones de felicidad y tristeza. La ecuación (3.2) expresa la función analítica representada en la figura 3.10.

$$f(t) = \begin{cases} -\frac{t}{t_1}(\frac{t}{t_1} + 2), & 0 < t < t_1 \\ \frac{1}{t_2^2}(t - t_0)^2, & t_1 \leq t \leq t_0 \end{cases} \quad (3.2)$$

El tiempo de activación t_1 será el tiempo que tarda la AU en alcanzar la máxima intensidad y el tiempo de desvanecimiento t_2 es el que tarda en volver a la posición de reposo. En la figura 3.10 se muestra un ejemplo del comportamiento de la actividad del servomotor en función del tiempo.

Los resultados de la ejecución de unidades de acción se presentan en el capítulo cinco.

3.3.3. Modulo gestor de emociones

El gestor de emociones es el encargado de gestionar la realización de expresiones complejas: disgusto, tristeza, enfado, alegría, miedo y sorpresa.

Para la realización de expresiones complejas, el gestor de emociones indica las AUs que deben activarse para cada expresión emocional. Cada expresión emocional viene definida por un vector de unidades de acción de dos componentes: la intensidad I_j de la emoción j y el tiempo de activación t_j . Matemáticamente se puede caracterizar por la ecuación 3.

$$[I_j, t_j] = \begin{bmatrix} i_{AU_1} & t_{AU_1} \\ \dots & \dots \\ i_{AU_N} & t_{AU_N} \end{bmatrix} \begin{bmatrix} Ki_j \\ Kt_j \end{bmatrix} \quad (3.3)$$

Donde i_{AU_i} y t_{AU_i} son respectivamente la intensidad y el tiempo de activación de cada unidad de acción i , y Ki_j y Kt_j son dos constantes que permiten regular la intensidad y duración global de la expresión emocional.

Listing 3.2: Archivo de configuración de gestos

```
sorpresa_1 :
  aus: [1, 5, 2, 2, 5, 5, 9, 5, 26, 3]

sorpresa_2 :
  aus: [1, 5, 2, 2, 5, 5, 9, 5]

...
```

El módulo gestor de emociones, dispone de una base de expresiones, configuradas en un archivo yaml. Un ejemplo de tal archivo se puede ver en la lista 3.2. El archivo yaml guarda en un vector cada una de las unidades de acción con su respectiva intensidad. Esto permite configurar y guardar diferentes expresiones.

El resultado final de las expresiones básicas se lo puede ver en el capítulo 5.

3.4. Operación remota

Para la operación remota se ha utilizado un joystick modelo Play Station 3 conectado vía USB. A través del joystick se pueden controlar tanto unidades de acción como expresiones complejas. Los mandos analógicos controlan el cuello y los ojos, en los botones están incorporadas unidades de acción diversas y con el mando digital se operan las expresiones básicas.

El control del joystick se lo realiza por medio de un nodo ROS. El nodo imprime un mensaje, el cual es un arreglo de valores tanto de botones como ojos, el mensaje se actualiza cada vez que el joystick es presionado.

3.5. Animación de expresiones

El control de expresiones se realiza a través de una estructura modular, lo que integra facilidad al desarrollo de expresiones. Sin embargo, la tarea de animar expresiones no es menos ardua. Por lo cual, es necesario implementar herramientas que permitan un mejor trabajo con las secuencias.

3.5.1. Estructura de fotogramas

El desarrollo de secuencias está estructurado en fotogramas claves (*key frames*), un fotograma clave es un punto dentro de la animación en donde nos interesa que el robot tome una posición determinada. En la figura 3.11 se aprecia la estructura del desarrollo de secuencias implementada.

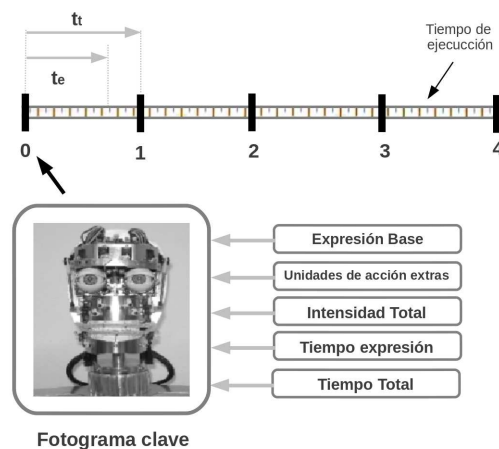


Figura 3.11: Estructura de fotogramas

Cada fotograma clave contiene: la expresión principal o base, la cual se obtiene de un grupo de expresiones preconfiguradas (módulo gestor de emociones). Las unidades de acción extras son las AUs agregadas o modificadas a la expresión base. La intensidad total de la expresión (0-1), es un porcentaje que multiplica a todas las intensidades de las unidades de acción del fotograma. El tiempo de expresión (t_e), es el tiempo que tarda la expresión en realizarse. El tiempo total (t_t), es el tiempo que existen entre fotogramas, como mínimo este tiempo tiene que ser igual al tiempo de expresión.

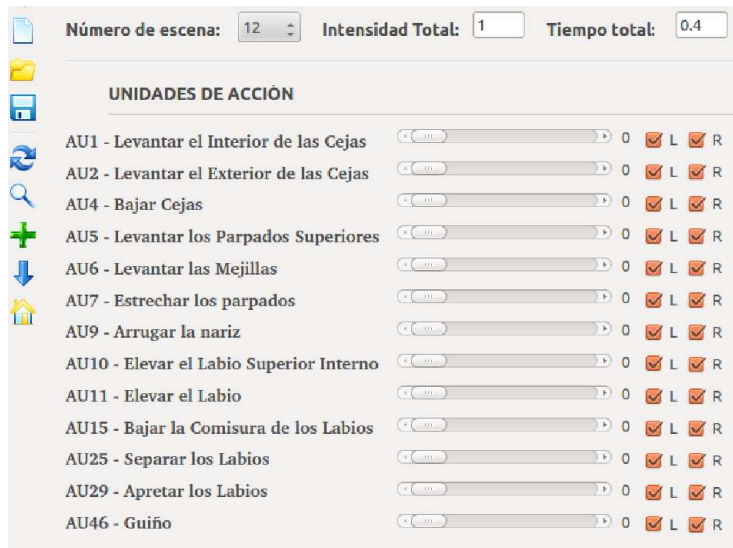


Figura 3.12: Programa para generar expresiones - Unidades de acción.

3.5.2. Integración de unidades de acción y expresiones básicas

Basados en trabajos similares [31] [41], se ha desarrollado un software que permite generar secuencias en tiempo real.

El software tiene tres partes fundamentales, ejecución de unidades de acción, ejecución de expresiones y por último ejecución de visemas (que se analizará en el capítulo 4).

La ejecución de unidades de acción se muestra en la figura 3.12. Permite el desarrollo de expresiones en forma detallada basándose en unidades de acción. El software permite modificar la intensidad y lateralidad de cada una de las unidades de acción.

La segunda parte del software analiza las expresiones básicas y complejas. Se pueden modificar las seis expresiones básicas: miedo, tristeza, sorpresa, disgusto, alegría y enfado en un porcentaje, este porcentaje afectará la intensidad de cada unidad de acción. Se dispone de la posibilidad de agregar expresiones predefinidas de una base de gestos preconfigurados. Por ejemplo, gestos de estar dormido, guiños o formas particulares de las expresiones básicas se pueden encontrar en la base. En la figura 3.13 se muestran algunas de estas expresiones.

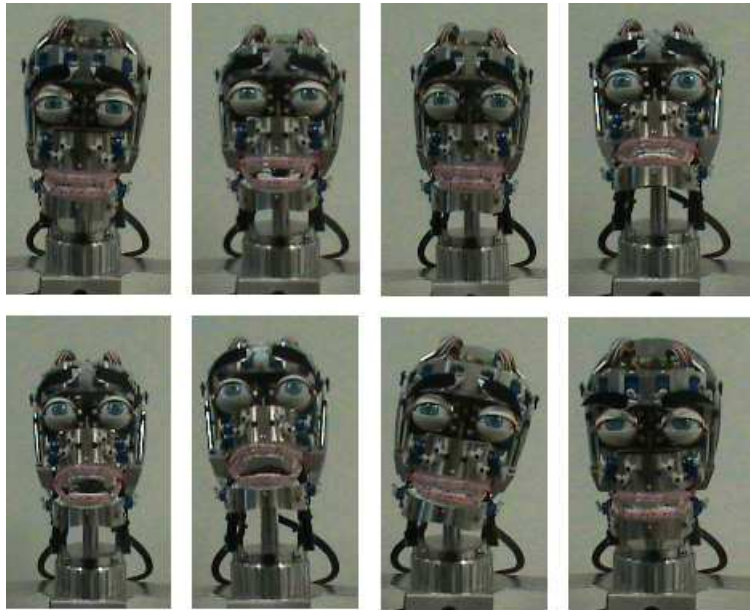


Figura 3.13: Ejemplos de expresiones de la base.

El programa contiene la posibilidad de editar directamente el archivo generado, esto permite un control pormenorizado de los diferentes fotogramas así como los cambios realizados en tiempo real. En la figura 3.14 se muestra la parte del software dedicada a las expresiones básicas y complejas.



Figura 3.14: Programa para generar expresiones - Expresiones básicas.

El software dispone de herramientas de edición, visualización, retorno

a la posición neutra, agregar fotograma, quitar fotogramas, etc. que facilitan la edición de los movimientos. También dispone de diferentes modos de ejecución, puede reproducir la secuencia completa, expresiones individuales, detener su reproducción, etc.

Al igual que los anteriores elementos este software genera un archivo yaml, que se puede ver en la lista 3.3.

Listing 3.3: Archivo de configuración obtenido con el programa de gestos

```
# Archivo de animacion

esc_0:
  expresion:          ninguna_1
  ua_extras:          []
  intensidad_total:  1
  tiempo_total:      0.2
  tiempo_espera:     0.2

esc_1:
  expresion:          guinio_boca
  ua_extras:          []
  intensidad_total:  1
  tiempo_total:      1
  tiempo_espera:     1.5

esc_2:
  expresion:          guinio_boca
  ua_extras:          []
  intensidad_total:  0
  tiempo_total:      1
  tiempo_espera:     1.5

...
```

Las unidades de acción obtenidas por los tres elementos del software se pueden traslapar entre sí, sin causar problemas, se tiene un orden de prioridad el cual no permite que haya ningún tipo de problema al momento de ejecutar la secuencia.

La interacción directa entre el software y el robot permite el obtener resultados más realistas además el proceso es bastante rápido. Las secuencias

se guardan en un archivo lo cual puede permitir que en cualquier momento se pueda editar y volver a utilizar. En el capítulo cinco se analizan los resultados obtenidos.

3.6. Conclusiones

En este capítulo se ha concluido que:

- El diseño mecánico tiene tres partes principales: frontal, mandíbula superior y mandíbula inferior. Se detallaron las diferentes aproximaciones realizadas en ciertos mecanismos importantes (ojos, cuello y boca).
- El control de expresiones posee tres módulos: el controlador de servos, de unidades de acción y gestor de emociones, se han implementado una arquitectura flexible que permite configurar fácilmente tanto los servos como las expresiones.
- Se ha desarrollado un software generador de expresiones, el cual facilita la programación de gestos. El programa permite trabajar en tiempo real las expresiones, además de poseer diferentes herramientas que ayudan a generar fácilmente un amplio rango de expresiones y comportamientos.

Capítulo 4

Módulos agregados

“Todo lo complejo puede dividirse en partes simples.”
– *René Descartes*

El capítulo trata acerca de los módulos desarrollados para la interacción social tales como visión, conocimiento, habla, etc. Se detalla la estructura de los mismos, sus características principales, el modo de funcionamiento. Por último se habla acerca de la estructura general del robot desarrollada en ROS.

4.1. Visión

El sistema de visión consiste en distintos módulos que tienen diferentes propósitos. Los módulos desarrollados son:

- Detección de movimiento.
- Detección de rostros.
- Detección de objetos.
- Tracking de rostros y objetos.

El estudio que concierne a la visión se hace con una sola cámara. El equipo se ha diseñado con dos cámaras, aunque para el presente proyecto no se ha planteado como objetivo el estudio de la visión estéreo.

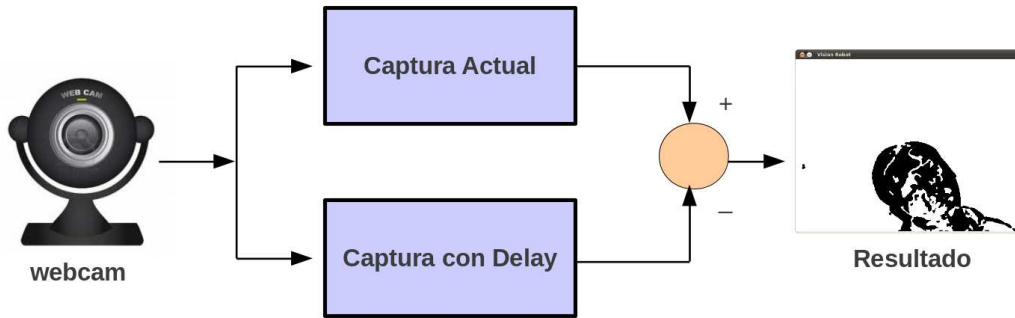


Figura 4.1: Esquema de la detección de movimiento

4.1.1. Detección de movimiento

La detección de movimiento se ha realizado empleando la técnica de captura óptica de movimiento por sustracción de vídeo en fotogramas anteriores. Se puede definir en los siguientes pasos:

1. Se realiza un fotograma inicial a un t_0 .
2. En t_1 se captura un nuevo fotograma.
3. Se convierten los fotogramas a blanco y negro.
4. Se ecualizan los fotogramas (para eliminar los efectos de las zonas más luminosas).
5. Se restan los fotogramas.
6. Se calcula el porcentaje de píxeles oscuros que se tienen en la imagen restada.
7. Se actualiza el fotograma en t_0 con el fotograma en t_1 .
8. Se repite el proceso a partir de 2.

En la figura 4.1 se puede ver un esquema del proceso. La técnica es sencilla de implementar y tiene un buen grado de robustez a las perturbaciones. Presenta ciertos inconvenientes, es susceptible a la iluminación del lugar, el robot cambia constantemente de posición, donde se afectan la sustracción

de fotogramas. Corresponde a los niveles superiores validar el porcentaje de movimiento y el grado de confiabilidad de lo recibido. En la figura 4.2 se observa el resultado de la técnica.



Figura 4.2: Detección de movimiento

4.1.2. Detección de rostros

Para la detección de rostros se ha implementado el algoritmo planteado en [42] propuesto por Viola y Jones (la explicación detallada de este método en castellano se encuentra en [43]), la figura 4.3 representa el algoritmo del método. Los diferentes pasos del algoritmo son:

Imagen Integral

Una imagen integral permite calcular la suma de las subregiones de la imagen rápidamente. La imagen integral es obtenida al agregar una fila y una columna $(w+1, h+1)$ estos nuevos elementos reciben el valor de 0, cada pixel en la imagen integral estará definido como la suma de los pixeles anteriores. La imagen integral tiene diferentes aplicaciones blurs, gradientes, sumatorias, promedios, derivaciones.

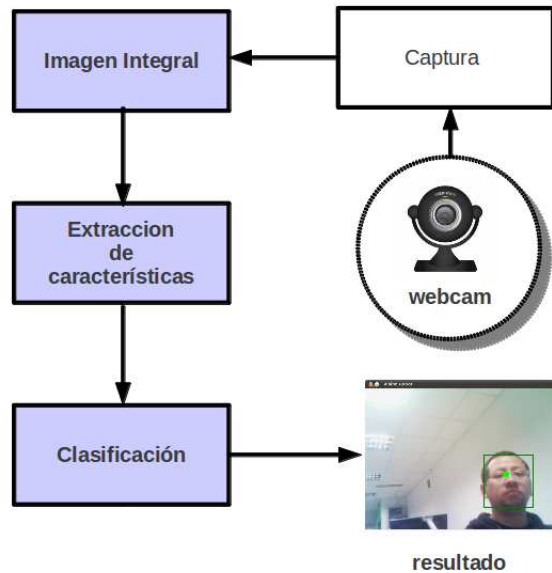


Figura 4.3: Flujograma de la detección de rostros

Extracción de características

Para obtener las características se utilizan filtros con base Haar (estos filtros tienen una forma rectangular simple), estos filtros son calculados sobre la imagen integral. Los filtros generan características de puntos y líneas en la imagen por el contraste que existen entre las regiones del fotograma. Se pueden cambiar diferentes características dentro de los filtros: escala, orientación, tamaño y frecuencia.

Clasificación

En este punto se procederá a la detección en base a un modelo inducido durante el entrenamiento, el entrenamiento es en base de ejemplos positivos y negativos. Para clasificar se utiliza un método denominado *boosting* que utiliza varios clasificadores sencillos que reunidos forma uno más robusto, de ahí el nombre de clasificador en cascada. En el método propuesto se utiliza el clasificador AdaBoost [44].

Las ventajas de este método son su rapidez y gran eficiencia. Por otra parte no reconoce color ni se puede detectar otro modelo que no se haya inducido en el entrenamiento, además de no detectar cambios o rotaciones en el rostro. Se pensó en un principio el agregar varios clasificadores para

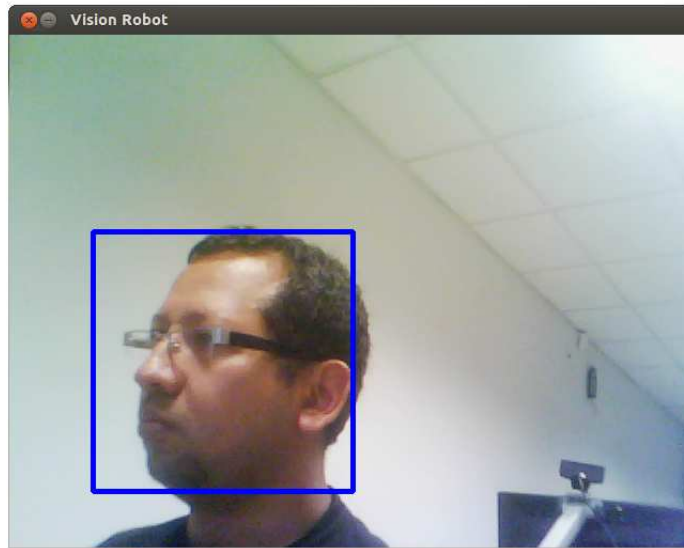


Figura 4.4: Detección de rostros

diferentes posiciones del rostro, pero en las pruebas se rechazó la idea debido a que la clasificación era muy lenta. La figura 4.4 presenta los resultados de la técnica propuesta.

4.1.3. Detección de objetos

Camshift viene del inglés “*Continuously Adaptive Mean Shift*”, combina el algoritmo básico *Meanshift* [45] con una región de tamaño adaptable que la modifica hasta converger. En la figura 4.5 se puede ver un esquema de la técnica empleada.

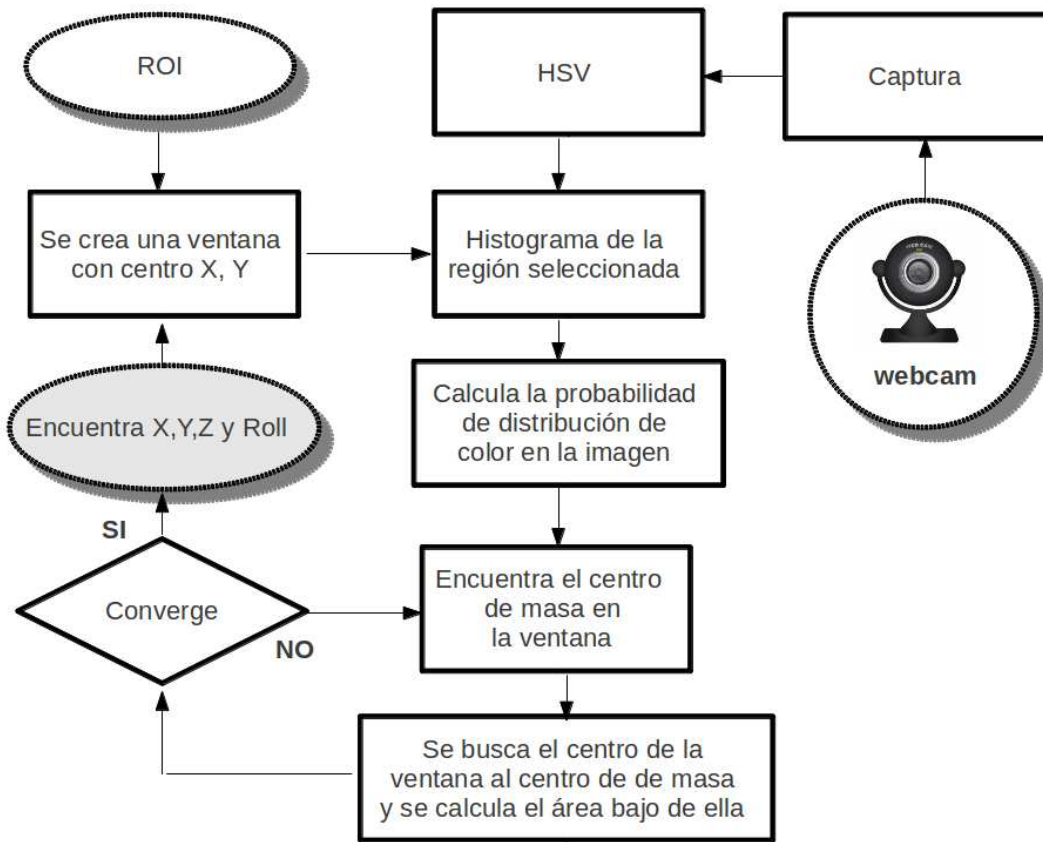


Figura 4.5: Flujograma de la detección de objetos

Se empieza estableciendo una ventana de búsqueda, luego se define un histograma para cada nueva imagen. Se calcula la retroproyección del histograma (back projection [46]). La retroproyección consiste en generar una imagen en escala de grises donde cada uno de los pixeles tendrá como intensidad el valor del histograma. Lo que se consigue será la probabilidad de que el pixel pertenezca al objeto.

Se calcula el centroide de la retroproyección mediante el Meanshift. También, se calcula el centroide en la ventana de la búsqueda actual. El siguiente paso establece el centroide obtenido como el nuevo centro de la ventana de búsqueda, vuelve a encontrar el centro de masa de la ventana hasta que converja. Se trata de un algoritmo iterativo que converge a máximos de probabilidad

Las ventajas de este método son: rapidez, relativamente robusto ante el ruido y es exacto. Como desventajas tenemos que al trabajar con color puede reconocer zonas erróneas con colores parecidos como posibles candidatas con

lo cual se perdería el objeto o rostro. La figura 4.6 muestra el resultado de la técnica empleada.



Figura 4.6: Detección de objetos

4.1.4. Flujo óptico

El aplicar ciertas técnicas (detección de rostros y objetos, reconocimiento de objetos, aprendizaje, etc.) iterativamente consume muchos recursos del sistema. Además en algunos casos no tienen la suficiente robustez a ciertas variaciones, es por esto necesario recurrir a algoritmos de seguimiento o *tracking*.

El método Lucas-Kanade desarrollado en [47] fue implementado en el robot. El método asume que el flujo es constante en los píxeles cercanos, y resuelve el flujo óptico por el criterio de mínimos cuadrados. El método tiene algunas características importantes:

- El brillo será constante, el punto en movimiento no cambiara sus características.
- Entre dos fotogramas consecutivos el movimiento será pequeño.
- El conjunto de puntos se moverán congruentemente, una serie de puntos tendrán la misma dirección.

La librería openCV provee del método iterativo Lucas-Kanade con pirámides [48], que consiste de los siguientes pasos:

1. Realizar una imagen piramidal de dos fotogramas.
2. Se inicializa los vectores de flujo para cada uno de los pixeles deseados
3. Para cada nivel en la pirámide, se calcula la x y y derivadas del primer fotograma.
4. Igualmente en cada nivel en la pirámide, empezando desde el más pequeño se computa el flujo óptico para cada nivel de la pirámide.
5. Para cierto número de iteraciones, se resuelve el algoritmo Lucas-Kanade a través de mínimos cuadrados en la selección, y lo calculado asume como la solución. Se repite las iteraciones hasta que cumpla el número máximo de iteraciones hay sido alcanzado.
6. Propaga las velocidades desde los niveles más bajos hacia arriba, hasta que el nivel máximo de la pirámide es alcanzado.

Este método es relativamente robusto al ruido, además asume que el desplazamiento de un objeto entre dos imágenes consecutivas es constante y pequeño. El algoritmo necesariamente debe iniciarse en un punto determinado, este punto se suministra con algunas de las técnicas anteriormente descritas.

Se ha realizado diferentes pruebas y se concluyó que el mejor resultado se obtiene poniendo un punto en el centro del rostro o del objeto detectado y de ahí aplicar la técnica. La figura 4.7 muestra el resultado de la técnica empleada.



Figura 4.7: Flujo óptico

La técnica posee ciertos problemas como son el trabajar en cortas distancias entre fotogramas, en el caso actual no existe problemas debido a que el operador no necesita hacer movimientos bruscos. Otro problema es la calidad del tracking, esta se va degradando con el tiempo, para corregir este error se ha optado por garantizar un mínimo de calidad en el punto, en caso que esta calidad se pierda se volverá a recurrir a uno de los algoritmo anteriores para encontrar el objeto o rostro. Por último la técnica tiene problemas inherentes a la visión como son sombras, rotaciones, iluminación etc.

4.2. Seguimiento de rostros y objetos

Para el seguimiento de rostros y objetos se toman en consideración ocho unidades de acción. Cuatro que son del cuello (izquierda AU51, derecha AU52, arriba AU53, abajo AU54) y cuatro que son de los ojos (izquierda AU61, derecha AU62, arriba AU63, abajo AU64). En la figura 4.8 se puede apreciar el esquema del programa desarrollado para el seguimiento.

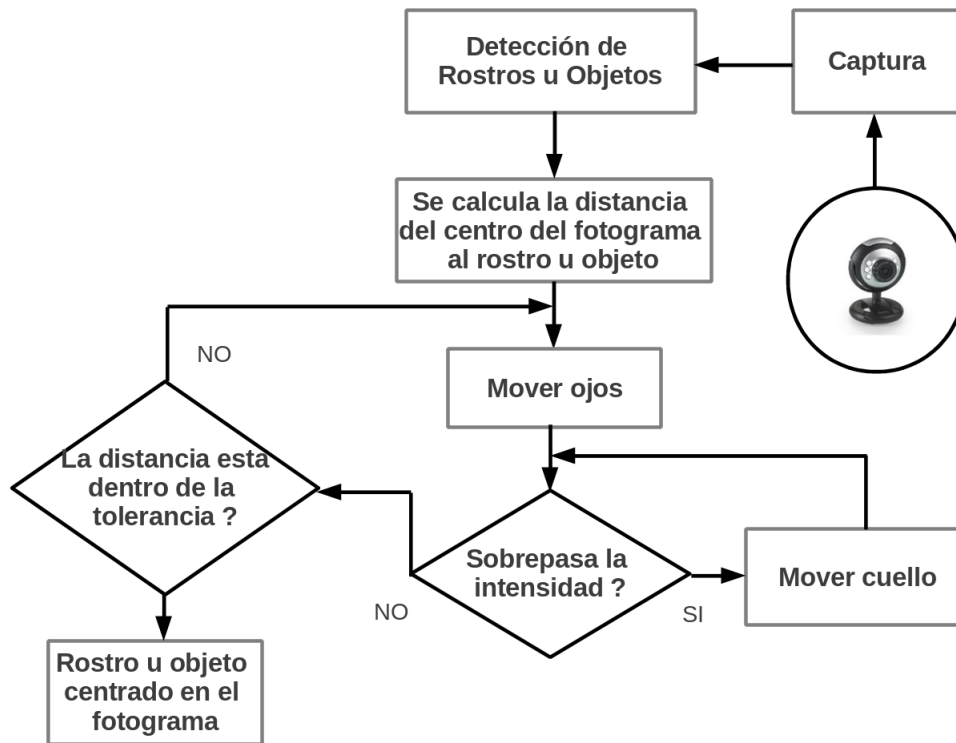


Figura 4.8: Flujograma del tracking

Para el seguimiento se ejecuta el siguiente algoritmo:

1. Se obtiene un fotograma a partir de la cámara web.
2. Aplicando una de las técnicas anteriormente descrita. Si se encuentra un rostro o un objeto conocido se lo identifica como una “región de interés” (ROI).
3. El programa calcula la distancia del centro del fotograma al centro de la ROI, si es este mayor a la tolerancia deseada activa las unidades de acción de los ojos.
4. Las unidades de acción de los ojos se desplazan a encontrar el objeto.
5. En caso de que la intensidad de las unidades de acción de los ojos excede un cierto valor, el robot activará las unidades de acción del cuello.
6. Al disminuir la intensidad de las unidades de acción de los ojos a un valor deseado, las unidades de acción del cuello dejan de ejecutarse.

7. Al centrarse el rostro u objeto en el fotograma con cierta tolerancia el algoritmo finaliza.

4.3. Habla

Es importante que el robot dé una correcta información visual al momento de comunicarse. El ser humano aprende desde pequeño a relacionar ciertas expresiones a posiciones de labios y boca, siendo así, es necesario que exista tanto coherencia temporal y espacial cuando el robot se exprese. Para el módulo del habla se ha desarrollado un sistema semiautomático de generación de visemas y una herramienta con el fin de facilitar la animación de secuencias con habla.

Los elementos importantes para el desarrollo de un sistema de habla en un robot son los visemas y fonemas. Un fonema es la unidad sonora básica de un idioma. En cambio, un visema es la representación gráfica de cada uno de los fonemas y está directamente relacionado con la posición de la boca y los labios. Existen un número reducido de visemas a implementar, tomando como referencia [49] el número de visemas a implementar en español es 10, en la figura 4.9 se puede ver los visemas para el español.

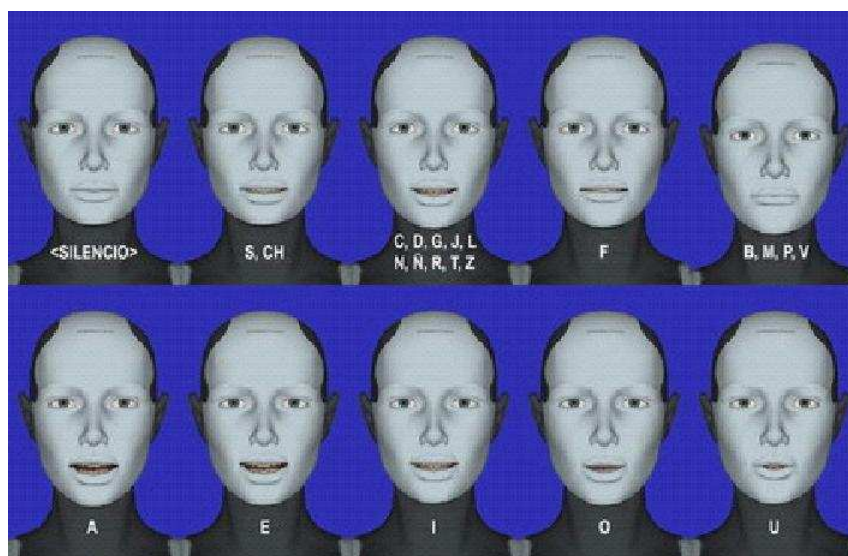


Figura 4.9: Visemas en español

Los visemas se configuraron como expresiones básicas. A pesar de esto, agregar los visemas en una secuencia de manera manual resulta difícil de

implementar, por lo tanto, es necesario generar automáticamente la secuencia de visemas.

4.3.1. Generación automática de visemas

Existen muchos modelos para definir la relación que existe entre la señal auditiva y la posición de la boca, el modelo implementado para el robot se lo puede ver en la figura 4.10.

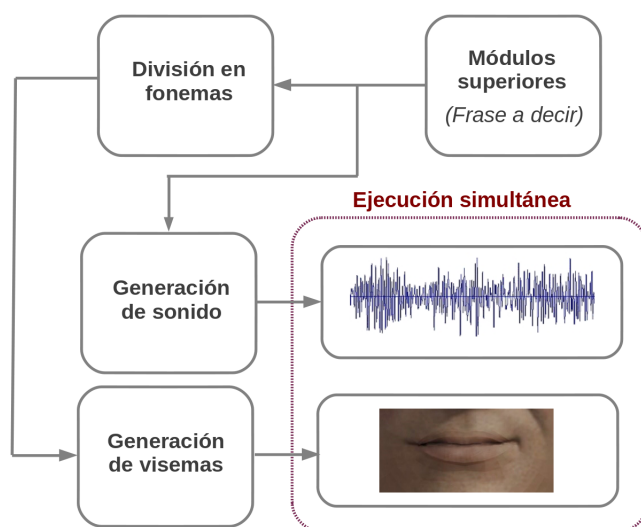


Figura 4.10: Esquema del proceso de habla en el robot

El proceso analiza la “*frase a decir*” a partir de la cual se obtiene la composición fonética. Para realizar el análisis fonético se ha utilizado el programa *espeak*, con él se genera un archivo de extensión *pho*, este archivo contiene los fonemas necesarios. En la lista 4.1 se puede ver la estructura del archivo.

Listing 4.1: Archivo *pho*

```

o      36      0 116 80 108 100 108
l      65
a      38      0 108 80 105 100 105
  
```

En el ejemplo se dice la palabra hola, la h no se la considera en el análisis fonético por lo tanto el número de fonemas queda reducido a tres, se especifican en la primera columna el fonema y la segunda columna el tiempo de duración, esta información resultara importante para generar el archivo de visemas. El resto es información prosódica necesaria para sintetizar la voz.

La creación de los visemas se realiza a partir del archivo pho antes generado. A partir de la información anterior se agregan los visemas necesarios de la base de expresiones, los cuales serán guardados como una expresión en un fotograma clave. El tiempo total de duración de la expresión será el mismo tiempo que tarda el fonema, cuando los tiempos de los fotogramas son muy pequeños se unifican los fotogramas en uno solo.

En cuanto a la generación de la señal audible, existen diferentes programas en Ubuntu para *Text-To-Speech*, programas como orca, espeak, ubuntu text reader, festival, etc. Para sintetizar la voz se utiliza el programa gratuito MBrola desarrollado por la *Faculté Polytechnique de Mons* (Bélgica). Tiene una licencia gratuita para aplicaciones no comerciales o militares y soporta muchos lenguajes, es un sintetizador de lenguaje basado en la concatenación de dífonos (un dífono abarca la mitad de la realización de un fonema hasta la mitad de la realización del fonema siguiente). El programa recibe el archivo de fonemas y genera el audio en bases a voces preconfiguradas del mismo programa.

Por último se ejecutan de forma paralela el archivo de animación de visemas como el archivo de sonido generado. El resultado es una correcta sincronización de los labios y la boca con la señal audible.

4.3.2. Integración de visemas

Se dispone de una interface gráfica que a partir de “la frase a decir”, realiza la división, generación automática de visemas y gráfica el sonido producido. La interface presenta los diferentes visemas separados en fotogramas claves. Un ejemplo de la interface se puede visualizar en la figura 4.11.

Con el software se pueden realizar diferentes acciones: ajustar los tiempos para obtener resultados perfectos, alterar tanto el tiempo de ejecución entre fonemas, como el fonema en sí. Se pueden agregar otras unidades de acción además de las generadas para los visemas e incrementar el grado de realismo, por ejemplo si se quiere resaltar la expresión con un guiño o movimiento de cuello con lo cual se lograría un mayor efecto de naturalidad. La figura 8 muestra la herramienta de sincronización de labios.

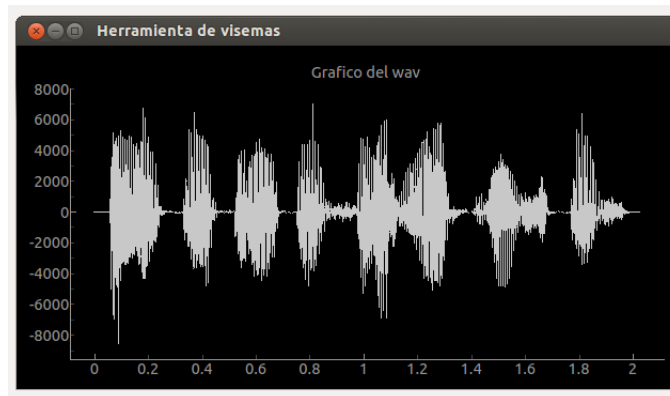


Figura 4.11: Programa para generar expresiones - Integración de visemas

4.4. Conocimiento

El nodo de conocimiento obtiene información acerca de diferentes temas que pueden interesar a las personas que interactúan con el robot. Actualmente el módulo de conocimiento puede obtener la siguiente información:

- Información de la Wikipedia.
- Contar chistes.
- Fechas importantes.
- Dar Fecha y Hora.

Para la información de la Wikipedia se accede a la URL: *es.wikipedia.org/*. En donde a través de una "palabra clave" se encuentra la información, si está disponible se despliega la primera frase de la información de la wiki, en caso de encontrar un término similar comenzara la búsqueda nuevamente. Por último, si no encuentra ninguna información desplegara un mensaje de que no ha encontrado información al respecto.

Para "contar chistes" se dispone de una base de datos con chistes cortos. Lo que se hace es generar un número randómico a partir de la cantidad de chistes que existen en la base y el seleccionado se lo envía al nodo del habla. La base de chistes puede ser modificable y se encuentra en un fichero txt.

La fecha y hora se obtiene de la información adquirida en el sistema con la cual puede hacerse el resto de cálculos necesarios. Las fechas importantes se basan en un fichero de formato txt que el robot dispone para recordar al usuario en caso de ser necesario información de fechas especiales, feriados u otros datos importantes que se hayan guardado anteriormente.

4.5. Arquitectura desarrollada

En general los sistemas robóticos son complejos. Una forma de trabajar con esta complejidad es proveer de modularidad al sistema. De esta manera, todo el sistema ha sido dividido en módulos que se han analizado anteriormente con un pequeño nivel de abstracción. La figura 4.12 muestra el esquema de la estructura implementada en el robot.

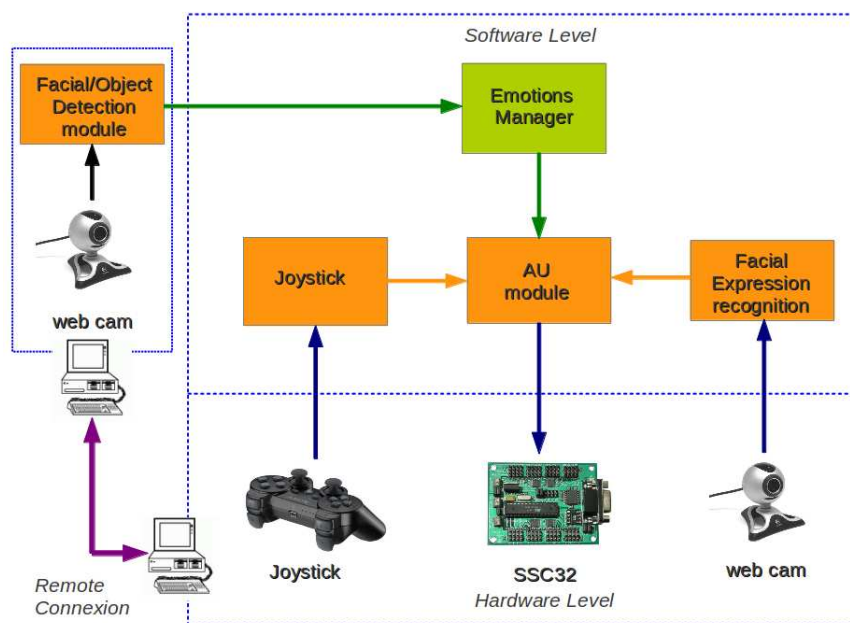


Figura 4.12: Arquitectura modular

Como podemos ver se puede distinguir dos niveles de software y uno de hardware, en donde el nivel de hardware incluye tanto sensores del robot (cámara web), controladores (tarjeta ssc32), y actuadores (joystick). Los niveles de software que ya sea analizado previamente, cabe mencionar que estos módulos pueden ser ejecutados in situ o desde una computadora externa.

Para el nivel de decisión se han implementado diferentes máquinas de estado. Se han programado tres máquinas de estado: acciones esporádicas, búsqueda e información. En la figura 4.13 se observa un esquema.

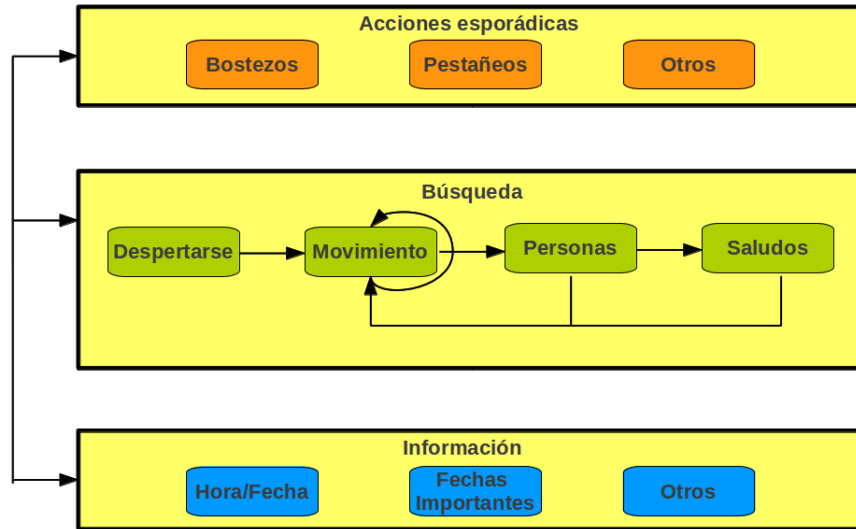


Figura 4.13: Máquina de estados

Las tres máquinas de estado funcionan de manera paralela. La primera para las acciones involuntarias, se encarga a un tiempo aleatorio de cerrar los ojos, bostezar, hacer pequeños movimientos con la cabeza y el cuello. La segunda máquina virtual incluye una pequeña secuencia que es: despertarse, busca algún movimiento si encuentra movimiento, busca un rostro por un tiempo caso contrario vuelve a buscar, si encuentra un rostro sonríe y dice “hola como estas”, después de lo cual continúa buscando. La tercera máquina virtual se encarga de enviar la información requerida por el usuario, lo que corresponde al módulo de conocimiento.

La herramienta utilizada para el desarrollo de la arquitectura es ROS. Por medio de ROS se ha podido integrar los diferentes módulos explicados en este trabajo y en un futuro se integraran otros módulos que nos permitirán mejorar el desempeño del robot. Para información de acuerdo a paquetes y módulos programados de la arquitectura revisar el apéndice B.

Esta flexibilidad da importantes ventajas por ejemplo, se puede remplazar fácilmente un componente por otro, de la misma manera se puede agregar nuevo dispositivos. Otra de las ventajas es poder operar remotamente el sistema, con este fin se ha desarrollado una interface de usuario que permiten acceder a los diferentes módulos programados. En la figura 4.14 se puede ver la interface de usuario con la cual se opera remotamente al robot.

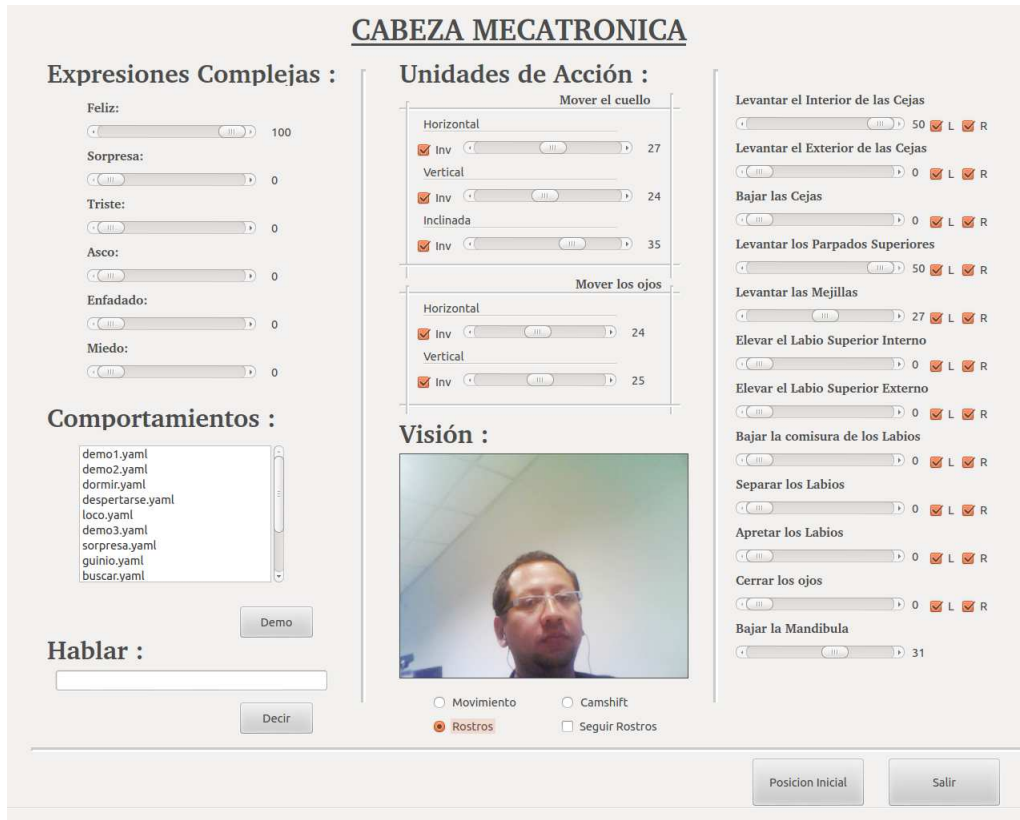


Figura 4.14: Interface de usuario para operación remota

La interface de usuario para operación remota puede accionar diferentes nodos en el robot. Por ejemplo puede modificar las diferentes unidades de acción en tiempo real, también ejecutar diferentes comportamientos preconfigurados. La interface permite recibir información del sistema a través de las cámaras. Por último, La interface permite hablar al robot diferentes frases con sincronización automática de labios en tiempo real.

4.6. Conclusiones

En el presente capítulo se ha concluido que:

- El sistema de visión está constituido por los siguientes nodos, un sistema de reconocimiento de movimiento, detección de rostros y colores así como un sistema de seguimiento. Los cuatro módulos conforman el sistema de visión que es procesado por una máquina virtual la cual ayuda al robot a tener una percepción de su entorno.

- El sistema de habla del robot fue descrito, en particular la obtención semiautomática de visemas, así como las herramientas que se han creado con el fin de facilitar el trabajo con los visemas.
- Las siguientes funciones básicas fueron implementadas al robot: visión, habla, conocimiento. Actualmente se trabaja en otros nodos que permitirán mejorar el nivel de interacción que tendrá el robot con la personas.
- Para trabajar con la complejidad inherente a los sistemas robóticos se ha implementado una arquitectura modular. El sistema modular está constituido de tres niveles: el nivel hardware, el nivel software y un tercer nivel de decisión, en cada nivel se ha descrito su función así como los diferentes elementos que lo conforman.

Capítulo 5

Resultados y evaluación del proyecto.

“Uno se alegra de resultar útil.”
– *El hombre bicentenario*

En este capítulo se expondrá los resultados obtenidos en diferentes partes del proyecto. También se explicara la evaluación realizada por medio de una encuesta. Se detallara la muestra, medios utilizados y por último se expondrá los resultados obtenidos de la evaluación.

5.1. Resultados de los diferentes elementos en el robot

A continuación se presenta los resultados obtenidos de los diferentes elementos desarrollados en el proyecto. Se presentan resultados de:

- Unidades de acción.
- Expresiones básicas.
- Animación de expresiones.

5.1.1. Unidades de acción

En la figura 5.1 se muestra los resultados obtenidos para algunas unidades de acción y la combinación de las mismas. En la figura 5.1a se observa la posición neutra del robot. El resultado de ejecutar la unidad de acción 1 se puede ver en la figura 5.1b y de ejecutar la unidad de acción 2 se puede ver

en la figura 5.1c. Por último la figura 5.1d muestra el resultado de unir las dos unidades de acción anteriores.

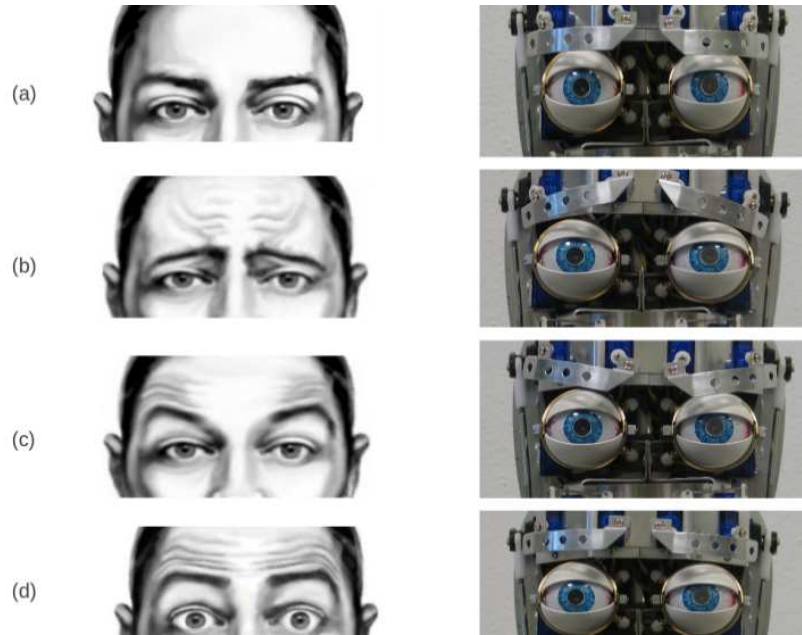


Figura 5.1: Diferentes unidades de acción

5.1.2. Expresiones básicas

En la figura 5.2 se muestra cómo la combinación de unidades de acción permite la realización de expresiones básicas. La figura 5.2a indica la posición neutra del robot. Por ejemplo la expresión de disgusto se genera mediante la combinación de las unidades de acción (AU4 + AU5 + AU10L) mientras que la expresión de enfado se ha obtenido con las unidades de acción (AU7 + AU10L + AU15L + AU62). La figura 5.2d muestra la expresión de felicidad que se ha obtenido con la suma de las siguientes unidades de acción (AU5 + AU12 + AU26 + AU6 + AU53).

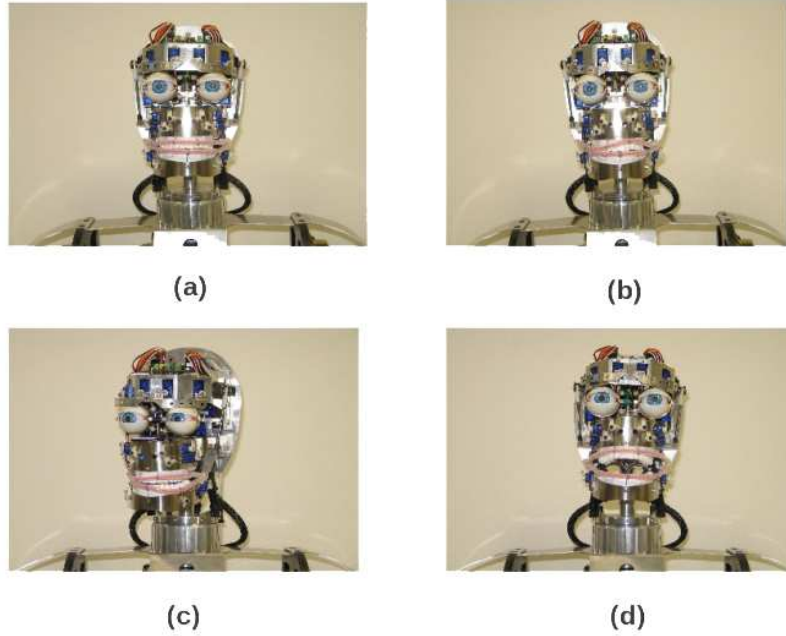


Figura 5.2: Diferentes expresiones obtenidas con la cabeza mecatrónica. (a) Neutral (b) Disgusto (c) Enfado (d) Felicidad

5.1.3. Animación de expresiones

En la figura 5.3 se puede ver fotogramas de una secuencia que se obtuvo en unos pocos minutos de edición. La secuencia completa abarca más de 30 fotogramas clave, es interesante decir que tal animación hubiera sido muy difícil de lograr sin la ayuda del software desarrollado.

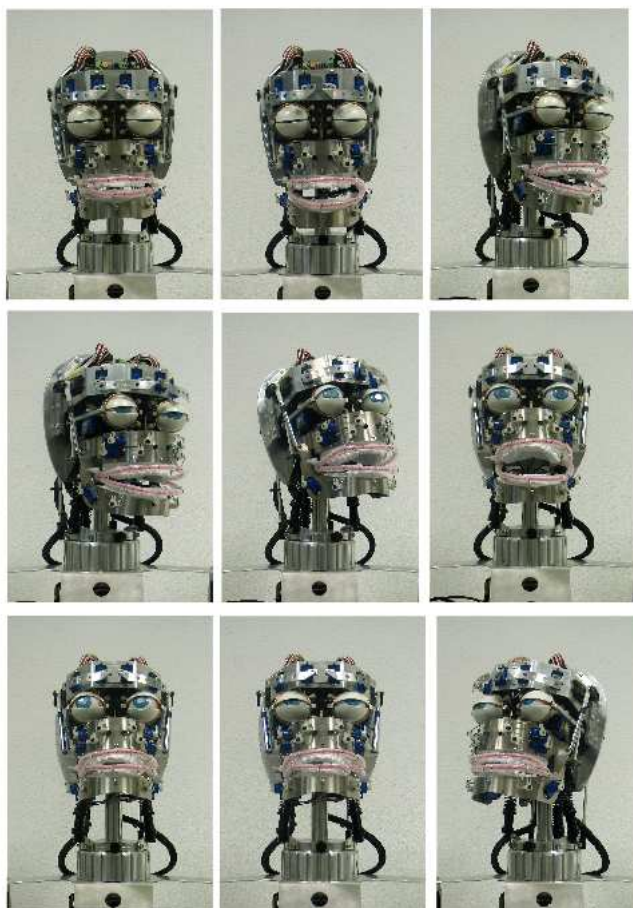


Figura 5.3: Secuencia animada

5.2. Evaluación de la cabeza mecatrónica

La evaluación de la cabeza mecatrónica se ha realizado a través de una encuesta a diferentes personas. Existen algunos criterios que se desean analizar al realizar esto:

- Reconocer las diferentes expresiones que realiza el robot.
- Analizar en qué manera la intensidad afecta el poder reconocer las diferentes expresiones.
- Evaluar si sus movimientos parecen naturales.

- Si el robot se expresa con claridad, si los visemas presentados dan la sensación de realismo.
- Si les ha producido agrado el interactuar con el robot.

5.2.1. Participantes

Las características de los participantes se muestran en la tabla 5.1.

Tabla 5.1: Análisis de la muestra

Sexo			
	Frecuencia	Porcentaje	Porcentaje Acumulado
Masculino	19	76	76
Femenino	6	24	100
	25	100	
Edad			
	Frecuencia	Porcentaje	Porcentaje Acumulado
Entre 20 y 40	20	80	80
Entre 40 y 60	5	20	100
	25	100	
Conocimiento Informático			
	Frecuencia	Porcentaje	Porcentaje Acumulado
Ofimático	6	24	24
Avanzado	19	76	100
	25	100	

5.2.2. Descripción de la prueba

Para la evaluación de la cabeza mecatrónica se ha diseñado una prueba en donde el usuario reconozca ciertos gestos que el robot realice. También se analizan otros parámetros como son: dinámica de los gestos, naturalidad, habla e interacción. Un ejemplo de la encuesta se puede ver en el anexo c.

La prueba consiste en mostrar doce rostros, se ha pedido al usuario que anote que expresión reconoce en el robot. Las expresiones utilizadas en la prueba son las seis expresiones básicas: alegría, tristeza, disgusto, sorpresa, ira y miedo a dos diferentes intensidades y en un orden aleatorio. Se cuenta con la posibilidad de que el encuestado marque “ninguna” en caso de no reconocer

la expresión. Con esto se ha buscado analizar si los rostros mostrados por el robot son fácilmente reconocibles.

5.2.3. Resultados

La tabla 5.2 presenta las frecuencias y el porcentaje de reconocimiento de las seis expresiones básicas con una intensidad del cincuenta por ciento.

Tabla 5.2: Frecuencia y porcentaje de expresiones reconocidas al cincuenta por ciento de intensidad

	Frecuencia	Porcentaje
Miedo	5	20
Alegría	9	36
Tristeza	11	44
Sorpresa	17	68
Enfado	19	76
Disgusto	5	20

La figura 5.4 representa la tasa de reconocimiento de las diferentes expresiones al cincuenta por ciento de intensidad. En el eje x se indican las seis expresiones básicas, mientras que en el eje y se indica el porcentaje de reconocimiento de la expresiones obtenidas en la encuesta. Cada expresión contiene 7 columnas que indican las diferentes posibilidades de la encuesta.

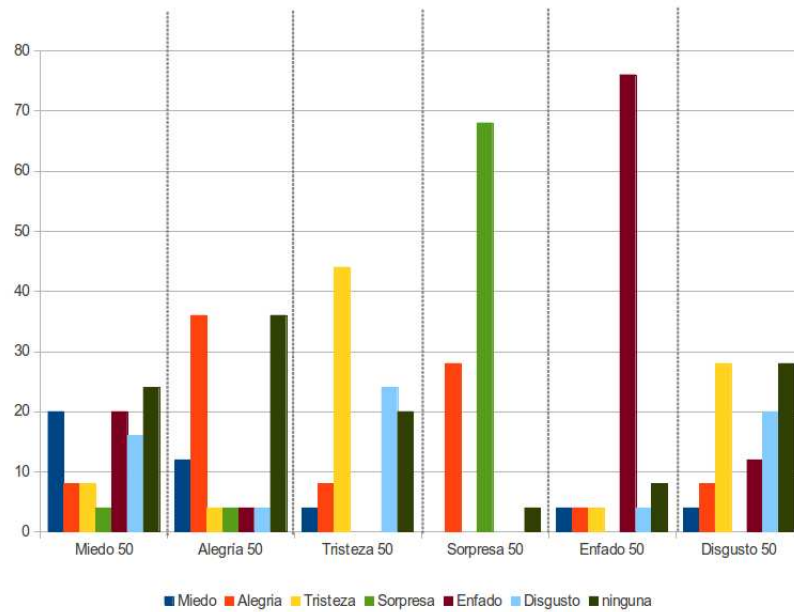


Figura 5.4: Expresiones básicas al cincuenta por ciento de intensidad frente al porcentaje de expresiones reconocidas por los usuarios.

La tabla 5.3 presenta las frecuencias y el porcentaje de reconocimiento de las seis expresiones básicas con una intensidad del cien por ciento.

Tabla 5.3: Frecuencia de reconocimiento de expresiones al cien por ciento de intensidad

	Frecuencia	Porcentaje
Miedo	9	36
Alegría	8	32
Tristeza	18	72
Sorpresa	21	84
Enfado	19	76
Disgusto	4	16

La figura 5.5 representa la tasa de reconocimiento de las diferentes expresiones al cien por ciento de intensidad. En el eje x se indican las seis expresiones básicas, mientras que en el eje y se indica el porcentaje de reconocimiento de las expresiones obtenidas en la encuesta. Cada expresión contiene 7 columnas que indican las diferentes posibilidades de la encuesta.

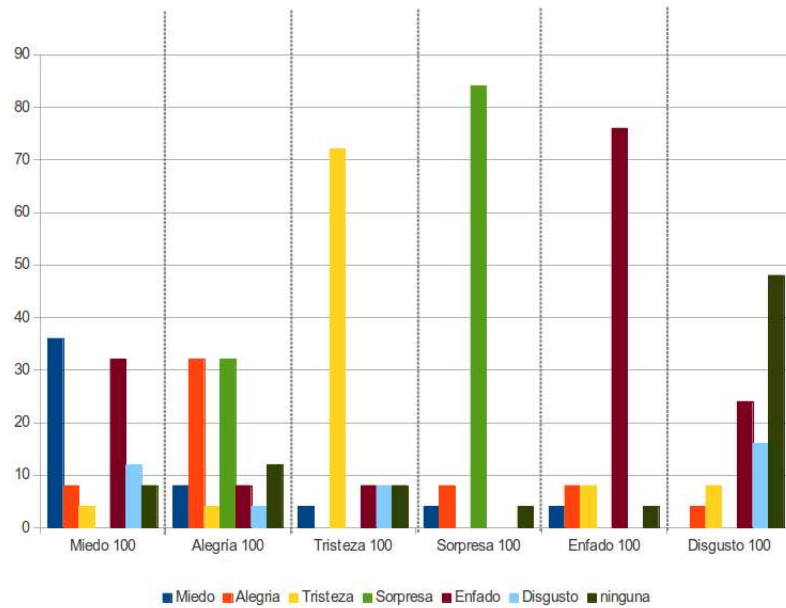


Figura 5.5: Expresiones básicas al cien por ciento de intensidad frente al porcentaje de expresiones reconocidas por los usuarios.

La tabla 5.4 presenta el análisis anova realizado entre las expresiones del mismo tipo a diferentes intensidades.

Tabla 5.4: Analisis anova entre grupos de expresiones

	F	Sigma
Sorpresa	3.122	0.026
Miedo	1.219	0.311
Alegría	0.144	0.988
Tristeza	2.040	0.127
Disgusto	1.985	0.127
Enfado	3.262	0.024

La tabla 5.5 presenta la valoración que se realizó por los usuarios a diferentes variables como son: naturalidad, habla, interacción. Se calculan máximos, mínimos, promedio y desviación estándar en cada una de las variables.

Tabla 5.5: Promedio y desviaciones estándar para cada variable analizada

	N	Mínimo	Máximo	Promedio	Desviación std.
Evaluación general de rostros	25	4	10	6.96	1.72
Fluidez	25	2	10	6.56	1.78
Habla	25	2	9	5.20	2.04
Interacción	25	5	10	7.80	1.78

En las variables analizadas se encuentra la evaluación general, en la que se pide al usuario que analice los diferentes rostros que ha mostrado el robot y califique el grado de significación que le ha dado. La fluidez tiene relación directa con los movimientos, si relativamente parecen naturales. En el habla se evalúa tanto los visemas generados así como la voz y si esta tiene un buen nivel de aceptación por parte del usuario. En la variable de la interacción se ha obtenido la valoración de todos los aspectos anteriores juntos.

5.2.4. Discusión

De los resultados obtenidos previamente se puede observar que el robot presenta ciertas expresiones claramente: sorpresa, tristeza y enfado, tanto al 100 como al 50 por ciento. Esto se debe a que existen unidades de acción muy notorias que identifican claramente a la expresión, tal es el caso de la AU26 (bajar la mandíbula) que identifica claramente a la expresión de sorpresa. Por otra parte, no poder mostrar ciertas unidades de acción afecta de manera negativa a ciertas expresiones tal es el caso del disgusto donde la AU9 (arrugar la nariz) no permite identificar claramente esta expresión.

A partir del análisis anova tanto en la sorpresa (σ 0.026) como en el enfado (σ 0.024) se ve una dependencia entre las expresiones a diferentes intensidades. Para las demás expresiones se tiene un σ mayor al 0.05 esto no descarta la hipótesis y por lo tanto se necesita realizar más pruebas.

En cuanto a los aspectos generales evaluados, de acuerdo al promedio y desviaciones estándar, se han obtenido resultados positivos. Teniendo mejores resultados en la interacción con la personas (7.8). El elemento menos apreciado es el habla con 5.20 esto se debe a que se evaluó en el robot una versión de prueba del sistema que actualmente se está implementado. Se espera en versiones futuras conseguir un sistema de habla más natural que mejore su aceptación.

A pesar del aspecto totalmente mecánico del robot, la evaluación ha resultado satisfactoria. Se necesita realizar diferentes pruebas en cuanto se tenga la máscara de silicona, comparar los resultados y ver como son afectados las diferentes variables del experimento.

Capítulo 6

Conclusiones y futuros desarrollos

“Si eliminamos lo imposible, lo que nos quede,
por improbable que parezca, será la solución”.
– *Conan Doyle*

En este capítulo se exponen las conclusiones y recomendaciones que se han obtenido a partir del trabajo realizado. También se abordan las futuras líneas de investigación que se plantean a partir de las incógnitas generadas en la investigación.

6.1. Conclusiones

Con la realización del presente trabajo se han extraído las siguientes conclusiones:

- A través del análisis del estado del arte en robots sociales se ha podido entender conceptos fundamentales que han sido aplicados en el trabajo fin de master. También se ha investigado las diferentes aportaciones recientes en este campo de la robótica con lo cual se ha obtenido una idea de lo últimos logros y adelantos en este campo. Por último se han analizado los desarrollos previos que se han realizado en el grupo de investigación de la Fundación Cartif en este campo lo cual ha servido para implementar algunas ideas interesantes en el proyecto desarrollado.
- Partiendo del análisis de la fisiología del rostro humano se ha realizado aproximaciones mecánicas de los principales huesos y músculos. De este proceso se ha obtenido un modelo mecánico simplificado de tres zonas

significativas de la cabeza: frontal, mandíbula superior y mandíbula inferior. También, se han detallado algunos mecanismos especiales de ojos, cuello y boca y las aproximaciones hechas a partir de la estructura muscular. De este análisis se ha construido un modelo mecánico con 23 grados de libertad.

- Para el control de expresiones se han realizado una arquitectura modular basada en el FACS. Para ello, se han desarrollado tres módulos principales: el módulo controlador de servos que es el encargado de controlar las limitaciones mecánicas del modelo, el módulo de unidades de acción configura 32 unidades de acción seleccionadas y el módulo gestor de emociones que configura las expresiones básicas. Con esto se ha obtenido un sistema que permite el desarrollo de expresiones básicas de forma parametrizada.
- La visión en el robot está integrada por tres elementos: detección de movimiento, rostros y objetos. Para ello se han implementado los algoritmos: captura óptica de movimiento por sustracción de vídeo en fotogramas anteriores para la detección de movimiento, Viola-Jones para la detección de rostros, Camshift para la detección de objetos y Lucas-Kanade para el seguimiento. Para cada uno de ellos se ha analizado el algoritmo con el cual se ha implementado. También se ha desarrollado un sistema de seguimiento que acciona las unidades de acción de ojos y cuello en el robot. A través de este módulo el robot es capaz de detectar diferentes elementos en su entorno y utilizar esta información en módulos superiores.
- Se han implementado diferentes módulos que han agregado diferentes funcionalidades al robot. Un módulo de habla el cual genera automáticamente los visemas y ejecuta de manera simultánea la señal auditiva y el archivo de animación. También se ha agregado un módulo de conocimiento en donde el robot busca información que el usuario podría necesitar, por el momento puede encontrar información en la Wikipedia, contar chistes, recordar fechas importantes, decir la hora, etc. Por último se ha implementado un sistema de decisión basada en máquina virtual el cual permite realizar tareas de manera autónoma al robot. Todos estos módulos desarrollados han ayudado a mejorar las capacidades sociales del robot.
- Se ha desarrollado un software que facilita el proceso de generar expresiones complejas y animaciones. El programa dispone de tres niveles

de operación: a nivel de unidades de acción, expresiones complejas, sincronización de labios. También se ha explicado el proceso de desarrollo de animaciones a través de fotogramas claves y su implementación en el robot. El desarrollo de esta herramienta a permitido el trabajar con el robot de forma más rápida y de manera más eficiente.

- Se ha evaluado la interacción del robot con las personas las cuales ha dado resultados positivos. Con este fin se ha diseñado una encuesta. Las expresiones más reconocidas han sido: sorpresa, tristeza y enfado. Otras expresiones no han sido del todo reconocidas debido a que el prototipo no tiene la máscara de silicona la cual facilitaría el reconocimiento de ciertas unidades de acción. Factores como habla, naturalidad, expresividad han dado criterios positivos bajo la misma evaluación.

6.2. Trabajos futuros

Se ha dotado al robot de las características sociales básicas, como son: la visión, detección de objetos, detección de rostros, seguimiento y habla. También se ha logrado incrementar la sensación de copresencia cuando se interacciona con el robot. De este proceso han surgido diferentes ideas e interrogantes que servirán para desarrollos futuros.

Para mejorar la aceptación del robot es necesario incrementar el nivel de realismo. Por ello se va implementar una máscara de silicona la cual resaltará las posibilidades gestuales del robot.

Con fines de lograr un mayor grado de naturalidad se ha visto la necesidad de mejorar la capacidad de habla del robot. Se trabaja en integrar un mejor sistema de habla así como un sistema de reconocimiento de voz. También se desea implementar un chatbot que permita interaccionar de manera más natural con los usuarios.

El robot tiene la capacidad de integrar un sistema de visión binocular que se podría explotar para dotarle de visión tridimensional. También el sistema gestual puede ser mejorado para obtener mejores respuestas y una interacción más natural.

Actualmente se trabajan en algunas de estas ideas para en un futuro integrarlas al proyecto. La plataforma permite el desarrollo de investigación en muchos campos de la robótica social lo cual permitirá en un futuro el desarrollo de robots con mejores capacidades sociales.

6.3. Publicaciones

- Loza David, Marcos Pablos Samuel, Zalama Casanova Eduardo, Gómez García-Bermejo y Jaime González José Luis. “Application of the FACS in the design and construction of a mechatronic head with realistic appearance”. *Journal of Physical Agents*. Vol. 7, No. 1 (Jan. 2013). ISSN 1888-0258, pp. 31-38
- Loza David, Marcos Pablos Samuel, Zalama Casanova Eduardo, Gómez García-Bermejo y Jaime González José Luis. “Aplicación del FACS en el diseño y construcción de una cabeza mecatrónica de aspecto realista”. XIII Workshop de agentes físicos, (Dic. 2012)
- Robótica social: Autómatas humanos. Robolid 2013, Abril 2013. Ponente: Eduardo Zalama Casanova. Salón de Grados EII (Paseo del Cauce). Presentación del proyecto.
- Robot 2013. Loza David, Marcos Pablos Samuel, Zalama Casanova Eduardo, Gómez García-Bermejo y Jaime González José Luis. “Animation of expressions in a mechatronic head”. First Iberian Robotics Conference, (Noviembre 2013).

Apéndice A

Técnica de modulación por ancho de pulso (MAP)

La modulación por ancho de pulso MAP (en inglés pulse-width modulation PWM) es una técnica muy utilizada para controlar servomotores, la cual modifica el ancho del pulso de acuerdo a la posición deseada.

Las principales características de la técnica son: que obtiene resultados analógicos a partir de un control digital, mantiene el par del motor constante y no se desaprovecha corriente eléctrica. El servo mantendrá la posición a pesar de que exista una fuerza externa, la magnitud de la fuerza que podrá resistir dependerán de las características dadas por el fabricante. Se puede ver una explicación de la técnica en la figura A.1.

La figura A.1 muestra las diferentes posiciones del servo a determinados anchos de pulso, a partir de esta relación se puede calcular el ciclo de trabajo de una señal periódica que es el ancho relativo de su parte positiva en relación con el período. La fórmula A.1 indica cómo se calcula el ciclo de servicio o “*Duty Cycle*”.

$$DT = t_{alta} / (t_{alta} + t_{baja}) \quad (A.1)$$

El ancho de pulso puede ser generado por un procesador z80, o microprocesador (por ejemplo el PIC 16F877A). La tarjeta controladora en el robot es la SSC32 que tiene un microcontrolador ATMega 168-20PU que permite entre otras funciones generar el pulso requerido.

Además de controlar servomotores existen muchas aplicaciones para la técnica de modulación por pulso como son: regulación de voltaje, efectos de audio, amplificación, telecomunicaciones, etc.

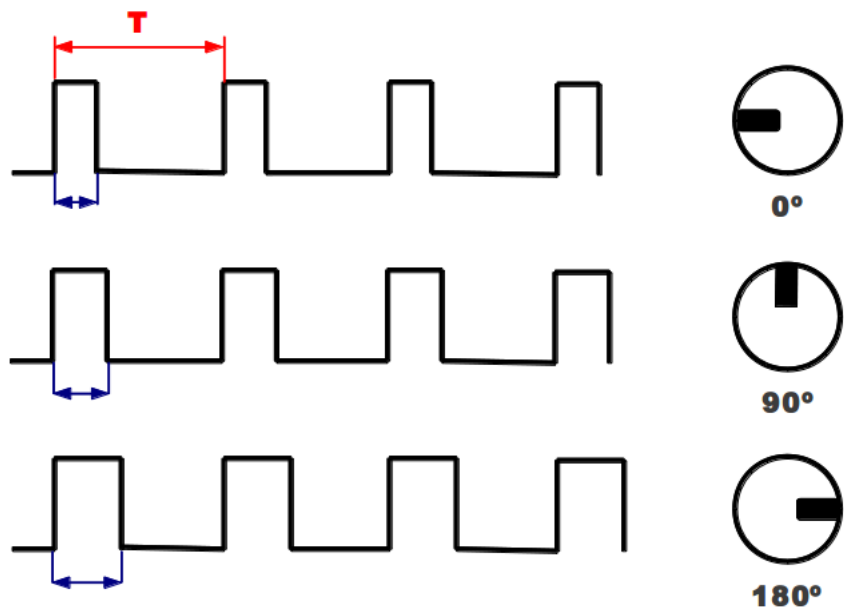


Figura A.1: Técnica de modulación por ancho de pulso

Apéndice B

Estructura en ROS del robot

En el apéndice se detalla la estructura de ROS, su estructura de archivos así como los medios de comunicación entre nodos en ROS. Después, se analiza la estructura implementada en el robot.

B.1. Estructura en ROS

ROS tiene un sistema de archivos estructurados de la siguiente forma:

- **Stacks:** Es una colección de paquetes que conforman una librería de alto nivel.
- **Paquetes:** Puede contener elementos diferentes, por ejemplo: librerías, ejecutables, herramientas, etc.
- **Nodos:** Es un ejecutable que utiliza ROS para comunicarse con otros programas o nodos.

Existen diferentes elementos que permiten la comunicación entre nodos:

- **Mensaje:** Un mensaje es una estructura predefinida que se utiliza cuando se suscribe o se publica algo.
- **Topics:** Un topic es una canal de comunicación entre nodos. Se puede tanto suscribir o publicar mensajes.
- **Servicios:** Los servicios son otra forma de comunicación entre nodos. Un servicio obliga a que los nodos envíen una solicitud y reciban una respuesta.

- **Actions:** Es un conjunto de servicios específicos que se utilizan con el fin de lograr un desempeño deseado. Se cuenta con elementos que permitan ver el objetivo, información y resultado del proceso.

B.2. Estructura del robot en ROS

A continuación se detallan los diferentes paquetes con una pequeña descripción y sus respectivos nodos programados.

- ***msg_cab_mec*:** Paquete dedicado a guardar los mensajes propios del robot.
- ***srv_cab_mec*:** Paquete dedicado a guardar los servicios propios del robot.
- ***hablar*:** Paquete que contiene los nodos dedicados a la configuración de visemas y reproducción de audio.
 - visemas.py: Programa independiente a ROS que sirve para obtener los visemas del archivo generado pho.
 - hablar.py: Nodo que ejecuta simultáneamente los visemas y fonemas de la "frase a decir".
 - decir.py: Nodo que envía una frase al nodo hablar.
- ***chatbot*:** Paquete que contiene el chatbot del robot.
 - chatbot.cpp: Nodo en c++ para el chatbot implementado en el robot.
- ***ia*:** Paquete de inteligencia del robot, contiene los siguientes nodos:
 - demo.py: Muestra una demo con las capacidades del robot.
 - statemachine.py: Nodo que contiene las máquinas de estados del robot.
 - machine.py: Librería de la máquina de estados.
 - showimage.py: Nodo que muestra una imagen en el escritorio.
- ***reconocimiento*:** Paquete de reconocimiento.
 - reconocimiento.cpp: Nodo que envía un mensaje al chatbot.

- ***servomotores***: Paquete que contiene los nodos necesarios para el manejo de los servo motores y unidades de acción.
 - `mua.py`: Nodo que tiene configurado todas las unidades de acción implementadas en el robot.
 - `gestor_gestos.py`: Contiene los gestos básicos, así como la base de gestos.
 - `comportamiento.py`: Ejecuta una secuencia de gestos predeterminada.
 - `teleop.py`: Modo del joystick, guarda la relación entre los diferentes botones del joystick y la unidades de acción y gestos.
 - `ros_ssc32.py`: Nodo en ROS para el control de la tarjeta ssc32
 - `libs_sc32.py`: Librería de la tarjeta ssc32.

- ***vision***: Paquete que contiene los nodos de vision.
 - `vision.py`: Nodo que contiene las diferentes configuraciones de visión del robot: detección de rostros, detección de movimiento y detección por color.
 - `tracker.py`: Nodo que maneja las unidades de acción tanto de los ojos como el cuello, permite el seguimiento de objetos.
 - `cam.py`: Nodo para verificar el buen funcionamiento de las cámaras en el robot.

- ***gui***: Paquete que contiene las interfaces de usuario.
 - `user.py`: Nodo que permite el manejo del robot a distancia a través de una interface de usuario.
 - `prog_gestos.py`: Nodo que permita la configuración de expresiones y de secuencias animadas, así como visualización de herramientas.

- ***info***: Paquete que contiene lo nodos de información.
 - `info.py`: Nodo que maneja la información en el robot. Entre sus funciones esta: buscar información en una url específica, contar chistes, recordar eventos y fechas importantes.
 - `pedir_info.py`: Nodo que sirve para adquirir información del nodo info.
 - `wiki2plain.py`: Librería que permite arreglar la información adquirida de la Wikipedia.

La figura B.1 presenta la estructura en ROS del robot.

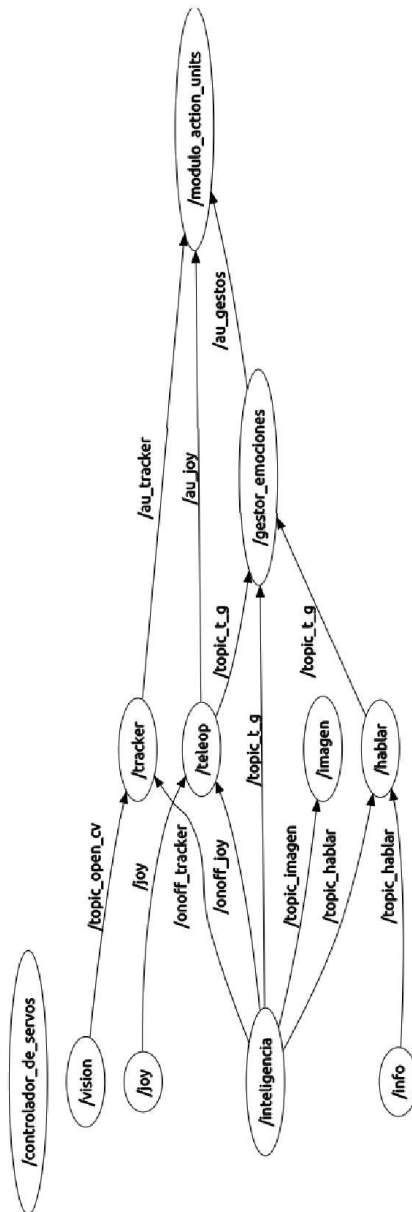


Figura B.1: Estructura en ROS del robot

Apéndice C

Formato de la encuesta realizada



ENCUESTA PARA RECONOCIMIENTO
DE EXPRESIONES CON LA CABEZA
MECATRÓNICA

Por favor, conteste la siguiente información básica:

Sexo : M F

Edad : Menos de 20 Entre 20 y 40 Entre 40 y 60 Más de 60

Conocimientos informáticos :

Nunca he usado ordenador

Solo navegar por internet

Experiencia ofimática

Avanzado

A continuación se presentan 12 gestos con la cabeza mecatrónica. Por favor marque con una 'x' la casilla de la expresión que usted a percibido por parte del robot. Solo es necesario una respuesta por cada expresión.

	Miedo	Alegría	Tristeza	Sorpresa	Enfado	Disgusto	No se reconoce
Expresión 1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Expresión 2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Expresión 3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Expresión 4	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Expresión 5	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Expresión 6	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Expresión 7	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Expresión 8	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Expresión 9	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Expresión 10	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Expresión 11	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Expresión 12	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

1. Valore en general del 1 al 10 (1=muy malas, 10=muy buenas) las expresiones que ha mostrado el robot.

1 2 3 4 5 6 7 8 9 10

2. Valore la naturalidad de las expresiones mostrados por el robot, del 1 al 10 (1=poco naturales, 10=muy naturales).

1 2 3 4 5 6 7 8 9 10

3. ¿Creé que el robot se expresa con naturalidad ? Valore la naturalidad con la que habla el robot (1=poco natural, 10=muy natural).

1 2 3 4 5 6 7 8 9 10

4. ¿Le ha gustado interactuar con el robot? (1=poco satisfactorio, 10=muy satisfactorio).

1 2 3 4 5 6 7 8 9 10

5. Si lo desea puede dejarnos un comentario de como mejorar nuestro trabajo.

.....
.....

Muchas gracias por su colaboración!!!!

Bibliografía

- [1] Masahiro Mori. *The Buddha in the robot*. Tuttle, Boston, MA, 1982.
- [2] A. Freitas-Magalhães. The psychology of emotions: The allure of human face. *Oporto: University Fernando Pessoa Press*, 2007.
- [3] Salvador Domínguez. *Contribución a la robótica social: Arisco, un robot con capacidad de interacción y aprendizaje*. PhD thesis, Universidad de Valladolid, 2007.
- [4] Samuel Marcos, Jaime Gómez-García-Bermejo, and Eduardo Zalama. A realistic, virtual head for human-computer interaction. *Interact. Comput.*, 22:176–192, 2010.
- [5] B. Duffy. *The Social Robot*. PhD thesis, Univ. College Dublin, 2000.
- [6] Kerstin Dautenhahn and Aude Billard. Bringing up robots or - the psychology of socially intelligent robots: From theory to implementation. pages 366–367, 1999.
- [7] Cynthia L. Breazeal. *Designing Sociable Robots*. MIT Press Cambridge, MA,, 2002.
- [8] C. Bartneck and J. Forlizzi. *A design-centred framework for social human-robot interaction*. IEEE, 2004.
- [9] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. A survey of socially interactive robots. *Robotics and Autonomous Systems - ELSEVIER*, 42:143–166, 2003.
- [10] Rodney A. Brooks. A robot that walks; emergent behavior from a carefully evolved network. *Neural Computation*, 1:2:253–262, 1989.
- [11] Rodney A. Brooks. Elephants don't play chess. *Robotics and Autonomous Systems*, 6:3–15, 1990.

- [12] Eric Bonabeau, Marco Dorigo, and Guy Theraulaz. *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press, 1999.
- [13] Cynthia Breazeal. *Sociable Machines: Expressive Social Exchange Between Humans and Robots*. PhD thesis, MIT, 2000.
- [14] Kazuko Itoh, Hiroyasu Miwa, Yuko Nukariya, Massimiliano Zecca, Hideaki Takanobu, Stefano Roccella, Maria Chiara Carrozza, Paolo Dario, and Atsuo Takanishi. Mechanisms and functions for a humanoid robot to express human-like emotions. In *IEEE International Conference on Robotics and Automation*, 2006.
- [15] Jochen Hirth, Norbert Schmitz, and Karsten Berns. Emotional architecture for the humanoid robot head roman. *Robotics and Automation, IEEE*, 53:2150 – 2155, 2007.
- [16] Shuuji Kajita, Kenji Kaneko, Fumio Kaneiro, Kensuke Harada, Mitsuharu Morisawa, Shin’ichiro Nakaoka, Kanako Miura, Kiyoshi Fujiwara, Ee Sian Neo, Isao Hara, Kazuhito Yokoi, and Hirohisa Hirukawa. Cybernetic human hrp-4c: A humanoid robot with human-like proportions. *Springer Tracts in Advanced Robotics*, 70:301–314, 2011.
- [17] W. V. Friesen and P. Ekman. Emotional facial action coding system. 1983.
- [18] Edward Tronick, Heidelise Als, and T. Berry Brazelton. Monadic phases: A structural descriptive analysis of infant–mother face to face interaction. *Merrill-Palmer Quarterly*, 26:3–24, 1980.
- [19] Carroll Izard. Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *American Psychological Association*, 115:288–299, 1994.
- [20] P. Ekman and W. Friesen. Facial action coding system: A technique for the measurement of facial movement. *Consulting Psychologists Press*, 1978.
- [21] CH Hjortsjö. *Man’s face and mimic language*. 1969.
- [22] S. Villagrana and A. Susin. Face! 3d facial animation system based on facts. *IV Iberoamerican Symposium in Computer Graphic*, 2009.
- [23] Kwang-Eun Ko and Kwee-Bo Sim. Development of a facial emotion recognition method based on combining aam with dbn. *International Conference on Cyberworlds*, 2010.

- [24] Jihun Hamm, Christian G. Kohler, Ruben C. Gur, and Ragini Vermaa. Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders. *J Neurosci Methods*, 200(2):237–256, 2011.
- [25] Brian R. Duffy. Anthropomorphism and the social robot. *Robotics and Autonomous Systems - ELSEVIER*, 42:177–191, 2003.
- [26] Brian R. Duffy. Anthropomorphism and robotics. *The Society for the Study of Artificial Intelligence and the Simulation of Behaviour*, pages 3–5, 2002.
- [27] K. L. Nowak and C. Rauh. Choose your "buddy icon" carefully: The influence of avatar androgyny, anthropomorphism and credibility in online interactions. *Computers in Human Behaviour*, 24:1473–1493, 2008.
- [28] R. Simmons. Grace: The social robot. In *Mobile Robot Competition and Exhibition, Edmonton,*, 2002.
- [29] K. L. Norwak. The influence of anthropomorphism and agency on social judgement in virtual environments. *Journal of Computer Mediated Communication*, 9(2), 9:2, 2004.
- [30] Laurel D. Riek, Tal-Chen Rabinowitch, Bhismadev Chakrabarti, and Peter Robinson. How anthropomorphism affects empathy toward robots. *HRI '09 Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 245–246, 2009.
- [31] Jelle Saldien, Kristof Goris, Selma Yilmazyildiz, Werner Verhelst, and Dirk Lefeber. On the design of the huggable robot probot. *Journal of Physical Agents*, 2:2, 2008.
- [32] Karl F. MacDorman and Hiroshi Ishiguro. The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, 7:297–337, 2006.
- [33] David Hanson, Andrew Olney, Ismar A. Pereira, and Marge Zielke. Upending the uncanny valley. *AAAI*, pages 1728–1729, 2005.
- [34] Karl F. MacDorman. Androids as an experimental apparatus: Why is there an uncanny valley and can we exploit it? 2005.
- [35] B. Tondu and N. Bardou. Aesthetics and robotics : which form to give to the human-like robot ? *World Academy of Science, Engineering and Technology*, 34, 2009.

- [36] M. Fabri. *Emotionally Expressive Avatars for Collaborative Virtual Environments*. PhD thesis, Leeds Metropolitan University, 2006.
- [37] Adams RB Jr and Kleck RE. Perceived gaze direction and the processing of facial displays of emotion. *Psychological Science*, 14:644–647, 2003.
- [38] Adams RB Jr and Kleck RE. Effects of direct and averted gaze on the perception of facially communicated emotion. *Emotion*, 5:3–11, 2005.
- [39] U. Hess, Jr. Adams, R. B., and R. E. Kleck. Looking at you or looking elsewhere: The influence of head orientation on the signal value of emotional facial expressions. *Motivation and Emotion*, 31:137–144, 2007.
- [40] Morgan Quigley, Brian Gerkey, Ken Conley, Josh Faust, Tully Foote†, Jeremy Leibs, Eric Berger, Rob Wheeler, and Andrew Ng. Ros: an open-source robot operating system. 2008.
- [41] Shin’ichiro Nakaoka, Fumio Kanehiro, Kanako Miura, Mitsuharu Morisawa, Kiyoshi Fujiwara, Kenji Kaneko, Shuuji Kajita, and Hirohisa Hirukawa. Creating facial motions of cybernetic human hrp-4c. *9th IEEE-RAS International Conference on Humanoid Robots*, pages 561–567, 2009.
- [42] Paul Viola and Michael Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.
- [43] Marta Lucía Guevara, Julian David Echeverry, and Wiliam Ardila Urueña. Detección de rostros en imágenes digitales usando clasificadores en cascada. *Scientia et Technica Año XIV*, 38, 2008.
- [44] Y. Freund and R. Shapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Second European Conference on Computational Learning Theory*:23–37, 1995.
- [45] D. Comaniciu, V. Ramesh, and P. Mee. Real-time tracking of non-rigid objects using mean shift. 2000.
- [46] M.J. Swain and D.H. Ballard. Color indexing. *International Journal of Computer Vision*, 7:1, 1991.
- [47] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. *Proceedings of Imaging Understanding Workshoh*, pages 121–130, 1981.

- [48] Jean-Yves Bouguet. Pyramidal implementation of the lucas-kanade feature tracker. 2000.
- [49] F. Perales. Inevai 3d: Agentes autónomos 3d, escenarios virtuales e interfaces inteligentes para aplicaciones de domótica y de realidad virtual. *Simposio de Computación - Ubicua e Inteligencia Ambiental*, 2005.