

TRABAJO FIN DE MÁSTER
(Máster en Tecnologías Informáticas Avanzadas)

**Conocimiento experto y técnicas de softcomputing
aplicadas a la animación de intérpretes virtuales de la
lengua de signos española**

Septiembre de 2009

Autor

Vanesa Herrera Tirado
(Ingeniera en Informática)

Director

Dr. Carlos González Morcillo



Universidad de
Castilla-La Mancha



Escuela Superior
de Informática

© Vanesa Herrera Tirado. Se permite la copia y la distribución de la totalidad o parte de este documento sin ánimo de lucro. Toda copia total o parcial deberá citar expresamente el nombre del autor, de la Universidad de Castilla-La Mancha y deberá incluir esta misma licencia, añadiendo, si es copia literal, la mención "Copia Literal".

Se autoriza la modificación y traducción de la obra sin ánimo de lucro siempre que se haga constar en la obra resultante de la modificación el nombre de la obra originaria, el autor de la obra originaria y el nombre de la Universidad de Castilla-La Mancha. La obra resultante también deberá ser libremente reproducida, distribuida, comunicada al público y transformada en términos similares a los expresados en esta licencia.

Este documento fue maquetado con \LaTeX . Imágenes compuestas con Gimp y OpenOffice.

Resumen

El uso de avatares 3D como asistentes virtuales se ha popularizado en diferentes campos de aplicación, como por ejemplo en el comercio electrónico, en la banca electrónica o en entornos de aprendizaje. En determinadas áreas su uso se convierte en un elemento imprescindible, como en el caso de la traducción automática de la lengua de signos. Este tipo de herramientas utilizan intérpretes virtuales con el fin de mejorar la accesibilidad de las personas sordas a la información. Un avatar (personaje virtual) simula los movimientos de manos, brazos y torso que realizaría una persona real. Gran parte de estas herramientas incluyen también la animación facial, que constituye un factor muy importante especialmente cuando la comunicación se realiza a través de la lengua de signos. Sin embargo, esta mejora de accesibilidad no es suficiente debido al descuido de ciertos factores clave para este tipo de sistemas, como el realismo de las animaciones, la adaptación a las necesidades del usuario, gestión del contexto y del medio de publicación.

En el presente trabajo de investigación se plantea una arquitectura que contempla los factores anteriormente descritos para permitir el uso de intérpretes virtuales de la lengua de signos en el mayor número de ámbitos posible. En primer lugar, se han utilizado diferentes técnicas de *softcomputing* con el fin de manejar la vaguedad e imprecisión que surgen en los distintos procesos de adaptación. Esta ambigüedad está relacionada con la simulación de emociones y que es utilizada con el fin de adaptar la aplicación al contexto en el que se emite el mensaje en lengua de signos. Se trata de un elemento muy importante ya que sin esta información la comunicación puede ser incompleta. Además hay que tratar con la imprecisión relacionada con la adaptación de la herramienta a las necesidades y preferencias particulares de cada usuario, por ejemplo el uso de subtítulos, la apariencia del avatar o el uso de audio en el caso de que el usuario no sufra sordera profunda. Además de estas técnicas, se ha prestado especial atención a la simulación de movimientos del avatar con un alto grado de realismo, siendo este punto esencial para que la información sea realmente entendible. Por otro lado, la aplicación puede utilizarse en una gran variedad de dispositivos como por ejemplo en ordenadores portátiles, teléfonos móviles, PDA's, puntos de información, etc, gracias a la adaptación del vídeo obtenido al dispositivo de visualización.

Finalmente, se ha prestado especial atención en la definición de la arquitectura para que sea directamente generalizable a otros ámbitos de aplicación con avatares virtuales 3D.

Índice general

Índice de figuras	V
Índice de tablas	VI
1. Introducción	1
1.1. Planteamiento general	1
1.2. Objetivos del Trabajo Fin de Máster	3
1.2.1. Marco de trabajo	3
1.3. Estructura del documento	3
2. Estado del arte	5
2.1. Uso de avatares 3D como intérpretes en la lengua de signos	5
2.2. Computación afectiva	7
2.2.1. Simulación de emociones en avatares	9
2.3. Adaptabilidad de avatares virtuales	15
3. Arquitectura propuesta	17
3.1. Introducción	17
3.2. Componentes de la arquitectura	18
3.2.1. Procesamiento del lenguaje natural	19
3.2.2. Representación 3D	20
3.2.3. Renderizado	22
3.2.4. Adaptación al rol del usuario	23
3.2.5. Adaptación al contexto y al mensaje	28
4. Conclusiones y trabajo futuro	35

Bibliografía	37
A. Resumen de las asignaturas cursadas en el Máster	42
A.1. Metodologías y técnicas de investigación en informática	42
A.2. Sistemas distribuidos avanzados: grid e inteligencia ambiental	43
A.3. Sistemas avanzados de interacción persona-computador: sistemas colabo- rativos y computación ubícuca	44
A.4. Pruebas y seguridad de Sistemas de Información	45
A.5. Cognición y colaboración	45
A.6. Técnicas de softcomputing	46
B. Currículum Vitae	48

Índice de figuras

2.1. Estructura cognitiva de emociones del modelo OCC [38]	8
2.2. Imagen superior: desilusión enmascarada con alegría. Imagen inferior: superposición de tristeza y alegría. Basada en la Figura del trabajo [33]	10
2.3. Esquema adaptado del propuesto por Raouzaïou et al. [44]	11
2.4. Curso temporal de una expresión facial según el trabajo de Bevacqua, Mancini y Pelachaud [9]	11
2.5. Espacio tridimensional de emociones según Araya et al. [6]	12
2.6. Conjuntos difusos definidos en el trabajo [6].	13
2.7. Esquema del sistema propuesto por [10].	14
3.1. Esquema general de la propuesta.	19
3.2. Diferentes personajes virtuales y descripción del mismo movimiento.	22
3.3. Red bayesiana obtenida	25
3.4. Perfil obtenido sin conocimiento de las preferencias del usuario.	26
3.5. Propagación de evidencia conociendo sólo la edad del alumno (entre 7 y 12 años).	27
3.6. Propagación de la evidencia en el conocida la edad y las respuestas del alumno.	27
3.7. Variable v_i , emoción del avatar. $DDV_i = \{N, A, MA\}$; donde $i = \{1, 2, 3, 4, 5, 6\}$. 30	
3.8. Manipuladores de alto nivel definidos para las expresiones faciales	31
3.9. Conjuntos difusos difusos para las variables de salida. Izquierda: <i>frames totales</i> . Derecha: <i>frames on</i> y <i>frames off</i>	32
3.10. Resultados obtenidos aplicando las reglas difusas. 1. Neutro. 2. Sorpresa alta 3. y alegría alta. 4. Tristeza alta. 5. Tristeza muy alta. 6. Asco muy alto	33

Indice de tablas

2.1. Valores dimensionales para las emociones estándar obtenidas del trabajo de Araya, DiPaol y Parush [6].	13
3.1. Parámetros de render.	18
3.2. Variables definidas para la recolección de evidencias.	25
3.3. Variables de entrada y su dominio de definición.	29
3.4. Variables de salida y su dominio de definición.	32
3.5. Emociones y su correspondiente descripción facial.	33

1

Introducción

1.1. Planteamiento general

Las personas con discapacidad auditiva se encuentran a diario con grandes barreras en el acceso a la información y a la educación. Esto se debe a que la sociedad es mayoritariamente oyente y por tanto utilizan la lengua oral como vía de comunicación, lengua que es inaccesible para las personas sordas.

A diferencia de la comunidad oyente, la lengua utilizada por las personas sordas es la lengua de signos. Su uso es imprescindible para que la comunidad no oyente tenga acceso a la información que les rodea. Se trata de una lengua completa y con grandes diferencias respecto a la utilizada por personas oyentes. A pesar de que la lengua de signos es la lengua natural de las personas sordas esto no significa que sea igual en todo el mundo, no se trata de una lengua universal [45], debido a que una misma seña no tiene que significar lo mismo en distintos puntos de la geografía. Estas variantes se producen incluso entre diferentes regiones de un mismo país. Sin embargo, si se establece comunicación entre dos personas de distintas partes del mundo, podrán entenderse puesto que gran parte de las señas son icónicas.

En cuanto a su representación, hay que tener en cuenta que no sólo requiere el movimiento de las manos y brazos, sino que los componentes no manuales como los ojos, expresiones faciales y corporales son necesarios para expresar adecuadamente la información [49]. Además, este elemento es esencial para evitar que la comunicación sea imprecisa e incompleta, así por ejemplo su uso permite diferenciar entre una pregunta, una negación o una afirmación. El comportamiento lingüístico del sordo está formado por diferentes componentes interrelacionados entre sí [46]:

- **Signos manuales:** referido al movimiento de las manos para representar palabras o letras en lenguaje de signos.
- **Quinésica facial:** referida al uso del movimiento de los labios.

- **Quinésica somática:** referida al movimiento corporal.
- **Quinésica oral:** referida al uso del movimiento de los labios junto a los gestuales.
- **Dactilología:** representación de cada letra del alfabeto por un signo.

A pesar de que la lengua de signos es una lengua completa (posee unas características gramaticales, sintácticas y léxicas propias), no ha sido reconocida legalmente como tal en España hasta el año 2007: *LEY 27/2007, de 23 de octubre, por la que se reconocen las lenguas de signos españolas y se regulan los medios de apoyo a la comunicación oral de las personas sordas, con discapacidad auditiva y sordociegas.*

La aprobación de esta Ley supone un gran paso para la comunidad sorda, ya que les garantiza el acceso a la información en *transportes, relaciones con las Administraciones Públicas, participación política, medios de comunicación social, telecomunicaciones y sociedad de la información.* Para hacer posible la accesibilidad se tiene en cuenta el uso de "medios de apoyo a la comunicación oral", es decir, medios de comunicación, recursos tecnológicos y ayudas técnicas (entre otros) que faciliten el *acceso a la expresión verbal y escrita de la lengua oral:*

"En la presente Ley se establecen las medidas y garantías necesarias para que las personas sordas, con discapacidad auditiva y sordociegas puedan, libremente, hacer uso de las lenguas de signos españolas y/o de los medios de apoyo a la comunicación oral en todos las áreas públicas y privadas, con el fin de hacer efectivo el ejercicio de los derechos y libertades constitucionales, y de manera especial el libre desarrollo de la personalidad, la formación en el respeto de los derechos y libertades fundamentales, el derecho a la educación y la plena participación en la vida política, económica, social y cultural".

En las últimas décadas han aumentado los estudios relacionados con el acceso a la información en personas con discapacidad auditiva. No sólo relacionado con la información oral, sino también con la escrita ya que el proceso de aprendizaje de la lengua escrita está muy ligado con la audición. Es por eso que los sordos prelocutivos, es decir, aquellos que sufrieron sordera antes de adquirir la concepción abstracta del lenguaje oral, tienen serios problemas en el aprendizaje de la lectura. Esto significa que el uso de subtítulos no es siempre la mejor opción a la hora de traducir la información oral.

Dentro de los *medios de apoyo de comunicación oral* descritos en la Ley anterior, podemos encontrar personajes virtuales que simulan los movimientos realizados por un intérprete real de la lengua de signos. Teniendo en cuenta las características comentadas anteriormente de la lengua de signos, en el desarrollo de este tipo de herramientas es necesario permitir la adaptación del mensaje en dos grupos de actuación:

- **Adaptación al usuario final,** es decir, a las necesidades particulares de cada usuario como puede ser el nivel de pérdida auditiva, la edad, el nivel de comprensión de la lengua de signos, etc.
- **Adaptación al mensaje a traducir.** En la comunicación realizada a través de la lengua de signos es esencial el uso de los componentes no manuales y dentro de ellos el uso de las expresiones faciales. A diferencia de las personas oyentes que se apoyan en el sonido para la transmisión de emociones, las personas sordas sólo cuentan con las expresiones faciales como único mecanismo para transmitir toda clase de sentimientos y emociones. De esta forma resulta crítico gestionar correctamente la expresividad de los avatares virtuales empleados en este área de aplicación.

1.2. Objetivos del Trabajo Fin de Máster

Como se ha descrito anteriormente, existe una creciente necesidad relacionada con la accesibilidad a la información por parte de las personas que sufren discapacidad auditiva. Tomando dicha reflexión como punto de partida, en este trabajo se plantea como objetivo principal el uso *de conocimiento experto y técnicas de softcomputing aplicadas a la animación realista de avatares 3D y más concretamente intérpretes virtuales de la lengua de signos española* con el fin de ofrecer un trato personalizado e individualizado dependiendo del usuario final y el contexto de aplicación. Para alcanzar el objetivo planteado se establecen dos subobjetivos:

1. Diseño de una arquitectura genérica y adaptable para la síntesis de la lengua de signos española que permita la adaptación al contexto, al usuario, al nivel de realismo y al medio de publicación final.
2. Definición de mecanismos basados en técnicas de softcomputing para el tratamiento de la incertidumbre, imprecisión e incompletitud existente en la información relativa al contexto y a las características del usuario.

Los componentes que constituyen la arquitectura propuesta se han definido en base a un estudio sobre sistemas existentes los cuales se dividen en tres diferentes áreas, que se integran en el modelo presentado en este trabajo. Por un lado se ha realizado un estudio de las herramientas que se basan en el uso de un personaje virtual que simula los movimientos realizados por un intérprete real de la lengua de signos. En segundo lugar, se ha realizado un estudio de la síntesis de emociones debido a que constituyen un elemento esencial cuando la comunicación se realiza utilizando la lengua de signos. Por último se han investigado diferentes aproximaciones de softcomputing para conseguir la adaptación de una aplicación determinada a las necesidades y preferencias particulares de cada usuario.

Una vez realizado el estudio del estado del arte, se llevará a cabo el diseño e implementación de una arquitectura que fusione las tres áreas descritas anteriormente. Esta arquitectura contará con una serie de módulos independientes que permiten la adaptación de la información (generada por un intérprete virtual) a las necesidades de cada usuario, las características del contexto en el que se produce la transmisión de la información y a las características del dispositivo de visualización.

1.2.1. Marco de trabajo

El trabajo de tesis de máster que aquí se presenta se enmarca dentro de la cátedra Indra-UCLM GANAS (Generador Automático de la Lengua de Signos).

1.3. Estructura del documento

El resto del trabajo está estructurado de la siguiente forma.

- **Capítulo 2.** En este capítulo el estado del arte se divide en tres áreas: uso de

avatares 3D en la lengua de signos, simulación de emociones en avatares y adaptabilidad de avatares virtuales.

- **Capítulo 3.** En este capítulo se expone el trabajo de investigación realizado. En concreto, se presenta una arquitectura en la que se utilizan técnicas de softcomputing con el fin de adaptar un intérprete virtual de la lengua de signos a las necesidades particulares de cada usuario, así como a las necesidades intrínsecas del contexto en el que se produce el mensaje.
- **Capítulo 4.** Se exponen las conclusiones y las líneas de trabajo futuro que darán lugar al desarrollo de una Tesis Doctoral.
- **Anexo A.** Resumen de las asignaturas cursadas en el Máster.
- **Anexo B.** Currículum Vitae.

2

Estado del arte

La revisión del estado del conocimiento llevada a cabo en este capítulo está enfocada principalmente en tres áreas: interpretación de la lengua de signos por medio de avatares 3D, síntesis de emociones en personajes virtuales y adaptación de sistemas, según las necesidades del usuario, basada en técnicas probabilísticas.

2.1. Uso de avatares 3D como intérpretes en la lengua de signos

En la última década se han desarrollado algunos intérpretes virtuales de la lengua de signos. Estos proyectos se pueden dividir en dos grandes grupos: los que se utilizan en un contexto fijo y los traductores. Los primeros cuentan con un número predefinido de palabras y frases en lengua de signos que son interpretados por un personaje, por lo que sólo pueden ser utilizados dentro de la aplicación para la que han sido desarrollados. Por otro lado, los traductores permiten convertir un texto escrito en lengua de signos a través de movimientos interpretados por un avatar para ser posteriormente utilizados en diferentes contextos.

La empresa *Vcom3D* [63] está orientada a la educación y accesibilidad de personas sordas a través de tecnologías 3D a través del uso de avatares que interpretan la lengua de signos americana (ASL). Las aplicaciones desarrolladas pueden dividirse en dos grandes grupos: *SignSmith* y *VCommunicator*. La primera de ellas, *Sign Smith*, engloba a un conjunto de herramientas orientadas a la creación y el uso de intérpretes de signos virtuales (3D): *Sign Smith Studio* que permite añadir intérpretes virtuales, *ASL animations* que consta de un conjunto de animaciones en lengua de signos (actualmente existen cinco volúmenes con diferentes temáticas) y en el diccionario *Sign Smith Illustrate Dictionary*.

ViSiCAST (Virtual Signing, Capture, Animation, Storage and Transmission) es un pro-

yecto financiado por “*European Union Fifth Framework*” que es parte del programa de “*Information Society Technologies*” (IST). El principal objetivo de ViSiCAST [7] es el de mejorar la calidad de vida de la comunidad sorda europea, permitiendo el acceso a información adaptándola a su mecanismo natural de comunicación. El proyecto identifica una serie de situaciones en las que sería posible la integración de las personas sordas si se dispusiera de comunicación en lengua de signos. Estas son: acceso a servicios públicos, transacciones comerciales, uso de la televisión interactiva, comercio electrónico y la World Wide Web. Con el fin de alcanzar el objetivo final, los desarrolladores plantearon tres subobjetivos:

- Comunicación “cara a cara” mediante un intérprete virtual de signos que sea utilizado en las oficinas de correos de Reino Unido.
- Uso del intérprete virtual en al menos cuatro programas de televisión.
- Instalación de las herramientas WWW y uso de las mismas en sitios web. Estas páginas serán evaluadas por un número representativo de personas sordas.

Una ampliación del proyecto VisiCAST es *eSIGN (Essential Sign Language Information on Government Networks)* [64]. Su principal objetivo es el de integrar a las personas sordas en la administración electrónica (eGovernment). Actualmente se encuentra implantado en diferentes páginas web de Reino Unido, Holanda y Alemania. Los desarrolladores de *eSIGN* se han centrado en tres áreas de aplicación: páginas web para mejorar la accesibilidad, conversaciones en internet (chat en lengua de signos) y pantallas táctiles utilizadas generalmente como puntos de información. A diferencia de ViSiCAST, en el que la animación se obtenía a través de técnicas de captura de movimiento, en *eSIGN* la animación signada se genera de forma sintética [58], es decir a partir de un script de entrada en notación SiGML (Signing Gesture Markup Language) se obtiene una representación 3D [21].

Otro proyecto que hace uso de intérpretes virtuales de la lengua de signos está siendo desarrollado por la Universidad de Tunis. Jemni y Elghougl pretenden crear una herramienta que facilite la comunicación con personas sordas. Este sistema, llamado *WebSign* [24] es una aplicación Web basada en el uso de un avatar virtual. La interpretación se construye a través de un diccionario de palabras y signos y el tratamiento lingüístico se realiza gracias al Xeros Web-Service. Una aplicación particular de *WebSign* es *MMSSign* [25] cuyo uso se centra en los teléfonos móviles, donde los mensajes de texto son traducidos, a través de *WebSign*, a un vídeo en el que un avatar representa la información en lengua de signos.

Otro sistema de traducción de lengua oral a lengua de signos ha sido desarrollado en la *Universidad Politécnica de Madrid* [?] [48]. Se aplica en un contexto muy determinado, la creación o renovación del DNI (Documento Nacional de Identidad) y pasaporte. El sistema está formado por tres módulos: reconocimiento de voz, traslación de la lengua natural y representación de la lengua de signos Española (LSE) a través de un intérprete virtual. El módulo de representación utiliza dos avatares creados por *eSIGN*.

La Universidad de Purdue (Indiana) está desarrollando programas educativos 3D interactivos [3] siendo *MathSigner* y *SMILE* los dos principales proyectos en los que se encuentran trabajando actualmente. *SMILE* [1] [4], es un juego inmersivo que utiliza un mundo virtual 3D y una interfaz bilingüe: inglés y lengua de signos americana (ASL). En él, el estudiante puede manipular y construir objetos de la ciudad de *Smilleville*. *Mathsigner* [2] es una herramienta interactiva 3D que contiene un conjunto de

actividades orientadas al aprendizaje de ciertos conceptos matemáticos y sus signos correspondientes.

Otra de las compañías orientadas a la generación de avatares virtuales para la lengua de signos es IBM. Esta aplicación se llama *SiSi* (Say It Sign It) [61] [40]. El prototipo *SiSi* utiliza *ViaVoice*, un módulo de reconocimiento de voz convierte cada palabra hablada a texto. Este texto se pasa a través de un módulo de traducción para convertir el contenido a la lengua de signos Británica (BSL). *SiSi* etiqueta las frases con marcadores de gramática (grammar markers) y los reestructura en gramática BSL. Después, para animar el avatar, el sistema convierte los comandos en gestos en BSL obtenidos de un diccionario de signos.

Todos los sistemas descritos anteriormente utilizan métodos de síntesis de imagen interactiva, empleando técnicas de *Scanline* basadas en el uso directo de la GPU (unidad de procesamiento gráfico). Esto permite obtener una representación rápida, pero con una mala calidad de imagen y un cálculo pobre de los movimientos. Dentro de estos proyectos, tan sólo *eSIGN* y *Sign Smith*, ofrecen la posibilidad de crear nuevos signos y ser añadidos a los que el sistema tiene predefinidos. Sin embargo ambos hacen uso de la animación sintética al vuelo, lo que produce una apariencia menos natural que la conseguida con otras técnicas. Por otro lado, tan sólo *Sign Smith* permite configurar algunas características del vídeo final como el tamaño de la ventana o el formato de salida.

Este tipo de sistemas requieren dos características esenciales: la animación facial del avatar y el uso de técnicas de animación que permitan obtener movimientos cercanos a los de un intérprete real. Si no se cumplen estas características es posible que el usuario no entienda las palabras signadas por el avatar 3D. Además de éstas, se pueden distinguir otras dos características deseables: independencia del dispositivo de visualización con el fin de aumentar la accesibilidad a este tipo de información y posibilidad de elegir la calidad de la imagen resultado para adaptarse a diferentes medios de publicación (elección del nivel de realismo).

2.2. Computación afectiva

La expresión facial permite en la lengua de signos otorgar diferentes significados a una misma palabra y/o expresión y mantener la atención del receptor. Estos distintos significados están relacionados con el estado de ánimo que representa el emisor de la comunicación. Siguiendo esta línea y extrapolándola a las aplicaciones basadas en computador se encuentra la *computación afectiva* (Affective Computing) [53] que pretende, entre otros, simular emociones en aplicaciones basadas en ordenador. El principal objetivo de este enfoque es mejorar el bienestar de los usuarios al utilizar un determinado software y por tanto, conseguir una mejor comunicación humano-computador.

A pesar de que la computación afectiva es un área reciente, las investigaciones relacionadas con este tema crecen día a día. Algunas de las más novedosas son:

- Estudio de la computación afectiva para el tratamiento de pacientes autistas [43].
- La computación afectiva y el arte digital. Esta línea de investigación tiene como objetivo que el público sea un participante activo del proceso creativo [51].

- En otras áreas podemos encontrar la actualización de aplicaciones tradicionales con las ideas de la computación afectiva, como por ejemplo la computación afectiva en el contexto de las bibliotecas digitales [54].

Existen numerosas teorías sobre cómo se generan las emociones y su simulación por computador. Una de ellas es el modelo cognitivo “Ortony Clore Collins” (OCC), que según Rosalind Picard [42] se trata de una teoría muy útil para la síntesis de emociones cognitivas. Este modelo asume que las emociones aparecen como reacciones (positivas o negativas) a situaciones consistentes en agentes, eventos y objetos, definiéndose para ello 22 tipos de emociones (Figura 2.1). Por otro lado, el sistema de codificación de acciones faciales (FACS) desarrollado por Ekman y Friesen [15] fue uno de los primeros esfuerzos orientados a determinar las unidades de acción (AU), es decir el conjunto más pequeño de acciones faciales posibles y asociarlos a expresiones faciales.

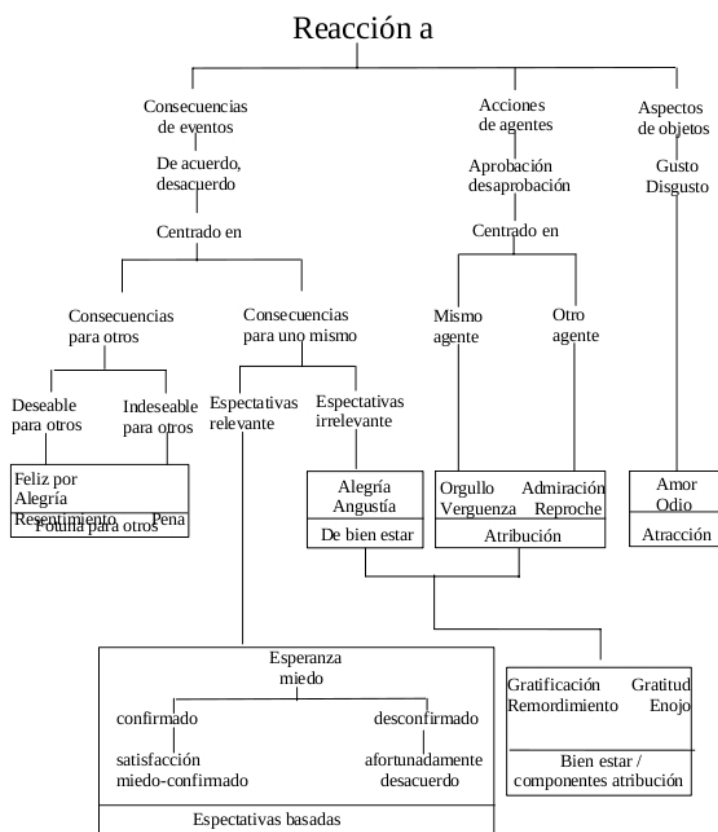


Figura 2.1: Estructura cognitiva de emociones del modelo OCC [38]

Por otro lado, una de las áreas de investigación más activas en el campo de la psicología del comportamiento consiste en establecer los parámetros a partir de los cuales se puedan definir todos los estados emocionales según los valores dados a dichos parámetros. Thayer [55] propuso dos parámetros de la energía y estrés que forman un estado de ánimo de dos dimensiones, mientras que Russell [47] define el lugar de los

estados emocionales comunes en un espacio circunplejo 2D utilizando los parámetros *excitación* y *valencia*.

2.2.1. Simulación de emociones en avatares

Existen diversos estudios que tratan de simular las expresiones faciales provocadas por distintas emociones. En los trabajos de Jean Claude Martin [33][11] se presenta un modelo computacional de expresiones faciales que aparecen como la combinación de distintas emociones. Dicho modelo no sólo computa expresiones faciales complejas sino que también distingue entre diferentes tipos de combinaciones de emociones para las que se definen un conjunto de reglas difusas basándose en la investigación previa del psicólogo Ekman [16] que consideraba un conjunto pequeño de seis emociones básicas (ira, asco, miedo, alegría, tristeza y sorpresa). El modelo por tanto se basa en el conjunto de reglas establecido por la combinación de estas seis expresiones e introduce la noción de similitud entre expresiones (empleando una medida difusa) teniendo en cuenta la cercanía entre una emoción dada y el conjunto de las básicas.

Sea $Exp(E_i)$ la expresión asociada a una emoción E_i y $Exp(N)$ la expresión neutral, el modelo parte de la suposición de que cualquier expresión facial se divide en n áreas (F_k) y cada F_k representa una parte única de la cara. Cada emoción E_i , $Exp(E_i)$ se compone de “ n ” áreas faciales diferentes siendo siete las áreas consideradas en este modelo: cejas, párpados superiores, párpados inferiores, mejillas, nariz, labio superior y labio inferior.

Por otro lado, si E_i y E_j son emociones que se mezclan y $Exp_{blend}(E_i, E_j)$ es la expresión mezclada resultante, esta mezcla puede ser tanto un enmascaramiento como una superposición. El enmascaramiento ocurre cuando una emoción no puede ser mostrada por alguna razón y se prefiere mostrar una expresión emocional distinta, por ejemplo cuando la emoción real es la de tristeza pero trata de esconderse tras una falsa apariencia de alegría. Por otro lado, la superposición ocurre cuando dos emociones diferentes se sienten y se muestran de forma simultánea (Figura 2.2). En este modelo, los autores consideraron que un área facial no puede mostrar diferentes expresiones al mismo tiempo, es decir cada $F^{(E_i, E_j)}$ siendo F_k una parte facial única (como los ojos o cejas) es igual a un área correspondiente de $Exp(E_i)$, $Exp(E_j)$ o $Exp(N)$ (expresión neutral).

Básicamente, el modelo trata de computar qué expresión básica es más similar a la expresión dada como entrada. Para ello, cada emoción se asocia a un conjunto de intervalos difusos en el que están definidas todas las expresiones posibles para esa emoción. Los pasos que se definen en el algoritmo son:

1. Para toda $Exp(E_j)$: se define su similitud con las seis expresiones básicas.
2. El valor más alto de similitud determinará la expresión básica más cercana.
3. Según el grado de similitud, la expresión final $Exp_{blend}(E_i, E_j)$ se elige en base a conjunto de reglas difusas.
4. La expresión final $Exp_{blend}(E_i, E_j)$ se obtiene como combinación de las áreas faciales de $Exp(E_i)$, $Exp(E_j)$ y/o $Exp(N)$

Debido a que los autores consideran dos tipos de combinaciones de emociones (enmascaramiento y superposición), se tratan también diferentes conjuntos de reglas difu-

sas para cada uno de ellos.



Figura 2.2: Imagen superior: desilusión enmascarada con alegría. Imagen inferior: superposición de tristeza y alegría. Basada en la Figura del trabajo [33]

Raouzaïou et al. [44] presentan un método que trata de enriquecer la interacción persona-computadora, a través del análisis y síntesis de expresiones faciales tanto primarias como intermedias. Para conseguir este objetivo, utilizan los Parámetros de Animación Facial (FAP), en el modelo de expresiones primarias y describen una técnica basada en reglas para el manejo intermedio. La relación existente entre los FAP y la activación de parámetros propuestos en los estudios de psicología clásica permite parametrizar el análisis de las expresiones faciales de forma compatible con el estándar MPEG-4.

El modelado de expresiones intermedias se logra como combinación, en el marco de un sistema basado en reglas, de un “parámetro de activación” con la expresión arquetipo de FAP. En la Figura 2.3 se muestra el esquema propuesto por los autores: el sistema de síntesis de las expresiones faciales funciona mediante la utilización de los valores estimados de FAP por un subsistema de análisis de imagen, o por el render de expresiones reconocidas mediante un conjunto de reglas difusas. En el primer caso, el movimiento de los puntos faciales relevantes son analizados y traducidos a variaciones de valor FAP, que a su vez se renderiza utilizando modelos sintéticos de cara (con el fin de reproducir la expresión dada). Los resultados del análisis deberían coincidir con la definición de las expresiones faciales. Después la expresión puede ser renderizada utilizando tablas predefinidas de FAP alterados. En este caso, las expresiones intermedias pueden ser renderizadas utilizando reglas de interpolación derivadas del conjunto de emociones.

Bevacqua et al. [9] se centran en el desarrollo de un agente conversacional capaz de interactuar de forma natural con los usuarios. Para ello desarrollaron un modelo que tiene en cuenta el comportamiento no verbal (emociones) asociado a las funciones comunicativas. El modelo no sólo añade expresividad a nivel de gestos faciales sino que

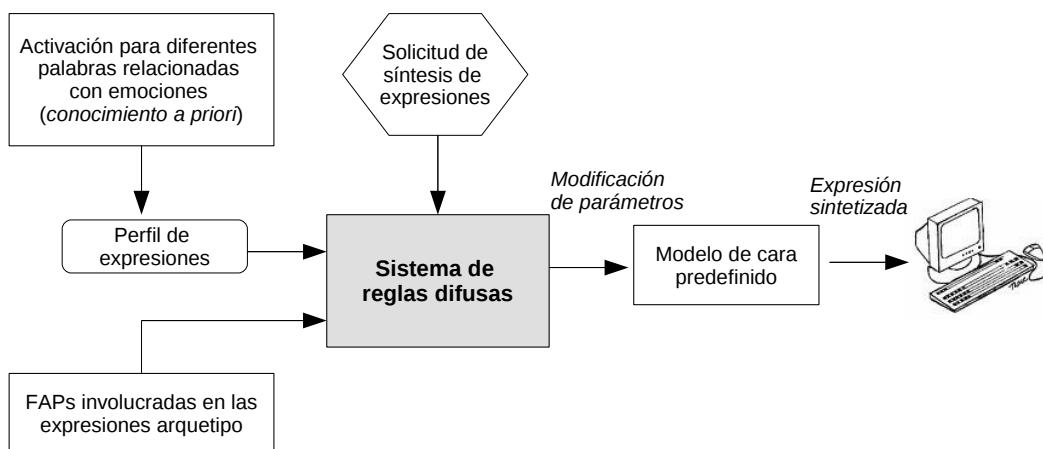


Figura 2.3: Esquema adaptado del propuesto por Raouzaïou et al. [44]

también tiene en cuenta la forma de los labios según las emociones.

En la simulación de emociones por parte del agente virtual, los autores también han tenido en cuenta la característica temporal de las expresiones. Un ejemplo de la representación utilizada se muestra en la Figura 2.4. En el eje de abscisas se muestra el tiempo y en el eje coordenadas se muestra la intensidad de la emoción, sabiendo que 0 significa que los músculos se encuentran en posición de reposo y 1 en posición final. Así, por ejemplo, las expresiones de tristeza generalmente desaparecen más despacio en el tiempo mientras que las de alegría aparecen muy rápidamente.

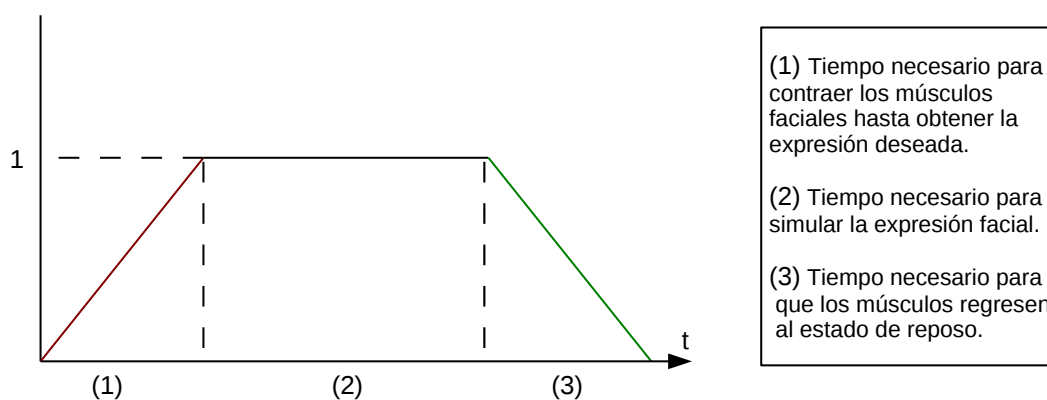


Figura 2.4: Curso temporal de una expresión facial según el trabajo de Bevacqua, Mancini y Pelachaud [9].

El sistema diseñado toma como entrada un texto etiquetado con información relacionada con funciones comunicativas. El lenguaje de etiquetas está basado en el denominado Affective Presentation Markup Language (APML) pero con algunas modificaciones, introduciendo un conjunto de cinco atributos que actúan tanto en el tiempo necesari-

rio para general las expresiones como en la intensidad relacionada con la misma. Para cada etiqueta APML se pueden especificar los atributos siguientes:

- *Delay*: especifica el porcentaje de retardo antes de que se produzca una expresión.
- *Duration*: determina la duración total de una expresión.
- *Onset*: especifica el valor de inicio medido en número de frames.
- *Offset*: valor que determina el momento en el que finaliza la expresión.
- *Intensity*: referido a un factor que multiplica la cantidad de movimiento por cada FAP involucrado en la expresión facial.

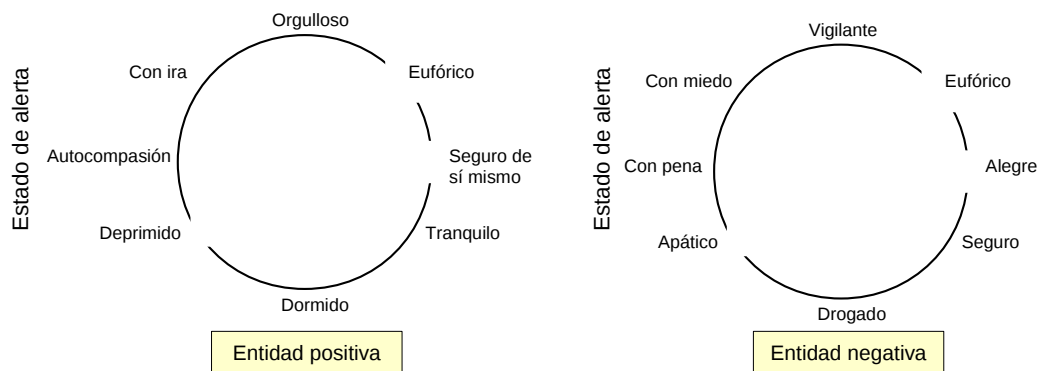


Figura 2.5: Espacio tridimensional de emociones según Araya et al. [6]

Araya, DiPaol y Parush [6] estudian el problema relacionado con la generación de expresiones faciales como mezcla de las emociones. La mayoría de los métodos utilizados emplean operaciones como la media para crear emociones resultantes de la combinación de otras dos. Aunque estos métodos pueden ser matemáticamente justificables, sin embargo no son necesariamente válidos desde el punto de vista perceptual. Por esta razón en el artículo se realiza una serie de experimentos cuyo objetivo es entender cómo las personas combinan las acciones faciales para expresar emociones mixtas y cómo la gente percibe un conjunto de acciones faciales en términos de emociones. A partir de los resultados obtenidos con los experimentos y utilizando un modelo de emociones tridimensional, los autores asocian acciones faciales con las dimensiones y regiones en el espacio de las emociones. De esta forma se crea una expresión facial basada en la localización de mezcla de emociones en el espacio tridimensional.

En la Figura 2.5 se muestra el espacio de emociones 3D con la *pose* utilizada como tercera dimensión. En la parte de la izquierda de la figura se muestra una organización positiva con emociones orientadas a uno mismo, por ejemplo el enfado, mientras que en la parte derecha los valores negativos se obtienen a partir de emociones no orientadas a uno mismo como puede ser el miedo. Este espacio tridimensional es una elección efectiva ya que separa las emociones universales con el menor número de parámetros que pueden ser visualizados, es decir relacionado con las acciones visuales. Las tres dimensiones del espacio de emociones son la excitación (el nivel de excitación), de valencia (el nivel de simpatía), y de la pose o gobierno (el nivel de auto-gobierno). En realidad, las

unidades de expresión son unidades de acción FAC (Facial Action Coding system) que están asociadas a las dimensiones de la emoción. En esta aproximación la asociación se ha realizado a través de un conjunto de reglas difusas.

Emoción	Estado de alerta	Valor	Entidad
Alegría	17.3	42.2	12.5
Tristeza	-17.2	-40.1	-52.4
Miedo	14.8	-44.4	-79.4
Ira	34.0	-35.6	20.0
Disgusto	60.0	-60.0	30.0
Sorpresa	60	0	0

Tabla 2.1: Valores dimensionales para las emociones estándar obtenidas del trabajo de Araya, DiPaol y Parush [6].

En la Tabla 2.1 se muestran las emociones estándar con el fin de localizarlas en el espacio 3D.

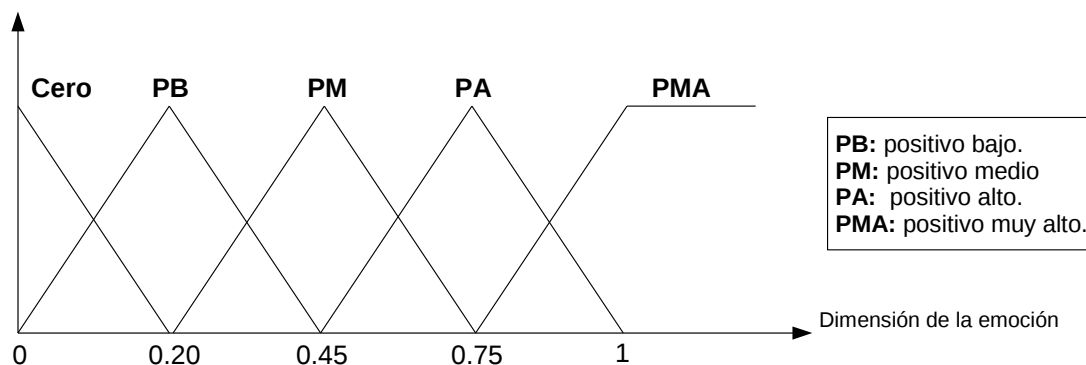


Figura 2.6: Conjuntos difusos definidos en el trabajo [6].

Debido a la incertidumbre de estas asociaciones, los autores proponen un sistema basado en lógica difusa. Para cada dimensión de la emoción se definen un conjunto de etiquetas difusas: cero, bajo, medio, alto y muy alto para los lados positivos y negativos. Los valores numéricos de la dimensión de la emoción (d_j) actuarán como entrada de las funciones de pertenencia asociadas a cada conjunto difuso representado mediante una etiqueta lingüística. Por ejemplo, el valor 30 tendrá un valor de pertenencia de 0,8 al conjunto positivo-bajo y 0,2 al conjunto positivo-medio y cero en otras.

En base a los experimentos realizados, cada etiqueta está asociada con un conjunto de acciones faciales con valores nominales. Por ejemplo, una excitación alta-positiva puede causar la apertura de la mandíbula en un 0,4. Los valores dimensionales de una emoción mixta se usarán para calcular los valores de pertenencia en cada etiqueta. Según las asociaciones observadas en los experimentos la regla puede tener como entrada una o dos etiquetas difusas:

IF Arousal is PositiveHigh THEN OpenJaw is 0,4 active.

IF Arousal is PositiveHigh AND Agency is PositiveMedium, THEN SqueezeEyebrow is 0,25 active.

Para cada regla, el valor total de pertenencia será el mínimo de todos los valores de pertenencia de entrada (mi). Este valor global, multiplicado por el valor nominal de salida ($anom$), determinará el nivel de activación de la acción facial de salida (a):

$$a = anom \times MIN(mi)$$

Bui et al. [10] proponen un sistema basado en reglas difusas para representar el estado emocional de un agente virtual a través de la contracción de los músculos. El sistema, que tiene en cuenta las seis emociones básicas (alegría, tristeza, miedo, sorpresa, ira y asco), determina cuándo la entrada en una expresión básica o la mezcla de dos de ellas y obtiene como salida un vector de contracción de músculos faciales (Figura 2.7). La intensidad de las emociones se describe a través de conjuntos difusos cuyas etiquetas son “muy bajo, bajo, medio, alto y muy alto”. De forma similar, el nivel de contracción de cada músculo se describe por medio de conjuntos difusos: muy pequeño, pequeño, medio, grande y muy grande

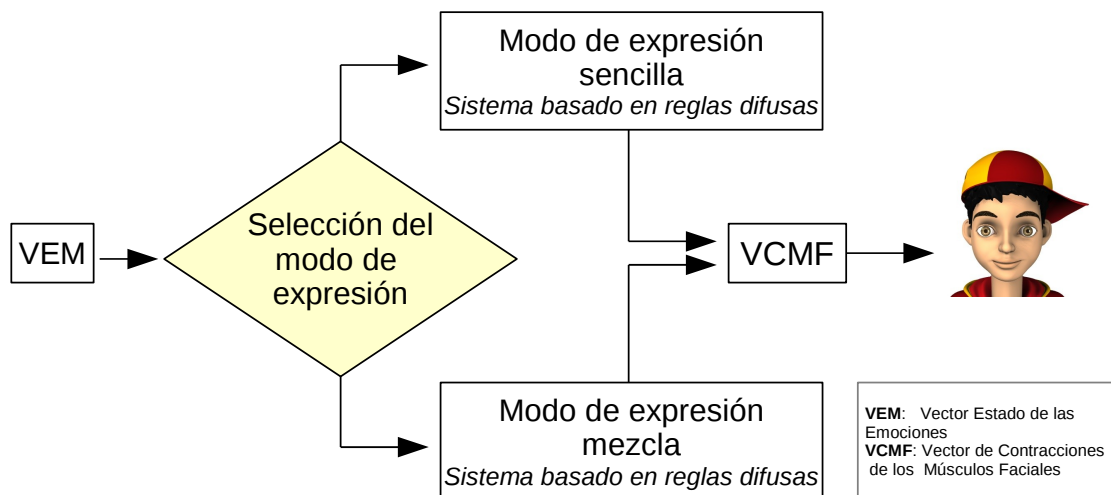


Figura 2.7: Esquema del sistema propuesto por [10].

La mayoría de los trabajos descritos anteriormente están orientados en la creación de avatares conversacionales con el fin de mejorar la interacción persona-ordenador. Cada uno de ellos selecciona el modelo psicológico de emociones que más se ajusta a sus necesidades, aunque uno de los más extendidos es el de seis emociones básicas. Este tipo de trabajos no suele estar orientado a personas con deficiencia auditiva, sino que la mayoría de las investigaciones que fusionan técnicas de lógica difusa con sistemas de apoyo a personas sordas suelen estar relacionados con tecnología de reconocimiento de gestos [28] [5] [60] [18] [12] [26] [52]. Sin embargo, como se comentó en la introducción, resulta esencial disponer de generación de emociones para intérpretes virtuales en la lengua de signos ya que al no poder apoyarse en el audio, no es posible sintetizarlas de ningún otro modo. Además las expresiones faciales son imprescindibles cuando se signa ya que, en caso contrario la comunicación puede llegar a ser confusa e incompleta. Un ejemplo se puede encontrar en los subtítulos de películas o series utilizados para

personas sordas en los que se muestran emoticonos junto con algunas frases con el fin de que se pueda entender el contexto de la sentencia (alegría, tristeza, enfado, etc.) y no exista confusión en su significado.

2.3. Adaptabilidad de avatares virtuales

La adaptación de los sistemas software al usuario final permite aumentar la usabilidad de los mismos [8]. Shackel [50] propone que la usabilidad puede definirse en base a cuatro criterios operacionales: eficacia, capacidad de aprendizaje, flexibilidad y actitud. La eficacia está relacionada con la realización de un conjunto de tareas por un porcentaje de usuarios dentro del entorno de la aplicación. La capacidad de aprendizaje está definida en términos del tiempo de entrenamiento que necesita el usuario, dentro de un determinado nivel, para llevar a cabo una tarea. La flexibilidad se refiere a la diversidad de tareas y entornos permitidos por el diseño de la aplicación. Por último, los niveles de aceptación del usuario vienen dados por la actitud.

Langley [30] define interfaz de usuario adaptativa como “*artefacto software que mejora su capacidad de interactuar con un usuario mediante la construcción de un modelo de usuario basado en la experiencia parcial con ese usuario*”. Un sistema adaptativo puede definirse como [23] “*un sistema interactivo que adapta su comportamiento a usuarios individuales en los procesos básicos de adquisición de modelo de usuario y aplicación que envuelven alguna forma de aprendizaje, inferencia o toma de decisiones*”.

En los sistemas adaptativos, el modelo de usuario supone una característica esencial ya que representa información de cada uno de los usuarios. Por esta razón, Brusilovsky, P. y Millán, E. [39] presentan una revisión de los modelos de usuario aplicados en sistemas web adaptativos. En el artículo describen cómo el uso de un método formal para el modelo del estudiante proporciona gran robustez, siendo las redes bayesianas el paradigma más utilizado.

Teniendo en cuenta estas pautas el sistema adaptativo debe configurarse según las necesidades y preferencias de cada usuario (o grupos de usuarios) en particular. En el caso concreto de la enseñanza y el aprendizaje podríamos definir la adaptación como una tutorización individual a cada estudiante. Para llevar a cabo de forma correcta dicha tutorización es necesario tener en cuenta un gran número de variables que pueden afectar al proceso de aprendizaje como las habilidades particulares o la capacidad de comprensión. Los Sistemas Tutores Inteligentes (STI) [37] [56] se basan en algún tipo de modelado de estudiante con el fin de orientar el comportamiento del tutor. De esta forma pretenden darle al alumno un aprendizaje individualizado como si fuera un tutor real.

Al contrario de los sistemas tradicionales de enseñanza asistida por ordenador, en los que la enseñanza se realizaba de manera estática, los STI [36] pueden guiar a los alumnos en el proceso de aprendizaje exploratorios así como adaptar la forma de enseñanza según las necesidades y características particulares de cada alumno. El modelo STI, está formado principalmente por cuatro componentes [34]: modelo de dominio, modelo de estudiante, modelo de enseñanza y entorno de aprendizaje.

Kathleen et al. [35] describen un prototipo de herramienta cuya finalidad es la de mejorar la alfabetización de las personas sordas en los institutos y colegios. El sistema pretende analizar los errores gramaticales de un fragmento de texto escrito por un

alumno sordo y que estos participen en un diálogo autorizado que le permita generar las correcciones necesarias.

Otra aplicación desarrollada es DeSIGN [57], se trata de un software educacional orientado a aquellos estudiantes que se comunican en lengua de signos americana (ASL). El objetivo de la aplicación es reforzar el vocabulario y los signos en ASL a través de un conjunto de lecciones y test. El sistema además adapta los test a la evolución del estudiante teniendo en cuenta sus conocimientos.

Casi todos los esfuerzos en los que se relacionan sistemas adaptativos orientados a personas sordas [32] [14] [13] se centran en un entorno muy concreto y con vocabulario en lengua de signos predefinido. Es por eso que el esquema que proponemos supone un enfoque diferente y novedoso, ya que permite integrar un sistema de traducción de lengua escrita con un sistema que se adapte a las necesidades particulares de cada usuario y del contexto aplicación. Nuestra aproximación no sólo está pensada para aquellos cuya lengua nativa es la lengua de signos, sino que el esquema propuesto se adapta al grado de discapacidad auditiva y al grado de lectura del usuario final. Además se tienen en cuenta otros aspectos menos relacionados con las capacidades y más orientados a los gustos personales, como puede ser la apariencia del avatar.

3

Arquitectura propuesta

3.1. Introducción

Uno de los problemas que se encuentran a la hora de definir el perfil del usuario y el contexto de aplicación de un mensaje es que este tipo de información suele ser imprecisa e incompleta. Por esta razón se hace necesaria la utilización de técnicas que trabajen correctamente con incertidumbre. La investigación que se ha llevado a cabo se centra en la generación automática de la lengua de signos de manera que se adapte a las necesidades del usuario y del contexto de aplicación. Para ello se han tenido en cuenta una serie de requisitos relacionados con la calidad en la generación y adaptabilidad de la síntesis en tres niveles:

- **Adaptación al medio de publicación.** Para llevar a cabo este tipo de adaptación es necesario que la etapa de síntesis de imagen esté desacoplada de la parte de descripción de movimiento, permitiendo generar vídeos para diferentes dispositivos y con diferentes calidades de imagen, es decir, con diferentes métodos de render y resoluciones (Tabla 3.1).
- **Adaptación al perfil (rol) del usuario.** Esto requiere que diferentes personajes virtuales (avatares) realicen los mismos movimientos. La descripción del movimiento, por tanto, debe estar almacenada en una estructura que sea independiente tanto del personaje como del mecanismo de generación. También es necesario adaptar la velocidad de signado, el tamaño del personaje y otras características que se expondrán más adelante.
- **Adaptación al entorno y al tipo de mensaje.** El contexto en el que se mostrará al personaje virtual viene determinado por el tipo de mensaje que puede ser generado (formales, tristes, alegres, urgentes, etc). La adaptación relacionada con el contexto y el mensaje propiamente dicho se obtiene gracias a la capacidad del personaje virtual para expresar diferentes tipos de emociones y sentimientos. De este modo

Parámetro	Valores	Descripción
Formato de la imagen	PNG, BMP, JPEG90	Formato de las imágenes renderizadas.
Formato del vídeo	AVIRAW, AVIJPEG, AVI-CODEC, QUICKTIME, TARGA, FLASH	Formato de los vídeos generados.
Frames por segundo	Valor por defecto 24 fps	Número de frames que contiene un segundo del vídeo generado.
Calidad de la imagen	[1-100]	Calidad de la imagen renderizada.
Tamaño de la ventana		Resolución del vídeo/imagen de salida.
Partes X	[1-512]	Número de partes (X) en las que se realiza el render.
Partes Y	[1-512]	Número de partes (Y) en las que se realiza el render
Antialiasing	Desactivado/activado ([5,16])	Valores de sobremuestreo utilizados para evitar el efecto sierra
Motion blur	[0,5]	Efecto desenfocado de movimiento
Sombras	Desactivado/activado	Cálculo de sombras.
Métodos de render	Path tracing, ray tracing, scanline, toon shading	Diferentes métodos para realizar el renderizado.
Capas	Lista de enteros	Capas de la escena que serán renderizadas (capas activas).

Tabla 3.1: Parámetros de render.

resulta imprescindible contar con mecanismos para la generación de expresiones faciales de forma independiente a los movimientos manuales y corporales propios de la lengua de signos.

3.2. Componentes de la arquitectura

La arquitectura del sistema se divide en cinco módulos principales (ver Figura 3.1): *editor de signos*, *procesamiento del lenguaje natural*, *representación 3D*, *adaptación al usuario* y *adaptación al contexto*. Cada uno de ellos es independiente del resto, de tal forma que aumenta la flexibilidad y reutilización del sistema en entornos diferentes al de la lengua de signos.

El *editor de signos* permite la creación de nuevas palabras o signos a partir de la manipulación manual de los huesos que conforman el esqueleto del avatar. Una vez realizado el posicionamiento de los huesos, se extrae de forma automática la localización de los mismos en el espacio 3D para poder ser utilizados posteriormente (representación 3D). El módulo de *procesamiento del lenguaje natural* emplea una aproximación basada en los métodos de traducción estadística y de transferencia. Dicho módulo permite convertir el texto escrito dado como entrada en un conjunto de tokens que equivalen a signos en la LSE. Estos tokens son enviados al módulo de *representación 3D* que se encarga de representar en el espacio 3D la localización y rotación de los huesos que intervienen en el movimiento completo de la animación con el fin de obtener la frase o expresión deseada. Gracias a la alta configurabilidad de los parámetros del render, el vídeo de salida puede adaptarse a las necesidades del dispositivo de visualización que será utilizado. Por otro lado, el módulo de *adaptación al usuario* permite configu-

rar diferentes elementos del vídeo y del avatar que muestra de forma que se ajusta a las necesidades y preferencias particulares de cada usuario. Por último, el módulo de *adaptación al contexto* simula las expresiones faciales o emociones que mejor se adaptan al contexto donde se emite el mensaje. A continuación se describirán cada uno de estos componentes en detalle.

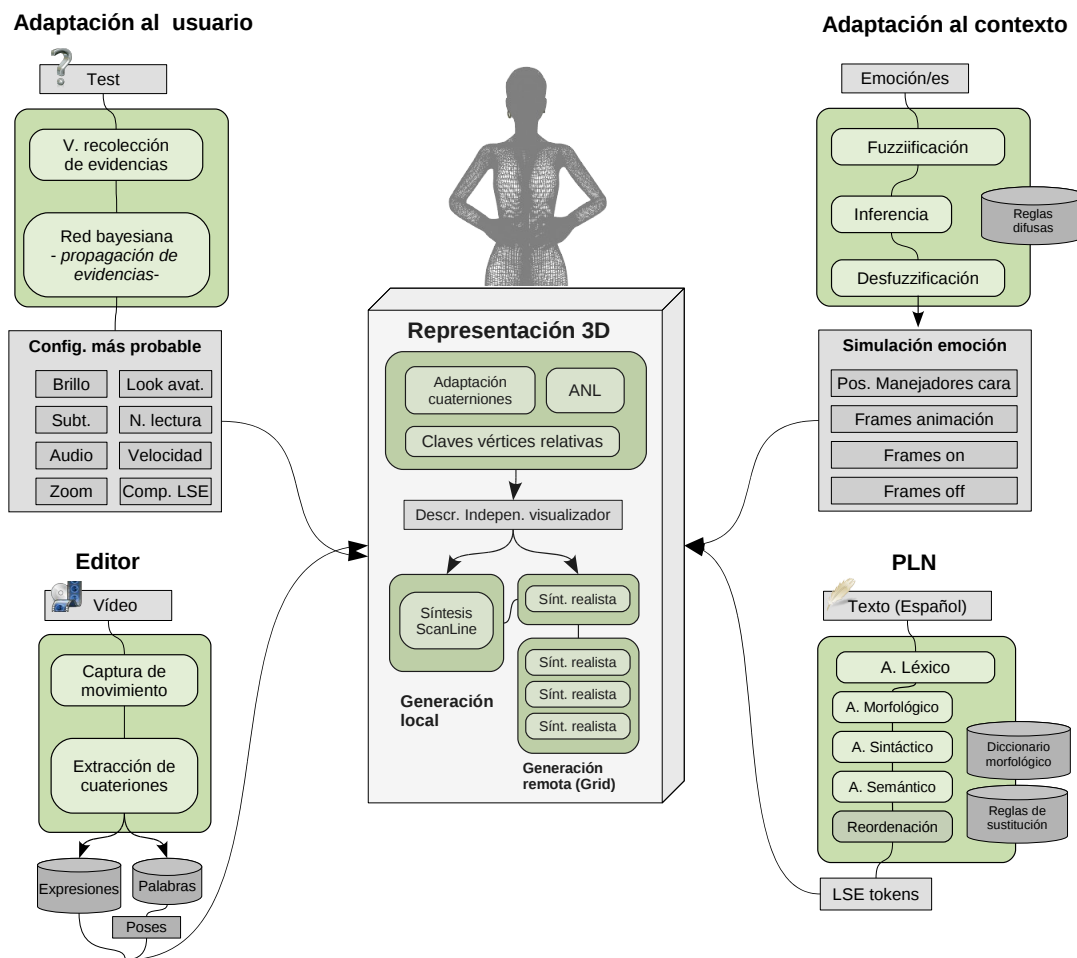


Figura 3.1: Esquema general de la propuesta.

3.2.1. Procesamiento del lenguaje natural

La traducción a lengua de signos presenta una serie de características particulares que hacen el procesamiento del lenguaje especialmente complejo. Además, las fuertes características locales de cada región requieren una aproximación flexible que permita una fácil ampliación y configuración de casos específicos. Un ejemplo de casos difíciles es la traducción de determinantes demostrativos acompañados de sustantivos que reordenan los sintagmas de forma totalmente diferentes si el nombre es alienable o no lo es.

A grandes rasgos, se pueden distinguir dos técnicas de traducción automática [65], los métodos de traducción estadística y la de transferencia. Los métodos de traducción estadística emplean métodos probabilísticos y de la teoría de la información a partir del análisis de corpus de textos bilingües. Este corpus bilingüe asigna una probabilidad de traducción entre cada par de frases. En el caso del sistema que se plantea, estos pares de frases están formados por documentos en español y su correspondencia a lengua de signos. Los métodos de traducción por transferencia por su parte definen un conjunto de reglas estructurales, un par de diccionarios monolingües para cada lengua, y un diccionario bilingüe. Con estos elementos se identifica el tipo de palabra en la frase y se realiza la reordenación y la traducción.

Los sistemas de traducción estadística funcionan muy bien en dominios acotados en los que se dispone de un amplio conjunto de frases de entrenamiento. El principal inconveniente radica en la necesidad de un conjunto amplio de frases de entrenamiento. Por su parte los sistemas de traducción por transferencia ofrecen peores resultados, pero son fácilmente generalizables (únicamente necesitan la definición de los diccionarios y las reglas de reordenación).

El módulo de procesamiento del lenguaje natural definido emplea una aproximación mixta basada en las dos técnicas; por un lado permite el uso de traducción estadística en ciertos dominios de aplicación donde el sistema esté entrenado empleando *Moses* [29], y el uso de un traductor de transferencia para frases que se traduzcan con una probabilidad muy baja empleando la aproximación estadística. La traducción de transferencia se basa en dos etapas, construidas en módulos empleando la plataforma de traducción libre *Apertium* [17]. En el módulo de procesamiento de Chunks se identifican los sintagmas y se produce la traducción léxica empleando el diccionario bilingüe. Además, se procede además a realizar una primera etapa de reordenamiento interno a nivel de chunk, obteniendo sintagmas. En el módulo de procesamiento *Interchunk* se reordenan los sintagmas construidos en el módulo anterior.

Finalmente en la fase de transferencia se realizan los cambios estructurales, detectando patrones en la frase de entrada y realizando acciones sobre ese patrón detectado. Este módulo de procesamiento del lenguaje natural es fácilmente ampliable para detectar patrones específicos y configurar nuevas reglas de acción.

3.2.2. Representación 3D

En la animación del avatar se han utilizado técnicas de animación no lineal (NLA) con el fin de obtener acciones complejas fluidas como resultado de la mezcla de acciones sencillas individuales. Utilizando esta metodología se han definido, en primer lugar, las acciones necesarias asociadas a cada parte del cuerpo y de la cara. Las acciones, asociadas al esqueleto del avatar, son creadas en diferentes capas de la escena a renderizar y posteriormente, todas las capas activas se mezclan para obtener la animación global. Se ha prestado especial interés en la composición de diferentes signos, de manera que se obtengan transiciones fluidas entre la representación de dos signos consecutivos.

Para facilitar la tarea del animador y obtener resultados realistas se ha utilizado técnicas de cinemática inversa e interpolación de rotaciones basada en cuaterniones [?] (extensión de los números reales). La cinemática se basa en el estudio de los movimientos independientemente de las fuerzas que los producen. Existen dos tipos: cinemática directa y cinemática inversa.

El método de cinemática directa tiene como objetivo representar la localización de un objeto en el espacio respecto a un sistema de referencia determinado. La posición final del modelo se determina especificando cada uno de los ángulos de sus articulaciones, es decir, hay que actuar sobre cada punto y producir un movimiento sobre su eje. La cadena cinemática se considera abierta y formada por eslabones unidos por articulaciones en la que el movimiento se transmite de padres a hijos.

El segundo tipo de cinemática es la cinemática inversa [31]. Este método permite reducir efectos no deseados en las animaciones. Antes de comenzar con la animación es necesario definir la estructura del objeto de forma jerárquica, típicamente esta estructura se puede ver como un árbol donde se comienza por el nodo principal (raíz) y se continúa por los nodos que dependen de él (hojas). Al revés que en el caso anterior, lo que se especifica es el punto final al que se desea llegar y el programa será el encargado de generar automáticamente los ángulos de las articulaciones implicadas en el movimiento.

Existen tres elementos esenciales para comprender el mecanismo de cinemática inversa: articulaciones, grado de libertad y nivel de rotación. Las articulaciones poseen características físicas que permiten describir el comportamiento cuando se efectúa el movimiento, el grado de libertad se refiere al número de direcciones en las que se puede mover una articulación y el nivel de rotación se refiere a los grados que una determinada articulación puede rotar.

El principal inconveniente que presenta la cinemática inversa, es que el movimiento es más complejo, ya que existen diferentes formas de rotar los elementos entre sí para obtener el movimiento final deseado. Un ejemplo claro es el caso del codo, que puede girar en un único sentido. Por esta razón en la arquitectura propuesta se emplean restricciones de rotación, de forma que se pueda identificar los movimientos prohibidos. En general, existen dos formas de abordar el problema de la cinemática inversa: analíticamente y mediante métodos iterativos. El principal inconveniente de los métodos analíticos reside en la complejidad computacional, pudiendo llegar a ser inabordable en modelos complejos. Para resolver este problema computacional, en el sistema propuesto, se usa el algoritmo CCD (Cyclic Coordinate Descent). Este algoritmo [59] se caracteriza por ser iterativo y ofrece la ventaja de minimizar el error ajustando, en cada instante, un ángulo determinado de la cadena jerárquica. El algoritmo comienza por el último elemento de la cadena y continúa hacia atrás para ajustar cada articulación. El ciclo se repetirá hasta que el efector esté a una distancia cercana de la posición final o bien se alcance un número determinado de iteraciones.

En la animación de personajes virtuales es muy importante que la deformación de la piel y los músculos asociados al esqueleto se haga de forma correcta. Una de las técnicas de deformación de la piel, *linear blend skinning*, puede producir efectos no deseados. Aunque existen diferentes aproximaciones que tratan de evitar este efecto no deseado, la mayoría de ellas no ofrecen la simplicidad y eficiencia de la técnica de mezclado lineal. Una solución propuesta es la basada en cuaterniones dobles [27]. Aunque los cuaterniones se han utilizado clásicamente para la representación de rotaciones, también pueden ser usados para el mezclado de rotaciones que incluye todas las transformaciones de los rígidos. Los algoritmos basados en cuaterniones dobles empleados en el sistema desarrollado, son computacionalmente más eficientes que los obtenidos por soluciones previas y ofrecen mejores propiedades como la velocidad constante. Por esta razón, en la animación de los avatares utilizados en el sistema se ha hecho uso de la interpolación basada en cuaterniones.

Para que los movimientos realizados por un personaje virtual puedan ser utilizados por otro diferente es necesario almacenar la descripción referente a estos movimientos en una estructura independiente al avatar. En el sistema propuesto, la posición, localización y rotación de los huesos involucrados en una determinada acción (entre otros) es almacenada en un fichero con formato XML. Estos archivos son independientes del avatar, del motor de render y de la interfaz gráfica (ver Figura 3.2)



Figura 3.2: Diferentes personajes virtuales y descripción del mismo movimiento.

3.2.3. Renderizado

Para hacer posible la adaptación al medio de publicación se han definido un conjunto de parámetros de configuración del render que se pueden observar en la Figura 3.1. Modificando estos parámetros se obtienen diferentes configuraciones según las necesidades particulares del dispositivo donde será visualizado (PDA, teléfono móvil, televisor, etc).

La etapa de renderizado es la más lenta del sistema sobre todo cuando se realiza de forma centralizada, en este caso un sólo ordenador es el encargado de renderizar la escena completa. Por esta razón, y con el fin de reducir el tiempo empleado en el proceso de renderizado, la arquitectura propuesta permite el renderizado distribuido a través del sistema Yafrid-NG [22]. Dicho sistema hace uso de un conjunto de computadores heterogéneos, distribuidos en internet, que ceden parte de sus ciclos de CPU para renderizar escenas tridimensionales.

Algunas de las ventajas que presenta la adaptación al medio de publicación, gracias a la parametrización del render, son las siguientes:

- Incrustación en páginas web: la animación obtenida puede ser exportada a formato flash y ser incluida en páginas web con el fin de facilitar la accesibilidad a personas con discapacidad auditiva.
- Uso en teléfonos móviles. El video en Lengua de Signos se puede adaptar al tamaño y capacidades 3D del teléfono móvil. Gracias a la técnica de frames clave, se pueden extraer aquellos más relevantes en la animación y generar así un gif animado, adecuado para la práctica de cualquier dispositivo móvil.
- Transmisión del vídeo a través del TDT. El estándar MHP [62] permite la recepción y ejecución en una televisión de aplicaciones basadas en Java. Gracias a este estándar los vídeos en Lengua de Signos Española generados pueden ser comprimidos a un formato que facilite su transmisión en el canal de datos del TDT y ser

utilizado a modo de “subtítulos” en películas, series y otros programas emitidos en televisión.

3.2.4. Adaptación al rol del usuario

Como se ha comentado anteriormente este tipo de adaptación se realiza en dos niveles diferentes: almacenamiento de los movimientos del esqueleto en una estructura independiente del personaje virtual (XML) y adaptación a las necesidades y preferencias del usuario.

La adaptación en tiempo de ejecución a las necesidades individuales del usuario se ha desarrollado utilizando redes bayesianas. De esta forma es posible identificar las características del avatar que sean más adecuadas para cada usuario, teniendo en cuenta no sólo sus necesidades sino también las preferencias particulares.

Una red bayesiana (RB) es un grafo acíclico dirigido cuyos nodos representan variables aleatorias y cuyos enlaces representan relaciones entre ellos [19]. Este tipo de red describe una distribución de probabilidad, sobre sus variables, que satisface la propiedad de *d-separación* [41] lo que implica que la distribución de probabilidad conjunta de una RB ($P(X_1, \dots, X_n)$) puede ser factorizada como el producto de la probabilidad de cada nodo condicionado a sus padres ($P(X_i | pa(X_i))$):

$$P(X_1, \dots, X_n) = \prod P(X_i | pa(X_i))$$

El razonamiento probabilístico, que se basa en la aplicación del teorema de Bayes, puede realizarse de dos formas:

- **Propagación de la evidencia.** Consiste en calcular la probabilidad a posteriori de variables no observadas dada la evidencia disponible.
- **Abducción.** En el caso de abducción total, consiste en buscar la configuración o configuraciones para todas las variables no observadas que tienen máxima probabilidad. En la abducción parcial se fija previamente un subconjunto de las variables no observadas y por tanto la configuración de la probabilidad máxima contiene sólo las variables de este conjunto.

La construcción de la RB se puede realizar básicamente de dos maneras: automática y manual. El proceso automático consiste en extraer la información de una base de datos y aplicar algoritmos con el fin de obtener tanto la estructura de la red como las probabilidades condicionales. En cuanto al proceso manual, puede dividirse en dos etapas, construcción de la estructura de la red (variables y enlaces causales) con ayuda del experto en la materia, e introducción de las probabilidades condicionales correspondientes.

La red bayesiana propuesta con el fin de dar soporte personalizado a cada usuario está basada en el trabajo de Millán et al. [36]. En este trabajo presentan un enfoque para el diagnóstico del modelado de estudiante basado en RB y *Computer Adaptive Test* (CAT) definiendo dos tipos de variables: variables para medir el grado de conocimiento del alumno y variables para recolectar evidencia. En nuestro caso particular se definen variables para medir las capacidades y gustos del usuario así como variables para recolectar las evidencias. Las primeras se definen en diferentes niveles de granularidad.

Variables para la medición de las capacidades y gustos del usuario

Las variables definidas están relacionadas por un lado con las capacidades auditivas, visuales y grado de lectura y por otro con gustos particulares como es la apariencia del personaje virtual:

- *Discapacidad visual.* El usuario puede sufrir algún grado de discapacidad visual, por lo que se le han asignado cuatro posibles estados: ninguno, bajo, medio y alto.
- *Grado de sordera.* En las personas con deficiencia auditiva se pueden considerar cuatro tipos diferentes de grado de sordera: deficiencia auditiva ligera (pérdida auditiva entre 20db y 40db), deficiencia auditiva media (pérdida entre 40db y 70db), deficiencia auditiva severa (entre 70 y 90 db) y deficiencia auditiva profunda (pérdida auditiva mayor que 90), en este último caso el individuo no puede adquirir el lenguaje oral. También se ha tenido en cuenta la posibilidad de que el usuario no sufra ningún tipo de sordera, el sistema por tanto estaría orientado al aprendizaje de la lengua de signos por parte de personas oyentes (intérpretes).
- *Velocidad del signado.* Una velocidad de signado no apropiada puede impedir que la información se transmita de forma adecuada.
- *Nivel de lectura.* Como se ha comentado anteriormente, el aprendizaje de la lectura está muy ligado con la audición. Por esta razón el sistema distingue entre distintos grados de lectura (bajo, medio y alto).
- *Edad.* La edad del usuario final afecta a características, capacidades y gustos del mismo. El valor de esta variable será introducida directamente por el usuario.
- *Zoom.* Con esta variable se puede aumentar y disminuir el tamaño del personaje virtual que interpreta la lengua de signos.
- *Brillo.* Si el avatar no se ve de forma adecuada puede dificultar el grado de comprensión por parte del usuario.
- *Audio.* En nuestra aproximación se ha tenido en cuenta el audio, ya que no sólo está dirigido a personas con sordera profunda. El audio supone una gran ayuda en el caso de que el usuario tenga capacidad de escuchar la información que se da representada en lengua de signos.
- *Apariencia del avatar.* Esta variable no se ha utilizado para definir las capacidades del usuario, sino los gustos. Una interfaz atractiva permite que el usuario se sienta más cómodo con la aplicación. Por esta razón se dan definido distintos tipos de personajes virtuales: infantil, juvenil y adulto.
- *Complejidad de la LSE (complejidad gramatical y complejidad del vocabulario).* Si alguna de estas dos variables (o ambas) resulta demasiado compleja para el usuario, el aprendizaje sería demasiado complejo y si por el contrario fuera (o fueran) demasiado sencillas, el usuario desaprovecharía su capacidad de aprendizaje.

Variables de recolección de evidencias

Con estas variables se obtendrá información sobre las características del usuario. Las evidencias recogidas son las respuestas que el usuario da a un conjunto de preguntas que el sistema realiza, una vez conocidas las preferencias y gustos es posible adaptar la aplicación al perfil del usuario.

Pregunta	Posibles respuestas
P1 ¿Cómo me ves?	Claro, bien, oscuro.
P2 ¿Cómo ves mi tamaño?	Pequeño, normal, grande.
P3 ¿Has oído lo que te he dicho?	Nada, poco, casi todo, todo.
P4 ¿Cómo crees que signo?	Lento, normal, rápido.
P5 Frase escrita: <i>El niño pequeño juega en su casa.</i> ¿Has entendido la frase?	No, si .
P6 Frase escrita: <i>El niño pequeño juega con una pelota en el patio trasero de su casa.</i> ¿Has entendido la frase?	No, si .
P7 Frase signada: <i>El último verano fui al campo de mis amigos.</i> ¿Has entendido la frase?	No, si.
P8 Frase signada: <i>El último verano fui al campo con mis amigos de Ciudad Real</i> ¿Has entendido la frase?	No, si.
P9 Palabras signadas: <i>inflexible, aturdido, juez, atónito, estatura, resto</i> ¿Entiendes lo que significan las palabras?	No, si.

Tabla 3.2: Variables definidas para la recolección de evidencias.

La interfaz de usuario muestra a un personaje virtual estándar presentándose y describiendo brevemente la finalidad de la aplicación. La información que da el personaje en su presentación se realiza tanto en lengua de signos como con voz. Una vez que la presentación ha concluido se realizan al usuario nueve preguntas con el fin de adaptar la aplicación conforme a sus necesidades y gustos (ver Tabla 3.2)

Cálculo de probabilidades a priori

Una vez definidos los nodos y enlaces de la red, es necesario establecer las probabilidades a priori. Las probabilidades a priori se han estimado subjetivamente según la experiencia de personas sordas y personas oyentes expertas en la lengua de signos, más concretamente en la lengua de signos española (LSE).

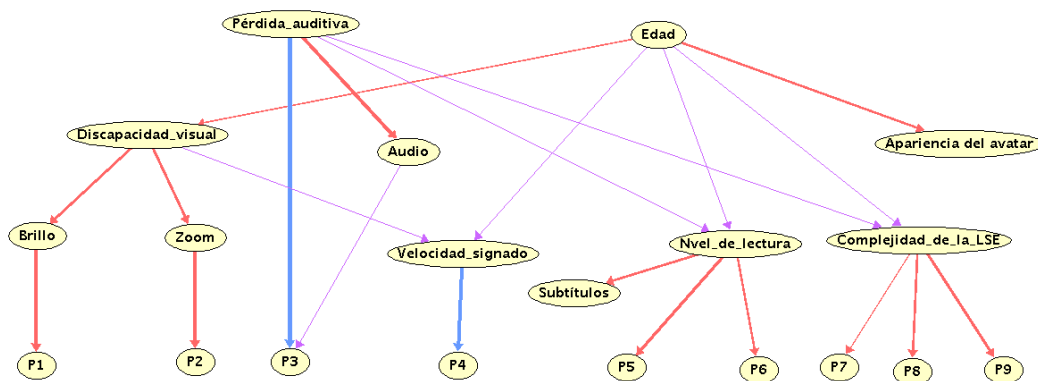


Figura 3.3: Red bayesiana obtenida

El grosor y color de los arcos de la red mostrados en la Figura 3.3 describen la influencia realizada sobre cada nodo. En la definición de las probabilidades asociadas con el *nivel de lectura* es necesario tener en cuenta la relación existente entre comprensión

de la lectura y audición. Por tanto es más probable que personas con un grado de sordera profundo presenten más problemas en la lectura que otras que presenten un grado de deficiencia auditiva leve. Lo mismo ocurre con la edad, los niños suelen tardar más tiempo en adquirir un grado elevado de lectura y por otro lado, las personas mayores probablemente no habrán sido instruidas de forma adecuada en la lectura.

Debido a que el esquema propuesto no está restringido a una edad determinada del usuario, éste puede ser tanto un niño como un adolescente o joven y un adulto. Para que el usuario se sienta cómodo con la herramienta, el avatar que interpreta la lengua de signos varía según la edad, esto significa que es más probable que un niño prefiera un personaje infantil, un adolescente uno juvenil y el resto un avatar con apariencia de adulto.

Ejemplo de ejecución

A continuación se describen algunos experimentos realizados con el esquema anteriormente propuesto. Los resultados obtenidos han sido satisfactorios, lo que indica que el sistema es capaz de adaptarse de una forma adecuada las necesidades particulares de cada usuario.

En primer lugar, mediante un proceso de abducción, se obtiene el perfil del usuario en el caso de no tener ninguna evidencia sobre él, es decir se obtiene la configuración de las variables que tienen la máxima probabilidad conjunta. En este caso, las variables que han sido estudiadas son: *audio*, *contraste*, *zoom*, *velocidad de signado*, *nivel de lectura*, *complejidad de LSE*, *apariciencia del avatar* y *subtítulos*. En la Figura 3.3 se puede observar que la configuración obtenida está sujeta al hecho de que la edad a priori más probable del usuario se encuentra en el rango entre 35 y 64 años, y con un grado de sordera profunda. Si el usuario no percibe sonidos no tiene sentido el uso del audio (*audio = ausente*), por otro lado el nivel de lectura más probable es medio ya que las personas con discapacidad auditiva profunda suelen tener dificultades en el aprendizaje de la lectura. En cuanto a la apariencia del avatar correspondería a un personaje adulto. Los valores asociados con el contraste y el zoom se mantienen tal y como están. Dado que el usuario por defecto es adulto y con un grado de sordera profunda, probablemente domine la lengua de signos, por esta razón la variable *complejidad LSE* toma el valor *alto*.

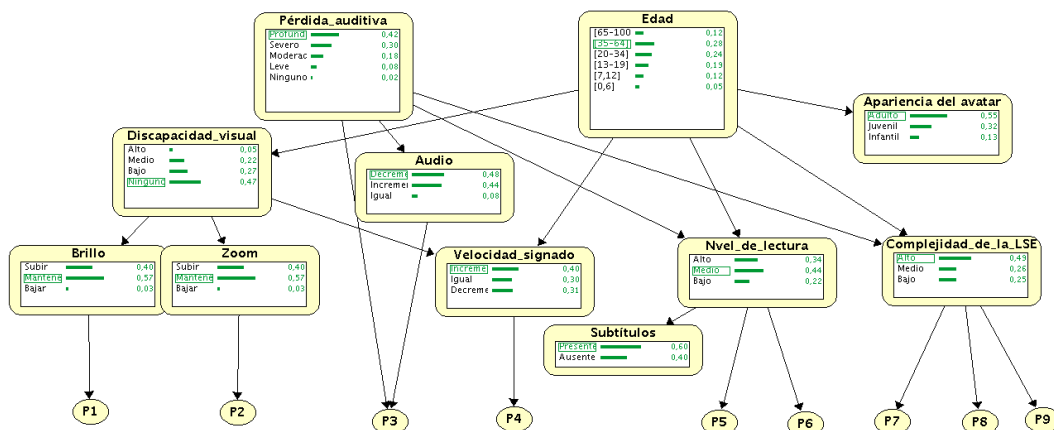


Figura 3.4: Perfil obtenido sin conocimiento de las preferencias del usuario.

Si se introducen hallazgos y se realiza la propagación de la evidencia observada, podemos analizar determinadas variables de forma individualizada. Suponemos que el usuario tiene 8 años y que aún no ha respondido a ninguna de las preguntas propuestas en el sistema, la red obtenida sería como la mostrada en la Figura 3.5: los valores de brillo y zoom serían los correctos ya que es poco probable que un niño tenga problemas en la vista, lo más probable es que fuera sordo profundo ya que la única información que se tiene del alumno es su edad.

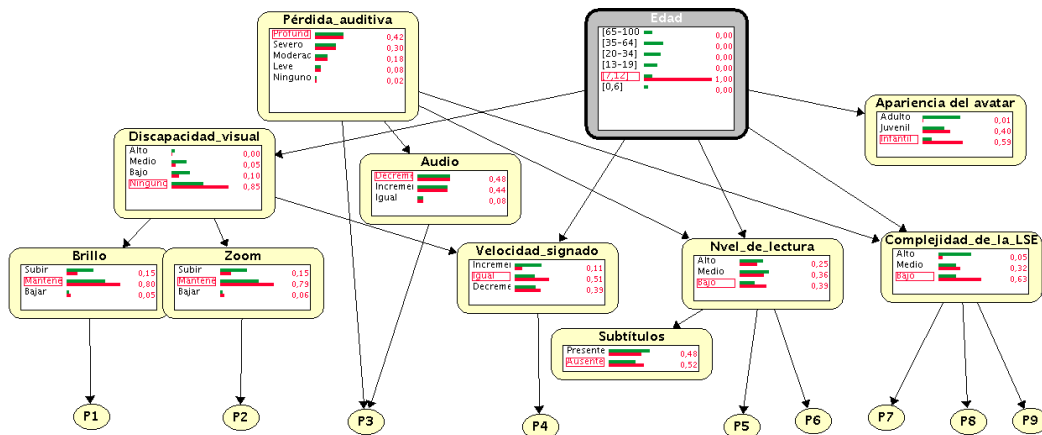


Figura 3.5: Propagación de evidencia conociendo sólo la edad del alumno (entre 7 y 12 años).

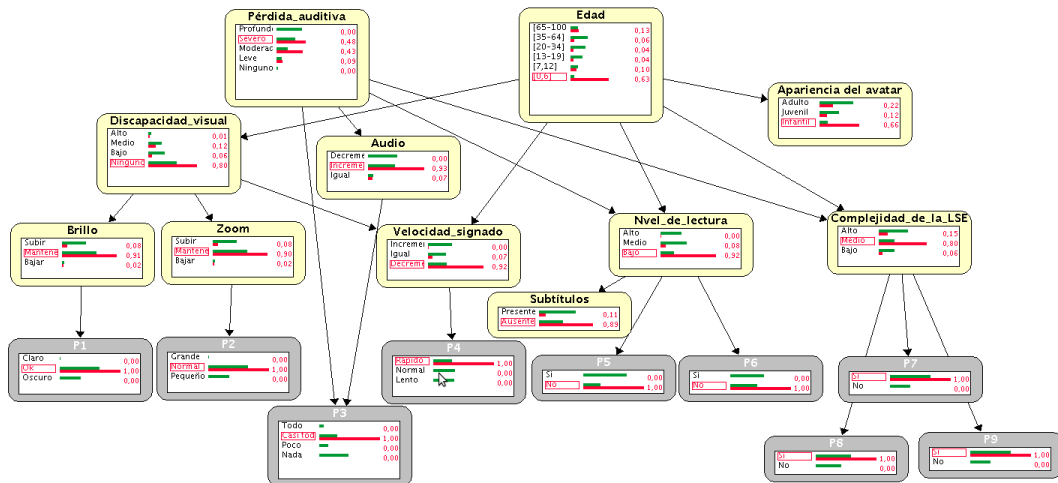


Figura 3.6: Propagación de la evidencia en el conocida la edad y las respuestas del alumno.

Ahora el alumno ha contestado de la siguiente forma a los test (Fig. 3.6): P1: bien, P2: normal, P3: casi todas, P4: rápido, P5: si, P6: no, P7: si, P8: si, P9: si. La evidencia se propaga a los nodos de medición de las capacidades del usuario, como el alumno

considera que el brillo y el zoom es el adecuado, el sistema no modificará estos parámetros. El usuario ha escuchado casi todas las preguntas realizadas por el personaje virtual, siendo así el modelo subirá el nivel del audio. También se puede observar que no se utilizarán subtítulos puesto que el usuario no ha comprendido los textos mostrados. Por último el nivel de complejidad de la LSE será medio ya que el alumno no ha comprendido las palabras signadas por el avatar.

Una de las ventajas que presenta el uso de este tipo de técnicas probabilísticas, es que aunque el usuario no conteste a todas las preguntas es posible obtener el perfil del mismo. Pero hay que tener en cuenta que cuanto mayor sea el número de preguntas contestadas mayor será también la probabilidad de obtener correctamente las preferencias y gustos del usuario.

3.2.5. Adaptación al contexto y al mensaje

La propuesta de adaptación se realiza a través de la generación de las expresiones faciales que más se adecúen al contexto de aplicación y al tipo de mensaje que será emitido. Si nos fijamos en los subtítulos (para personas sordas), suele aparecer emoticonos junto con determinadas frases con el fin de entender el contexto de la sentencia. Esta misma situación podemos extrapolarnos a los avatares cuya finalidad es transmitir información a personas con discapacidad auditiva utilizando la lengua de signos.

El esquema propuesto se trata básicamente de un controlador difuso cuyas partes son:

- Normalización de la entrada.
- Fuzzificación.
- Mecanismo de inferencia - base de reglas.
- Defuzzificación.
- Ajuste del valor de salida.

Variables de entrada

Para determinar las emociones que tendrá en cuenta el sistema, nos hemos centrado en seis emociones básicas propuestas por Ekman [16].

- Miedo: Anticipación de una amenaza o peligro que produce ansiedad, incertidumbre, inseguridad.
- Sorpresa: Sobresalto, asombro, desconcierto. Esta emoción es muy transitoria y puede dar una aproximación cognitiva para saber qué pasa.
- Disgusto, asco: solemos alejarnos del objeto que nos produce aversión.
- Rabia, enojo, resentimiento, furia, irritabilidad.
- Alegría, diversión, euforia, gratificación, contentos, da una sensación de bienestar, de seguridad.
- Pena, soledad, pesimismo.

A partir de las emociones básicas se pueden definir un gran conjunto de emociones mixtas: alegría-sorpresa, miedo-rabia, tristeza-miedo, etc. También pueden aparecer emociones que enmascaran a otras, por ejemplo cuando una persona se siente triste pero quiere mostrar alegría aparecería la emoción alegre-triste. En el trabajo actual no vamos a considerar este último tipo de emociones sino aquellas que pueden existir simultáneamente sin enmascarar una emoción principal.

Otro elemento que se ha tenido en cuenta es la edad del usuario que utilizará el modelo. Esto se debe a que cuando se le habla a un niño se suele enfatizar más o añadir más intensidad que cuando se trata de un adulto. Como resultado se han obtenido las siguientes variables difusas de entrada:

1. Emoción tristeza
2. Emoción alegría.
3. Emoción sorpresa.
4. Emoción ira.
5. Emoción miedo.
6. Emoción disgusto/asco.
7. Edad del usuario.

Una vez definidas las variables de entrada, la segunda fase consiste en normalizar los valores que toman. En este caso cada una de las variables pueden tomar valores del intervalo $[0, 1]$, un ejemplo sería tristeza = 0,2, sorpresa = 0,8, edad del usuario = 0.4. Cuanto menor sea el valor que tome la emoción menor será también su grado de influencia, es decir si miedo toma el valor 0.1 significa que la emoción de miedo es muy pequeña, mientras que tomando un valor de 0.9 se produciría el caso contrario.

Después de normalizar, los valores de la entrada se convierten en un conjunto de etiquetas lingüísticas mediante el proceso de Fuzzificación en el que se estudia la pertenencia de cada valor numérico a cada conjunto difuso. Las etiquetas definidas para las emociones son *Neutro*, *Alta* y *Muy Alta*, y para el caso de la variable edad las posibles etiquetas son *Muy pequeño*, *Pequeño*, *Joven*, *Adulto*

Por lo tanto tenemos siete variables de entrada $V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$; cada una con su correspondiente dominio de definición $DDV = \{DDV_1, DDV_2, DDV_3, DDV_4, DDV_5, DDV_6, DDV_7\}$ tal y como se muestra en la Figura 3.3. Los conjuntos difusos desarrollados para las variables $v_1, v_2, v_3, v_4, v_5, v_6$ se muestra en la Figura 3.7

Variable	Correspondencia	DDV
v_1	Tristeza	Neutro, alta, muy alta
v_2	Alegría	Neutro, alta, muy alta
v_3	Sorpresa	Neutro, alta, muy alta
v_4	Ira	Neutro, alta, muy alta
v_5	Miedo	Neutro, alta, muy alta
v_6	Asco	Neutro, alta, muy alta
v_7	Edad	Muy pequeño, pequeño, joven , adulto

Tabla 3.3: Variables de entrada y su dominio de definición.

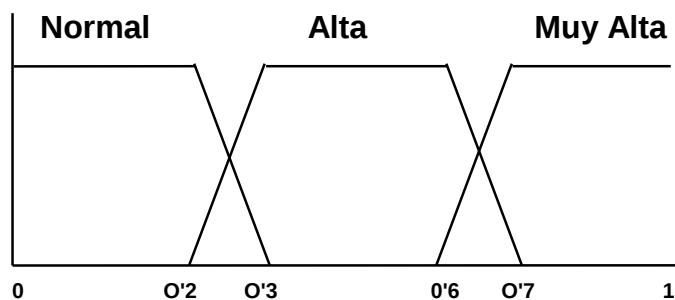


Figura 3.7: Variable v_i , emoción del avatar. $DDV_i = \{N, A, MA\}$; donde $i = \{1, 2, 3, 4, 5, 6\}$.

Variables de salida

Las variables de salida corresponden con los músculos involucrados en la emoción, número de frames totales de la emoción, número de frames desde que comienza la emoción hasta que está en su estado máximo (on) y número de frames necesario para que la emoción desaparezca (off).

Ekman y Friesen [15] desarrollaron una técnica para medir los movimientos faciales, el llamado Sistema de Codificación de Acciones Faciales (FACS). Este sistema descompone la expresión facial en 46 unidades de acción (AU), que se corresponden con los movimientos de cada uno de los músculos de la cara. En nuestro trabajo, con el fin de facilitar el movimiento de los músculos faciales, se han definido manipuladores de alto nivel asociados a uno o varios músculos faciales reales, por ejemplo en lugar de músculo orbicular se ha definido el manipulador “párpados”.

Los manipuladores definidos son los que se muestran en la Figura 3.8: entrecejo, parte interior de la ceja (derecha e izquierda), ceja (derecha e izquierda), párpados (derecho e izquierdo), dirección de la mirada, comisura del labio superior, labio superior, comisura labio (correspondiente a la parte derecha e izquierda), boca, comisura del labio, labios y boca. El movimiento de los manejadores se realizan en sentido vertical (bajar o elevar) excepto en el caso de la dirección de la mirada que puede realizarse tanto de forma horizontal como de forma vertical. El manejador *dirección de la mirada* permite mover la mirada del avatar a distintos puntos para dar mayor realismo a la expresión obtenida, por ejemplo si la emoción es *tristeza* el avatar bajará la mirada. La comisura del labio superior puede afectar sólo a la parte derecha, a la parte izquierda o a las dos, es principalmente utilizado para simular la expresión de *asco* que se caracteriza por elevar una única parte del labio superior. El manejador asociado con los labios permite que se saquen más hacia afuera, por ejemplo cuando se da un beso o hacia dentro. Por otro lado, el manejador asociado a la *boca* permite que el avatar pueda mover la mandíbula para abrir o cerrar con diferentes grados de apertura.

Por lo tanto se han definido diez variables de salida $Y = \{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}\}$; y su correspondiente dominio de definición $DDY = \{DDY_1, DDY_2, DDY_3, DDY_4, DDY_5, DDY_6, DDY_7, DDY_8, DDY_9, DDY_{10}\}$ tal y como se muestra en la Tabla 3.4. El dominio de definición de algunas de estas variables se muestran en la Figura 3.9. En la parte izquierda se muestra la variable y_{11} , número total de frames necesarios para

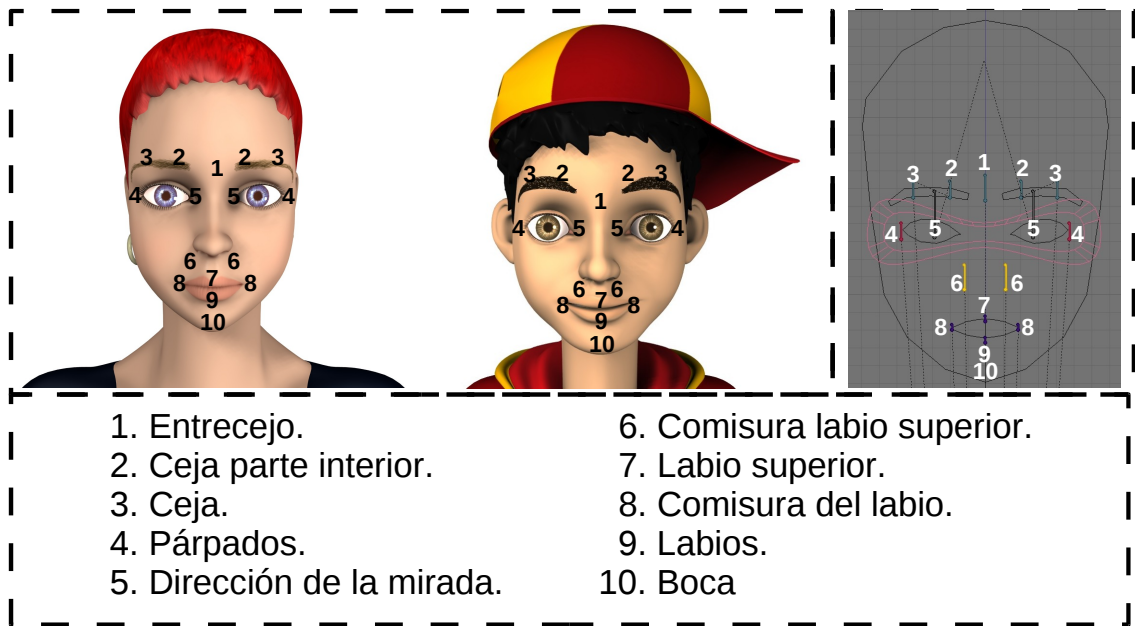
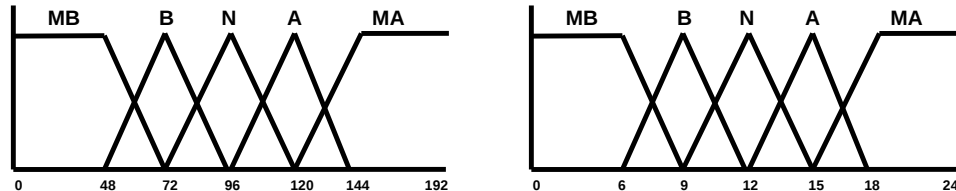


Figura 3.8: Manipuladores de alto nivel definidos para las expresiones faciales

expresar la emoción siendo su dominio de definición $DDY_{11} = \{MB, B, N, A, MA\}$. La figura de la derecha muestra los conjuntos difusos asociados a las variables y_{12} , y_{13} , número de *frames on* y número de *frames off*, siendo su dominio de definición $DDY_{12} = \{MB, B, N, A, MA\}$ y $DDY_{13} = \{MB, B, N, A, MA\}$.

Var.	Correspondencia	DDY
y_1	Entrecejo	Muy bajo, bajo, relajado, alto, muy alto
y_2	Parte interior de las cejas	Muy bajo, bajo, relajado, alto, muy alto
y_3	Cejas	Muy bajas, bajas, relajadas, altas, muy altas
y_4	Párpados	Muy cerrados, cerrados, un poco abiertos, abiertos, muy abiertos
y_5	Dirección de la mirada	Muy baja, baja, relajada, alta, muy alta
y_6	Labio superior	Muy bajo, bajo, normal, alto, muy alto
y_7	Comisura del labio superior	Relajada, alta, muy alta
y_8	Comisura del labio	Muy baja, baja, relajada, alta, muy alta
y_9	Labios	Relajados, un poco hacia afuera, muy hacia afuera
y_{10}	Boca	Cerrada, un poco abierta, bastante abierta, abierta, muy abierta
y_{11}	Frames totales	Muy bajo, bajo, normal, alto, muy alto
y_{12}	Frames on	Muy bajo, bajo, normal, alto, muy alto
y_{13}	Frames off	Muy bajo, bajo, normal, alto, muy alto

Tabla 3.4: Variables de salida y su dominio de definición.

Figura 3.9: Conjuntos difusos difusos para las variables de salida. Izquierda: *frames totales*. Derecha: *frames on* y *frames off*.

Reglas difusas definidas para la simulación de las emociones

En la definición de las reglas difusas se han tenido en cuenta la taxonomía facial de Ekman y Friesen [15] y la adaptación a un número determinado de músculos presentada por Cosi et al. [20]. A partir de este estudio se han adaptado los músculos faciales al conjunto de manejadores definidos en el avatar que intervienen en cada una de las seis emociones básicas (Tabla 3.5), de esta forma se ha definido el conjunto de reglas difusas que relacionan las variables de entrada con las variables de salida para generar las emociones deseadas. Un ejemplo de los resultados obtenidos puede verse en la Figura 3.10, en ella se muestra las emociones sorpresa y tristeza más la emoción neutra, así como dos niveles diferentes de intensidad para estos dos tipos de emociones básicas.

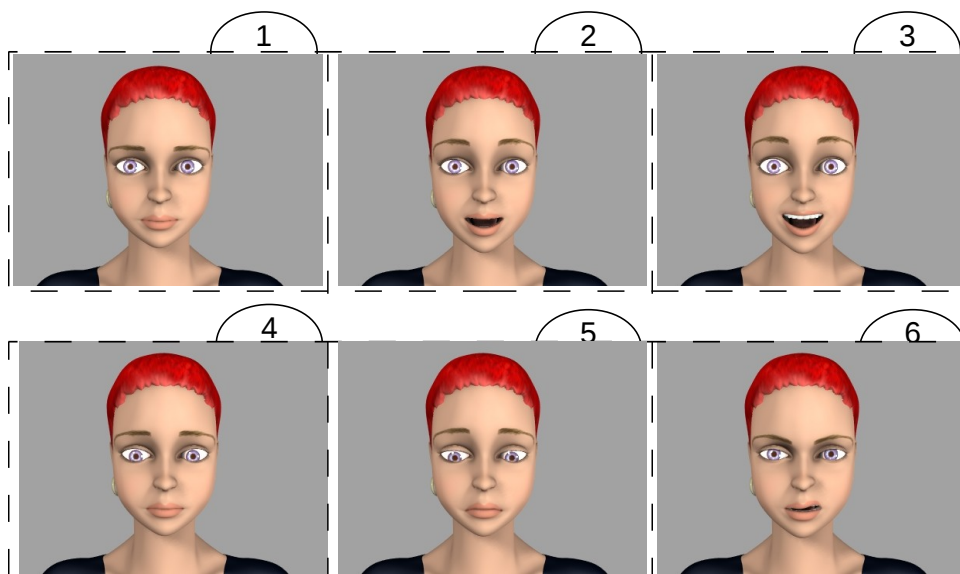


Figura 3.10: Resultados obtenidos aplicando las reglas difusas. 1. Neutro. 2. Sorpresa alta 3. y alegría alta. 4. Tristeza alta. 5. Tristeza muy alta. 6. Asco muy alto

Emoción	Descripción
Ira	Interior de las cejas hacia abajo. Ojos abiertos. Labios enseñando los dientes
Miedo	Cejas levantadas. Interior de las cejas hacia arriba. Ojos en tensión.
Asco	Cejas hacia abajo. Párpados relajados. Labio superior se eleva rizado.
Alegría	Cejas relajadas. Boca abierta. Esquina de la boca tira hacia las orejas.
Tristeza	Interior de las cejas doblada hacia arriba. Ojos un poco cerrados. Boca relajada. Esquina labio hacia abajo.
Sorpresa	Cejas levantadas. Ojos muy abiertos. Boca abierta.

Tabla 3.5: Emociones y su correspondiente descripción facial.

Mediante un proceso de inferencia difusa se comprueba el grado de satisfacción de las reglas activadas en función de los valores de las variables de entrada. El sistema seleccionará aquella regla cuyo grado de satisfacción sea mayor. En el consecuente de esta regla se indica como se debe modificar la expresión facial del avatar a través del uso de los manejadores faciales. Posteriormente, en la etapa de desfuzzificación se obtienen los valores numéricos para posicionar los manejadores faciales en el espacio 3D, determinar el número de frames totales de la animación, el número de frames de comienzo de la emoción y el número de frames de fin de la emoción.

A continuación se muestra un subconjunto de las reglas difusas definidas en el que se tienen en cuenta tanto la generación de emociones básicas como la combinación

entre ellas con diferentes grado de intensidad:

R1: If V_1 is MUY ALTA, V_7 is ADULTO then Y_1 MUY ALTO, Y_2 MUY ALTO, Y_3 MUY ALTAS, V_4 POCO ABIERTOS, Y_5 MUY BAJO, Y_{10} NEUTRA, Y_{11} NORMAL, Y_{12} NORMAL, Y_{13} ALTO.

R2: If V_2 is ALTA, V_7 is PEQUEÑO then Y_1 RELAJADO, Y_2 RELAJADO, Y_3 RELAJADAS, Y_4 ABIERTOS, Y_5 RELAJADA, Y_6 ALTO, Y_8 ALTA, Y_{10} POCO ABIERTA, Y_7 RELAJADA, Y_{11} ALTO, Y_{12} BAJO, Y_{13} NORMAL.

R3: If V_2 is MUY ALTA, V_7 is ADULTO then Y_1 RELAJADO, Y_2 RELAJADO, Y_3 RELAJADAS, Y_4 ABIERTOS, Y_5 ALTA, Y_6 MUY ALTO, Y_{11} MUY ALTA, Y_{13} ABIERTA, Y_7 RELAJADA, Y_{11} ALTO, Y_{12} MUY BAJO, Y_{13} NORMAL.

R4: If V_3 is ALTA, V_7 is ADULTO then Y_1 RELAJADO, Y_2 RELAJADO, Y_3 ALTAS, Y_4 MUY ALTAS, Y_5 ALTA, Y_{11} RELAJADA, $Y_{13} = \{\text{baste ABIERTA, MUY ABIERTA}\}$, Y_7 RELAJADA, Y_{11} ALTO, Y_{12} BAJO, Y_{13} ALTO.

R5: If V_4 is MUY ALTA, V_7 is ADULTO then Y_1 MUY BAJO, Y_2 MUY BAJO, Y_3 MUY BAJAS, Y_4 MUY ABIERTOS, Y_{11} BAJA, Y_{13} CERRADA, Y_7 RELAJADA, Y_{11} ALTO, Y_{12} NORMAL, Y_{13} ALTO.

R6: If V_5 is ALTA, V_7 is PEQUEÑO then Y_1 ALTO, Y_2 ALTO, Y_3 ALTAS, Y_4 ABIERTOS, Y_5 RELAJADA, Y_{13} POCO ABIERTA, Y_7 RELAJADA, Y_{11} ALTO, Y_{12} BAJO, Y_{13} ALTO.

R7: If V_6 is MUY ALTA, V_7 is ADULTO then Y_1 RELAJADO, Y_2 RELAJADO, Y_3 BAJAS, Y_4 ABIERTOS, Y_{11} RELAJADA, Y_{13} POCO ABIERTA, Y_7 ALTA, Y_{11} NORMAL, Y_{12} NORMAL, Y_{13} ALTO.

R8: If V_1 is ALTA, V_6 ALTA, V_6 is ADULTO then Y_1 ALTO, Y_2 ALTO, Y_3 ALTAS, Y_4 {POCO ABIERTOS, ABIERTOS}, Y_5 BAJA, Y_{11} RELAJADA, Y_{13} POCO ABIERTA, Y_7 ALTA, Y_{11} NORMAL, Y_{12} NORMAL, Y_{13} ALTO.

R9: If V_3 is ALTA, V_5 is MUY ALTA, V_7 is ADULTO then Y_1 MUY ALTO, Y_2 MUY ALTO, Y_3 MUY ALTAS, Y_4 POCO ABIERTOS, Y_{11} RELAJADA, Y_{13} BASTANTE ABIERTA, Y_7 RELAJADA, Y_{11} NORMAL, Y_{12} NORMAL, Y_{13} NORMAL.

R10: If V_4 is ALTA, V_6 is MUY ALTA, V_7 is PEQUEÑO then $Y_1 = \{\text{BAJO, MUY BAJO}\}$, $Y_2 = \{\text{BAJO, MUY BAJO}\}$, Y_3 RELAJADAS, Y_4 ABIERTOS, Y_6 ALTO, Y_{11} RELAJADA, Y_{13} POCO ABIERTA, Y_7 MUY ALTA, Y_{11} ALTO, Y_{12} BAJO, Y_{13} NORMAL.

R11: If V_2 is MUY ALTA, V_3 is MUY ALTA, V_7 is ADULTO then Y_1 RELAJADO, Y_2 RELAJADO, Y_3 MUY ALTAS, Y_4 MUY ABIERTOS, Y_{11} RELAJADA, $Y_{10} = \{\text{ABIERTA, MUY ABIERTA}\}$, Y_7 MUY ALTA, Y_{11} NORMAL, Y_{12} MUY BAJO, Y_{13} NORMAL.

4

Conclusiones y trabajo futuro

En el presente trabajo se ha descrito una arquitectura que pretende facilitar el uso de intérpretes virtuales de la lengua de signos (LS) en el mayor número de ámbitos posibles. Para conseguir este objetivo se ha hecho uso de técnicas de *softcomputing* mediante las que se han logrado diferentes niveles de adaptación.

En primer lugar se ha tenido en cuenta la *adaptación al usuario*, es decir a las preferencias y características que presenta cada uno de los receptores del mensaje. Aunque el modelo presentado está orientado al uso de la lengua de signos, es necesario tener presente que no todos los sordos sufren el mismo grado de discapacidad auditiva por lo que para aquellas personas que puedan percibir sonidos será recomendable acompañar el signado del avatar con información audible. Por otro lado, el uso de subtítulos no es siempre recomendable, ya que determinadas personas sordas (principalmente los prelocutivos) tienen serias dificultades en el aprendizaje de la lengua escrita, en este caso el uso de subtítulos es innecesario. La lengua de signos, al igual que el resto de lenguas requiere de un proceso de aprendizaje, es decir, no todos los usuarios que conocen la LS tienen el mismo nivel de comprensión, siendo éste otro parámetro que se puede ajustar a las necesidades del usuario. A parte de estos elementos existen otros menos relacionados con el problema de la sordera pero que influyen en el nivel de comprensión por parte del receptor del mensaje, por ejemplo si el usuario sufre de problemas en la visión será necesario adaptar determinados parámetros como el brillo del vídeo o el tamaño del avatar. Debido a que este tipo de información es imprecisa, se ha utilizado métodos probabilísticos (concretamente redes bayesianas) para obtener y ajustar las características de la aplicación.

En segundo lugar, es necesario resaltar la importancia de las expresiones faciales especialmente cuando la comunicación se realiza empleando la lengua de signos. Atendiendo a este elemento se ha considerado un nivel de *adaptación al contexto* mediante el cual, las expresiones del avatar deben ajustarse a la situación en la que se lleva a cabo el mensaje. A través de la simulación de emociones se consigue que la información transmitida adquiera el “tono” que tendría si se utilizara la lengua hablada. Debido a

que este tipo de simulación lleva intrínseco un alto grado de vaguedad e imprecisión, se han utilizado un conjunto de reglas difusas para obtener el movimiento de los músculos faciales del avatar. Al igual que ocurre en las emociones reales, las simuladas también pueden ser fruto de mezcla de diferentes emociones, de esta manera partiendo de un número limitado de emociones básicas se obtiene un número elevado de emociones mixtas (alegría-sorpresa, miedo-rabia, tristeza-miedo, etc.). Independientemente del tipo de emoción, cada una de ellas se puede expresar con diferentes grados de intensidad ya que no es lo mismo, por ejemplo, estar triste que estar muy triste, o estar un poco asustado y muy asustado. La intensidad es representada asignando diferentes grados de posicionamiento de los manejadores faciales. Estos manejadores han sido definidos para facilitar el proceso de animación, ya que el movimiento de cada uno de ellos afecta a un músculo o conjunto de músculos de la cara del avatar.

Únicamente con estos niveles de adaptación no se conseguiría el objetivo planteado si la animación del avatar no fuera realista. En el módulo de render de la arquitectura presentada se presta especial atención a la obtención de transiciones suaves y fluidas entre dos signos consecutivos y en definitiva, en obtener una animación con un alto grado de realismo. Además se han definido un conjunto amplio de parámetros de render que permite la adaptación del video generado a las necesidades particulares del dispositivo de visualización (televisor, PDA, teléfonos móviles, etc.).

En cuanto a posibles ampliaciones o mejoras de la arquitectura presentado en este trabajo y líneas de trabajo futuro, podemos destacar las siguientes:

- Uso de mecanismos más complejos para obtener el perfil del usuario y conseguir por tanto una adaptación más eficiente. El uso de preguntas puede entrañar dificultades en ciertos dominios de aplicación.
- Ampliación del número de factores a los que afecta el contexto del mensaje tales como la ropa del avatar (traje, deportivo, informal,..etc), la edad del mismo o el fondo utilizado (color utilizado, uso de una imagen,...).
- Definición de mecanismos que identifiquen de forma automática el contexto en el que se emite el mensaje es decir, identificar cuando se más aconsejable utilizar emociones de alegría, tristeza, seriedad, etc. Para ello se pueden utilizar técnicas de minería de texto.
- Estudio de los mecanismos para lograrla adaptación gramatical del mensaje según la región, el nivel cultural del receptor, etc.
- Mejora de algunos aspectos técnicos de la arquitectura: soporte de nuevos métodos de render, manejadores cinemáticos de más alto nivel, soporte para la captura automática del movimiento, mejoras en el procesamiento del lenguaje natural, etc.

Bibliografía

- [1] N. Adamo-Villani, E. Carpenter, and L. Arns. An immersive virtual environment for learning sign language mathematics. pages 20–23, 2006.
- [2] N. Adamo-Villani, J. Doublestein, and J Martin. The mathsigner: An interactive learning tool for american sign language k-3 mathematics. *International Conference on Information Visualisation*, pages 713–716, 2004.
- [3] N. Adamo-Villani and R. Wilbur. Novel approaches to deaf education. In *First International Conference on Technology-Based Learning with Disability*, pages 13–21, Julio 2007.
- [4] N. Adamo-Villani and K. Wriugh. Smile: an immersive learning game for deaf and hearing children. pages 17–22, 2007.
- [5] O. Al-Jarrah and A. Halawani. Recognition of gestures in arabic sign language using neuro-fuzzy systems. *Artificial Intelligence*, 133(1-2):117–138, 2001.
- [6] A. Arya, S. DiPaola, and A. Parush. Perceptually Valid Facial Expressions for Character-Based Applications. *International Journal of Computer Games Technology*, 20:1–13, 2009.
- [7] J. Bangham, S. Cox, R. Elliot, J. Glauert, and I. Marshall. Virtual signing: Capture, animation, storage and transmission-an overview of the visicast project. In *IEEE Seminar on Speech and Language Processing for Disabled and Elderly People*, IEEE Press, pages 1–7, 2000.
- [8] D. Benyon. Adaptive systems: A solution to usability problems. *User Modeling and User-Adapted Interaction*, 3(1):65–87, 1993.
- [9] E. Bevacqua, Mancini, and C. Pelachaud. Speaking with Emotions. In *Proceedings of the AISB Symposium on Motion, Emotion and Cognition*, pages 197–214. Leeds, U.K, March 2004.
- [10] T.D. Bui, D. Heylen, M. Poel, and A. Nijholt. Generation of facial expressions from emotion using a fuzzy rule based system. In *14th Australian Joint Conf. on Artificial Intelligence*, pages 83–94, 2001.
- [11] S. Buisine, S. Abrilian, R. Niewiadomski, J.C. Martin, L. Devillers, and C. Pelachaud. Perception of blended emotions: From video corpus to expressive agent. *Intelligent Virtual Agents. LNCS, Springer Berlin*, 4133:93–106, 2006.

- [12] B. R. Callejas, G.P. Dimuro, and A. C. Rocha. Interval fuzzy rule-based hand gesture recognition. In *SCAN '06: Proceedings of the 12th GAMM - IMACS International Symposium on Scientific Computing, Computer Arithmetic and Validated Numerics*, pages 12–19. Washington, DC, USA, 2006.
- [13] A. S. Drigas, J. Vrettaros, and D. Kouremenos. Teleeducation and e-learning services for teaching english as a second language to deaf people, whose first language is the sign language. *Education and Information Technologies*, 1(3):1–5, 2004.
- [14] A. S. Drigas, J. Vrettaros, and D. Kouremenos. An e-learning management system for the deaf people. In *AIKED'05: Proceedings of the 4th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering Data Bases*, pages 1–5. Stevens Point, Wisconsin, USA, 2005.
- [15] P. Ekman and W. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto., 1978.
- [16] P. Ekman, W. V. Friesen, and P. Ellsworth. *Emotion in the human face: guidelines for research and a review of findings*. 1972. ISBN 978-0195155327.
- [17] Sánchez Martínez F., Pérez Ortiz J.A., and M. Forcada. Using target-language information to train part-of-speech taggers for machine translation. *Machine Translation*, 22(1):29–66, 2008.
- [18] G. Fang, W. Gao, and D. Zhao. Large vocabulary sign language recognition based on fuzzy decision trees. *IEEE Transactions on Systems, Man and Cybernetics*, 34(3):305–314, 2004.
- [19] N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian network classifiers. *Machine Learning*, 29:131–163, 1997.
- [20] P. Fusaro. Data-Driven Tools for Designing Talking Heads Exploiting Emotional Attitudes. In *Proceedings of Tutorial and Research Workshop, Affective Dialogue Systems*, pages 101–112, Junio 2004.
- [21] J. Glauert, R. Kennaway, R. Elliott, and Theobald B. Virtual human signing as expressive animation. In *Symposium on Adaptive Agents and Multi-Agent Systems, Leeds, UK*, pages 88–106.
- [22] C. González Morcillo, J. A. Fernández Sorribes, and L. Jimenez Linares. Grid architecture for distributed rendering. pages 149–163, 2006.
- [23] A. Jameson. *Adaptive interfaces and agents*. Jacko and Andrew Sears, editors, Handbook of Human-Computer Interaction in Interactive Systems. Erlbaum, Mahwah, NJ, 2002. ISBN 0-8058-3838-4.
- [24] M. Jemni and E. Elghougl. Towards web-based automatic interpretation of written text to sign language. In *First International conference on ICT & Accessibility, Hammamet, Túnez*, pages 43–48, 2007.
- [25] M. Jemni, E. Elghougl, N.B Yahia, and M. Boulares. Sign language mms to make cell phones accessible to deaf and hard-of-hearing community. In *Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments, Granada, España*, pages 24–29.

- [26] K. Jong-Sung, W. Jang, and Z. Bien. A dynamic gesture recognition system for the korean sign language(ksl). *Autism Advocate, First Edition*, 26(2):354–359, 2002.
- [27] L. Kavan, S. Collins, J. Žára, and C. O’Sullivan. Geometric skinning with approximate dual quaternion blending. *IEEE Transactions on Robotics and Automation*, 27(4):1–23, 2008.
- [28] J. Kim, K. Park, W. Bang, and Z.Z. Bien. Continuous gesture recognition system for korean sign languagebased on fuzzy logic and hidden markov model. In *Proceedings of the 2002 IEEE International Conference on FUZZ-IEEE.*, pages 1574–1579.
- [29] P. Koehn, Hoang H., A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst. Moses: Open source toolkit for statistical machine translation. In *In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 177–180. Stevens Point, Wisconsin, USA, 2007.
- [30] P. Langley. User modeling on adaptive interfaces. In *Seventh International Conference on User Modeling, Secaucus, NJ, USA*, pages 357–370, 1999.
- [31] N. Lever. *Real-Time 3D Character Animation with Visual C++*. Focal Press An imprint of Butterworth Heinemann, Linacre House, Jordan Hill, 2002. ISBN 978-0-240-51664-6.
- [32] P. Markellou, M. Rigou1, S. Sirmakessis, and A. Tsakalidis. A web adaptive educational system for people with hearing difficulties. *Education and Information Technologies*, 5(3):189–200, 2004.
- [33] J.C. Martin, R. Niewiadomski, L. Devillers, S. Buisine, and C. Pelachaud. Multi-modal complex emotions: Gesture expressivity and blended facial expressions. In *Journal of Humanoid Robotics*, 3:831–843, 2006.
- [34] S. Mcroy and R. Freedman. What is an intelligent tutoring system? *Intelligence*, 11(3):15–16, 200.
- [35] L. N. Michaud, K. F. McCoy, and C. A. Pennington. An intelligent tutoring system for deaf learners of written english. In *Proceedings of the fourth international ACM SIGCAPH conference on Assistive technologies*, pages 92–100, Noviembre 2000.
- [36] E. Millán, M. Trella, J.L Pérez de la Cruz, and R. Conejo. Uso de redes bayesianas en test adaptativos computerizados. In *Informática y Educación en el siglo XXI. Congreso Nacional de Informática Educativa*, pages 17–19, Noviembre 1999.
- [37] T. Murray, S. Blessing, and Ainsworth S. *Authoring Tools for Advanced Learning Environments*. Fkluwer Academic Publishers, 2003. ISBN 978-1-4020-1772-8.
- [38] Andrew and Clore Gerald L. and Collins Ortony. *The cognitive structure of emotions*. Cambridge University Press. ISBN 978-0521386647.
- [39] Brusilovsky P. and Millán E. User models for adaptive hypermedia and adaptive educational. *The Adaptive Web, LNCS*, 4321:3–53, 2007.
- [40] L.D. Paulson. News briefs. *COMPUTER*, 41(2):23–25, 2008.

- [41] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. 1998. ISBN 978-1558604797.
- [42] R. W. Picard. *Affective Computing*. MIT Press, Cambridge, Massachusetts, 1997. ISBN 9780262161701.
- [43] R. W. Picard and M.S. Goodwin. Developing Innovative Technology for Future Personalized Autism Research and Treatment. *Autism Advocate, First Edition*, 50(1): 32–39, 2008.
- [44] A. Raouzaiou, N. Tsapatsoulis, and S. Kollias. Parameterized Facial Expression Synthesis For Videoconferencing Applications. *EURASIP Journal on Applied Signal Processing*, 2002:1021–1038, 2002.
- [45] I.R. Rodríguez. *Comunicar a través del silencio: las posibilidades de la lengua de signos española*. Universidad de Sevilla, 2005. ISBN 9788447208968.
- [46] M.A. Rodríguez. Lenguaje de signos. *Signa: revista de la Asociación Española de Semiótica, Edición digital de la Biblioteca Virtual Miguel de Cervantes*, (2):193–200, 1993.
- [47] J.A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.
- [48] R. San-Segundo, R. Barra, R. Córdoba, L. F. D'Haro, F. Fernández, J. Ferreiros, J. M. Lucas, J. Macías-Guarasa, J. M. Montero, and J. M. Pardo. Speech to sign language translation system for spanish. *Speech Commun.*, 50(11-12):1009–1020, 2008.
- [49] L. Semler and Carhart. The Automation of Non-manual Signals of Declarative Sentences in American Sign Language. DePaul CTI Research Symposium (CTIRS) & Midwest Software Engineering Conference (MSEC), Chicago, Illinois, 29 apr 2006.
- [50] B. Shackel. Human factors and usability. *Human-computer interaction, Prentice Hall Press*, pages 27–41, 1990.
- [51] Sosa, A. and Causa, E. Proyecto de investigación Desarrollo en multimedia del arte bio generativo y los sistemas de captación del gesto y la emoción humana.
- [52] M. C. Su. A fuzzy rule-based approach to spatio-temporal hand gesture recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 30(2):276–281, 2000.
- [53] J. Tao and T. Tan. Affective computing: A review. *Affective Computing and Intelligent Interaction, Lecture Notes in Computer Science*, 3784/2005:981–995, 2005.
- [54] C. Tecuanhuehue. Interfaces afectivas en el contexto de bibliotecas digitales udlap. tesis licenciatura. 2004.
- [55] R. E. Thayer. *The Biopsychology of Mood and Arousal*. Oxford University Press, New York, NY, USA, 1989. ISBN 978-0195051629.
- [56] K. Vanlehn. The behavior of tutoring systems. *Int. J. Artif. Intell. Ed.*, 16(3):227–265, 2006.

- [57] L. Varadharajan Xu, Maravich V., R. J. Tongia, and J. Mostow. Design: An intelligent tutor to teach american sign language. In *Proceedings of the SLATE Workshop on Speech and Language Technology in Education. Carnegie Mellon University and International Speech Communication Association (ISCA)*, pages 15–18, 2007.
- [58] M. Verlinden, I. Zwitterlood, and H. Frowein. Multimedia with animated sign language for deaf learners. In *World Conference on Educational Multimedia, Hypermedia and Telecommunications ED-MEDIA, Montreal, Canada*, pages 4759–4764.
- [59] T.L. Wang and C.C. Chen. A combined optimization method for solving the inverse kinematics problem of mechanical manipulators. *IEEE Transactions on Robotics and Automation*, 7(4):489–499.
- [60] G. Wen, C. Yiqiang, F. Gaolin, Y. Changshui, J. Dalong, G. Chunbao, and W. Chunli. Handtalker ii: a chinese sign language recognition and synthesis system. In *ICARCV*, pages 759–764, 2004.
- [61] www.engadget.com/2007/09/13/ibms-sisi-virtually-translates-speech-to-sign-language/. Artículo referente a la aplicación SiSi desarrollada por IBM.
- [62] www.mhp.org. Página oficial del estándar Multimedia Home Platform.
- [63] www.vcom3d.com. Página oficial de la empresa VCom3D.
- [64] www.visicast.cmp.uea.ac.uk/eSIGN/Public.htm. Página web del proyecto ESign.
- [65] Wilks Y. *Machine Translation: Its Scope and Limits*. Springer, Wilks, Yorick, 2009. ISBN 978-0-387-72773-8.



Resumen de las asignaturas cursadas en el Máster

A.1. Metodologías y técnicas de investigación en informática

Título: Metodologías y técnicas de investigación en informática.

Profesorado: Dra. Marcela Genero Bocco y Dr. Mario Piattini Velthuis.

Contenido:

El objetivo principal de la asignatura es enseñar al alumno cómo realizar una tesis doctoral. Este objetivo, a su vez, se divide en cuatro subobjetivos: (1) dotar al alumno de los conocimientos necesarios para estructurar el contenido de la tesis doctoral, (2) ofrecer guías sobre cómo escribir una comunicación científica, (3) cómo realizar revisiones sistemáticas, y finalmente, (4) presentar los métodos de investigación más adecuados para la validación y contrastación de las hipótesis de investigación.

Los métodos de enseñanza aplicados en la asignatura son básicamente tres: clases magistrales impartidas por los profesores citados anteriormente, lectura de artículos científicos y debate entre los alumnos sobre dichos artículos. En cuanto al método de evaluación se tiene en cuenta la asistencia y la participación en clase, así como la entrega de un trabajo, el cual debe ser expuesto por el alumno en un tiempo máximo de doce minutos y tres para preguntas. Dicho trabajo consiste en la realización de una revisión sistemática sobre un tema seleccionado por el alumno, haciendo especial hincapié en los diferentes estudios empíricos.

Resultados:

Una primera actividad para identificar las fases de la revisión sistemática en un ar-

título en concreto (*Motivation in Software Engineering: a systematic literature review*). Posteriormente se realizó una revisión sistemática sobre los sistemas e interfaces adaptativos existentes basados en redes bayesianas. Dicha revisión se utilizó como punto de partida en el desarrollo de un prototipo orientado a personas sordas que ofrece información en lengua de signos española, a través de un intérprete virtual, adaptándose a las características de cada usuario. El estudio realizado nos permitió conocer que las investigaciones existentes en el campo de los sistemas adaptativos han crecido en los últimos años, pero la mayoría de los resultados obtenidos son prototipos experimentales y dado que fueron encontrados un número reducido de estudios en los que se integran avatares conversacionales con la adaptación, indica el área a tratar a penas ha sido explotada aún.

A.2. Sistemas distribuidos avanzados: grid e inteligencia ambiental

Título: Sistemas distribuidos avanzados: grid e inteligencia ambiental.

Profesorado: Dra. Camelia Muñoz Caro, Dr. Alfonso Niño Ramos y Dr. José Bravo Rodríguez.

Contenido: La asignatura se divide en dos partes bien diferenciadas. Una primera parte titulada Sistemas Grid, impartida por los profesores Camelia Muñoz y Alfonso Niño. Y una segunda parte titulada inteligencia ambiental impartida por el profesor José Bravo. La primera de ellas tiene como gran objetivo dotar al alumno de los conocimientos necesarios para configurar y gestionar infraestructuras grid, así como diseñar sistemas grid a nivel de servicios básicos y servicios web. Respecto a la segunda parte, se estudian nuevas formas de interacción en donde se manifiesta la proactividad de los sistemas. Además, se pretende capacitar al alumno para el modelado de acciones de usuario en sistemas Aml, realizándose estudios reales sobre instalaciones en diferentes contextos, entre los que cabe destacar los centros de enseñanza y hospitalarios.

El método de evaluación de la asignatura se basa en la valoración de la asistencia a clase, participación en las actividades propuestas y, finalmente, en la entrega y defensa de dos trabajos teóricos.

Resultados:

Para superar la asignatura se realizaron y defendieron dos trabajos teóricos:

- *Computación GRID y su aplicación en el renderizado de escenas 3D.* En este trabajo se realizó una introducción a los sistemas Grid analizando su evolución histórica y aquellos elementos que lo distinguen de los sistemas distribuidos tradicionales y de otro tipo de redes como las P2P. La parte central del documento está enfocado hacia un tipo particular de Grid utilizado para el renderizado de escenas 3D.
- *Grid computing vs cloud computing.* En este trabajo se describieron las similitudes y diferencias existentes entre dichos paradigmas. Debido a la reciente aparición de la *computación nube*, no existe aún un consenso en su definición y puede ser confundida con la computación grid, por esta razón se recopilaron un conjunto de opiniones sobre lo que implica esta nueva aproximación así como las ventajas

e inconvenientes que supone su uso. El trabajo finalizaba con una comparación entre el sistema cloud de Amazon y EGEE (grid).

A.3. Sistemas avanzados de interacción persona-computador: sistemas colaborativos y computación ubicua

Título: Sistemas avanzados de interacción persona-computador: sistemas colaborativos y computación ubicua.

Profesorado: Dr. Manuel Ortega Cantero, Dr. Miguel Ángel Redondo Duque, Dr. Crescencio Bravo Santos

Contenido:

Este curso tuvo como objetivo principal la introducción de los conceptos teóricos y aspectos prácticos fundamentales para el diseño y desarrollo de sistemas colaborativos y/o ubicuos desde una perspectiva metodológica.

Se estudiaon varios aspectos relacionados con el desarrollo y la implementación de sistemas de colaboración entre usuarios para la ejecución de tareas de forma conjunta. Se ha estudiado el sistema colaborativo **COLLECE**, que es una herramienta que permite realizar programación colaborativa. Haciendo uso de esta herramienta se han estudiado aspectos relevantes en este tipo de aplicaciones como el *awareness*.

Finalmente, se estudian dos metodologías para el diseño de interfaces de usuario: ConcurTaskTrees (CTT) y CIAN. CTT es una notación creada por Fabio Paterno para modelar interfaces de usuario aplicadas a sistemas de soporte a la colaboración y cooperación. CIAN, por otro lado, es una metodología para el desarrollo de interfaces de usuario para sistemas Groupware en general. El método de evaluación de la asignatura se basa en la valoración de la asistencia a clase, participación en las actividades propuestas y, finalmente, en la entrega y defensa de dos trabajos teóricos.

Resultados:

Para superar la asignatura se realizaron diferentes trabajos y ejercicios:

- Búsqueda y resumen de un conjunto de artículos en los que se hablara de de computación ubicua y awareness.
- Desarrollo de un artículo que tratara algunos de los temas visto en clase y corrección de artículos escritos por compañeros de la asignatura. El título del trabajo fue “ECLSE. Editor Colaborativo de la Lengua de Signos Española”, en el que se describía un herramienta de edición colaborativa (inventada) para la generación de frases complejas en lengua de signos española a través de un avatar.
- Trabajo teórico de un entorno CSCL/CSCW en el que se diseñó un soporte de análisis de la interacción y colaboración siguiendo los pasos del framework de análisis proceso-solución visto durante la asignatura.
- Trabajo teórico en el que se exponían posibles mejoras en el soporte de awareness de COLLECE y presentación y propuesta de mejores del awareness en Google Docs.

- Realización el leguaje CTT y la herramienta CTTE un modelado de las tareas de usuario que soportaría una aplicación de gestión de correo electrónico para dispositivo móvil.
- Realización del modelado de tareas de usuario que soportaría la aplicación anterior siguiendo la metodología CIAM.

A.4. Pruebas y seguridad de Sistemas de Información

Título: Pruebas y seguridad de Sistemas de Información. **Profesorado:** Dr. Eduardo Fernández Medina Patón, Dr. Macario Polo Usaola.

Contenido: El bloque de pruebas de Sistemas de Información se centra en la adquisición de destrezas y conocimientos relacionados con los principales métodos y técnicas de las pruebas del software. Se estudian los distintos niveles de prueba: de caja negra, de caja blanca, unitarias, de integración y de sistema. También se profundiza en el **análisis de las medidas de cobertura** en las pruebas de caja blanca y en las estrategias de combinación para la generación de casos de prueba. **Resultados:** Se realizaron una serie de ejercicios teóricos y prácticos para profundizar en los planes de prueba y automatización de las mismas utilizando herramientas software como *testooj* y **mujava**.

En el segundo bloque de la asignatura, se aborda la problemática de la seguridad en los Sistemas de Información. Dado que la seguridad constituye un problema transversal, se estudian las diferentes dimensiones de la seguridad avanzada abordando desde la criptografía hasta las bases de datos pasando por la seguridad en procesos de negocio, almacenes de datos y servicios web. **Resultados:** Para esta parte de la asignatura, se analizó el artículo *An Engineering Porcess for Developing Secure DataWarehouses* y se escribió un trabajo relacionado con la seguridad y los sistemas de videovigilancia. En dicho trabajo se describen algunos elementos importantes que aparecen en la Legislación española en temas de seguridad y videovigilancia, y posteriormente se exponen algunos métodos utilizados para asegurar la privacidad de las personas en este tipo de sistemas. **Resultados:** .

A.5. Cognición y colaboración

Título: Cognición y Colaboración.

Profesorado: Dr. José Ángel Olivas Varela, Dr. Manuel Prieto Méndez y Dra. Carmen Lacave Rodero.

Contenido: La asignatura está estructurada en tres bloques temáticos bien diferenciados, cada uno de ellos impartido por uno de los profesores de la asignatura.

- Bloque I: Gestión del conocimiento y recuperación eficiente de la información en la Web mediante técnicas de Soft Computing (impartido por el Dr. José Ángel Olivas). El principal objetivo de este primer bloque es enseñar al alumno las principales técnicas basadas en Soft-Computing (tolerantes a la imprecisión e incertidumbre) para manipular información y extraer conocimiento de diversas fuentes.

- Bloque II: Gestión del conocimiento y computación gráfica (impartido por la Dra. Carmen Lacave Rodero). En este bloque se realiza una introducción al razonamiento probabilístico, haciendo mayor hincapié en el razonamiento con Redes Bayesianas.
- Bloque III: Gestión del conocimiento en grupo y aprendizaje comparado (impartido por el Dr. Manuel Prieto Méndez). El principal objetivo de este bloque es conocer las taxonomías y principios de funcionamiento de los sistemas de trabajo en grupo: análisis de los modelos CSCW, patrones para trabajo en grupo, aprendizaje automático y aprendizaje humano, etc.

Para superar el curso, el alumno debe realizar un trabajo encuadrado en alguno de los tres bloques anteriores.

Resultados: Se realizó un trabajo teórico encuadrado en el Bloque II: Gestión del conocimiento y computación gráfica, el cual fue supervisado y valorado por la Dra. Carmen Lacave Rodero. Dicho trabajo fue aceptado en el congreso internacional *Education and New Learning Technologies*. En el artículo se propone un sistema tutor inteligente que pueda adaptarse a las necesidades individuales del alumno sordo. El entorno de aprendizaje hace uso de un intérprete virtual de la lengua de signos que es el encargado de mostrar la información al alumno. El esquema propuesto está basado en el uso de redes bayesianas con el fin de identificar las características del avatar más adecuadas para cada usuario, adaptándolo a sus necesidades, para mejorar el proceso de aprendizaje en aquellas personas que sufren discapacidad auditiva

A.6. Técnicas de softcomputing

Título: Técnicas de softcomputing.

Profesorado: Dr. Luis Jiménez Linares, Dr. José Jesús Castro Sánchez y Dr. Juan Moreno García.

Contenido: En esta asignatura se presenta al alumno una nueva perspectiva para la definición de sistemas complejos mediante sistemas difusos, diferente a los métodos clásicos aprendidos durante la ingeniería en informática. Además, se pretende que el alumno obtenga cierta destreza en el uso y manejo de sistemas de reglas difusas, así como en el manejo de distintas metodologías para la definición de sistemas difusos. Para superar la asignatura se valora positivamente la asistencia a clase, así como la realización y la defensa de dos trabajos teórico-prácticos encuadrados en alguno de los temas principales enumerados anteriormente.

Resultados:

Se realizaron dos trabajos:

- Un trabajo teórico titulado “Razonamiento aproximado y Fuzzy Clips”. En el trabajo se realizó una introducción al tratamiento de la incertidumbre y la lógica difusa, así como a los métodos de inferencia. En la segunda parte se describía la inferencia con Fuzzy Clips para finalizar con un ejemplo teórico.
- Trabajo teórico-práctico en el que se realizó un estudio del estado del arte sobre los sistemas basados en lógica difusa cuya finalidad es la simulación de emociones.

En la segunda parte del trabajo se propuso un sistema basado en reglas difusas para la generación de un subconjunto de emociones en un intérprete virtual 3D.



Currículum Vitae

DATOS PERSONALES

- APELLIDOS: Herrera Tirado
- NOMBRE: Vanesa
- DNI: 5691153-X
- FECHA DE NACIMIENTO: 09-03-1980
- DIRECCIÓN PARTICULAR: c/Libertad 59A, 1ºB, Miguelturra (Ciudad Real).
- TELÉFONO: 676019773

SITUACIÓN PROFESIONAL ACTUAL

- Contratado por la Universidad de Castilla-La Mancha en el proyecto de investigación “GANAS: Generador Automático de la leNguA de Signos”, Cátedra Indra - UCLM.

FORMACIÓN ACADÉMICA

- Ingeniera en Informática por la Universidad de Castilla-La Mancha desde Junio de 2007.
- Actualmente cursando el Máster en Tecnologías Informáticas Avanzadas de la UCLM.

EXPERIENCIA LABORAL

- Beca en el Centro de Investigación y Desarrollo Soluziona en Ciudad Real en el proyecto "PATACS", rabajando en el desarrollo de un sistema de renta para Palestina. Desde el 3 de Enero de 2005 hasta el 10 de Marzo de 2006.
- Beca de Investigación del Vicerrectorado de la Universidad de Castilla la Mancha: *Visita virtual del área de nefrología del Hospital General de Ciudad Real*. Desde el 25 de Octubre de 2006 hasta el 25 de Febrero de 2007.
- Profesora de a academia *Tecnas* de Ciudad Real. Clases impartidas a nivel de E.S.O, Bachiller y Universidad. Desde el 14 de Febero de 2007 hasta el 15 de Junio de 2007.
- Contratada laboral en el proyecto I+D "GANAS" por la UCLM. Desde el 1 de Noviembre de 2007 hasta la fecha actual.

CURSOS DE FORMACIÓN DOCENTE Y OTROS CURSOS RECIBIDOS

- Conferencias X Aniversario ESI (Noviembre de 1998, UCLM).
- Sistemas de Radiocomunicaciones (Marzo de 1999, UCLM).
- Mantenimiento del Software (Marzo de 2001, UCLM).
- Clústers en Estaciones de Trabajo de Altas Prestaciones (Enero de 2002, UCLM).
- Desarrollo de componentes Enterprise JavaBeans (Marzo de 2003, UCLM).
- Programación de aplicaciones web con Python Zope (Enero de 2004, UCLM).
- Programación con Visual Basic .NET (Noviembre de 2004, UCLM).
- Taller de diseño gráfico: a todo color (Agosto de 2008, UPV)
- Miembro del comité organizador del "XIII Congreso Español sobre Tecnologías y Lógica Fuzzy y VII Jornadas sobre Transferencia de Tecnología Fuzzy" (Septiembre de 2006)

PUBLICACIONES

CONGRESOS Y PONENCIAS

- GANAS: a flexible architecture for 3D sign language rendering. Vanesa Herrera Tirado, Carlos González Morcillo, Maria de los Ángeles García Marín, Jose Ángel Mateos Ramos, Ignacio Arriaga Sánchez. *IADIS International Conference Interfaces and Human Computer Interaction*, páginas 61-68. Junio de 2009, Algarve (Portugal).
- An educational tool for people with hearing disabilities based on bayesian networks. Vanesa Herrera Tirado, Carlos González Morcillo, Carmen Lacave Rodeiro. *International Conference on Education and New Learning Technologies*, páginas 4194-4205. Julio de 2009, Barcelona (España).