



Sintetizador Paramétrico Multidispositivo de Lengua de Signos Española

por

D. Fernando Jesús López Colino

Tesis presentada para la obtención del
grado de Doctor

en el

Departamento de Ingeniería Informática
Escuela Politécnica Superior
Universidad Autónoma de Madrid

Director de la Tesis

Dr. José Colás Pasamontes

2009



«Cuando abrió el séptimo sello, se hizo silencio en el cielo...»
(Apocalipsis 8:1).

TÍTULO: Sintetizador Paramétrico Multidispositivo
de Lengua de Signos Española

AUTOR: D. Fernando Jesús López Colino

DIRECTOR: Dr. José Colás Pasamontes

El tribunal nombrado para juzgar la tesis doctoral antes citada, compuesto por
los doctores:

PRESIDENTE:

SECRETARIO:

VOCALES:

acuerda otorgarle la calificación de

Madrid, a de de 2009

Agradecimientos

En estos agradecimientos he optado por evitar todas esas frases típicas que suelen aparecer en este punto de las tesis doctorales. Por esta razón he decidido mencionar a todas las personas que me han apoyado, de una u otra manera, en este periodo. Creo que a todos ellos les he hecho llegar mi agradecimiento, de cualquier modo, gracias.

A Laura.

A Pilar y Fernando, también conocidos como Mamá y Papá.

A Pepe, como director de este trabajo.

A mi hermano Ramiro, mis abuelos Marta, Meli, Fernando Jesús y José y al resto de familia.

A la «Senior Staff» del HCTLab: Javier, Pepe, Ángel y Guillermo.

Al resto de compañeros del departamento: Alberto, Alejandro, Daniel, David, Fernando, Ismail, Javier, Jordi, Santiago y Susana.

A los profesionales de la FCNSE y de la academia SIGNAR, en particular a Almudena.

A todos los que han realizado experimentos.

A *Yuli* y *Daniel*, dado que sin su gran flexibilidad y tolerancia a mis fallos, no habría sido posible esta tesis.

Resumen

La síntesis de lengua de signos es un área de investigación con una antigüedad de apenas una década, periodo en el cual han surgido numerosas aproximaciones a este tema. La técnica más utilizada es la síntesis paramétrica, donde se parte de una descripción de los signos en función de los siete parámetros formativos quinésicos (PFQ) que los conforman. Aunque esta técnica no produce los resultados más naturales, si es la única que ofrece la flexibilidad y grado de control necesarios para gestionar la complejidad de la lengua de signos, razón por la que ha sido elegida para esta tesis.

Los sintetizadores paramétricos actuales utilizan una adaptación a XML de las notaciones iconográficas que describen los PFQ de los signos a sintetizar. Esta aproximación es equivalente a la síntesis de voz, en la que se utiliza una notación gráfica para representar los fonemas. El uso de esta aproximación presenta ciertos inconvenientes, tales como que la descripción de un mensaje a sintetizar es una tarea compleja o la falta de definición en la dimensión temporal de los signos: cada uno de los PFQ se definen como una secuencia de unidades en las que es necesario definir el instante en que se representan cada una de ellas; de igual modo, la síntesis de lengua de signos requiere de la definición de puntos de sincronización entre dichas secuencias.

En este trabajo afrontamos la síntesis paramétrica de lengua de signos desde una nueva perspectiva. Para obtener mayor precisión en la definición de los signos hemos optado por el uso de una base de datos relacional para su almacenamiento. Esta base de datos no sólo permite disponer de descripciones más precisas, detalladas y flexibles, sino que además permite definir representaciones paralelas, tanto de los propios signos como de los PFQ, gracias a los cuales se pueden contemplar en la misma base de datos la representación de los signos en distintos idiomas o dialectos y afrontar la síntesis de variantes emocionales para los signos. Una mayor flexibilidad de las descripciones de los signos, permite que los expertos en este modo de comunicación, especialmente signantes nativos, participen tanto en la evaluación como en la propia definición de los signos. Para ello se ha diseñado un conjunto de aplicaciones específicas para esta tarea. Dado que la notación de entrada ya no debe describir los signos mediante los PFQ, podemos definir una notación de entrada en un nivel superior de abstracción, lo que facilita la descripción de mensajes en lengua de signos. Esta nueva notación, que hemos denominado HLSML, contempla las tres unidades semánticas de la lengua de signos: diccionario dactilológico, signos establecidos y construcciones clasificatorias.

Junto con las novedades anteriores, este sintetizador presenta, además, dos características que lo diferencian de otras aproximaciones a la lengua de signos. La primera

se refiere a la arquitectura del sintetizador: la mayoría de los sintetizadores actuales se han planteado como aplicaciones de PC presentando determinados requisitos gráficos y computacionales que difícilmente pueden ser soportados por un dispositivo móvil; este trabajo aporta una arquitectura modular que permite distribuir el proceso de síntesis entre un servidor dedicado y el dispositivo del cliente, adaptándose a los recursos de éste. La segunda afecta al diseño del avatar: los avatares utilizados en otros sintetizadores utilizan una estructura esquelética estándar que únicamente contempla la animación del propio avatar; el diseño propuesto incluye un conjunto de huesos especiales que facilitan tanto la animación del avatar como la descripción de los signos, eliminando la dependencia de la definición de la malla del avatar.

Índice general

Acta	V
Agradecimientos	VII
Resumen	IX
Índice General	XI
Lista de Figuras	XVII
Lista de Tablas	XIX

I Introducción y Antecedentes

1. Introducción y motivación	3
2. Fundamentos teóricos	7
2.1. Parámetros formativos quinésicos	7
2.2. Clasificación de las unidades de las LS	10
2.2.1. Clasificadores o construcciones clasificatorias	12
2.2.1.1. Clasificadores nominales	12
2.2.1.2. Clasificadores introflexivos	13
2.2.1.3. Clasificadores icónicos	14
2.2.1.4. Clasificadores afijales	16
2.3. Notaciones para la descripción de los signos	18
2.3.1. Notación Stokoe	18
2.3.2. Notación SignWriting	19
2.3.3. Notación HamNoSys	20
2.3.4. Notación SEA	22
2.3.5. Notación Szczepankowski	23
2.3.6. Resumen de las notaciones	24
2.4. Conclusión	25
3. Estado del arte	27
3.1. Uso de la LS en los medios	27

3.1.1.	Uso de notaciones	27
3.1.2.	Videos de personas	28
3.1.2.1.	Síntesis de vídeo por concatenación de unidades	29
3.1.3.	Animación de personajes virtuales	30
3.1.3.1.	Técnicas de animación de personajes	31
3.1.3.2.	Alternativas en la síntesis basada en avatares	34
3.2.	Síntesis paramétrica de signos	35
3.2.1.	Avatares	36
3.2.1.1.	Avatares parciales	36
3.2.1.2.	Avatares completos	39
3.2.2.	Almacenamiento de signos	46
3.2.2.1.	Notaciones	46
3.2.2.2.	Bases de datos relacionales	48
3.3.	Síntesis de clasificadores	49
3.4.	Evaluación de mensajes sintéticos	51

II Descripción del Sintetizador Paramétrico

4.	Formalización	57
4.1.	Formalización de los PFQ	57
4.1.1.	Definición de los <i>signos reconocidos</i> y de la <i>dactilología</i>	66
4.2.	Formalización de los <i>clasificadores</i>	66
4.3.	Formalización de la síntesis emocional	69
4.4.	Conclusión	70
5.	Arquitectura del sintetizador	73
5.1.	Arquitectura del sintetizador	73
5.1.1.	Parser de HLSML	73
5.1.2.	Base de datos relacional	74
5.1.3.	Servidor web	75
5.1.4.	Síntesis de gestos	75
5.1.5.	Render	78
5.1.6.	Visualización	79
5.2.	Síntesis distribuida	79
5.3.	Resumen	82
6.	Avatar	83
6.1.	Esqueleto	83
6.1.1.	Descripción	84
6.1.2.	Nuevas clases de huesos	85
6.1.2.1.	Hueso de la muñeca	86
6.1.2.2.	Huesos direccionales	86
6.1.2.3.	Huesos para el PFQ <i>localización</i>	87

6.1.2.4.	Huesos para la detección de colisiones	88
6.2.	Malla	89
6.3.	Principales secciones del avatar	92
6.3.1.	Mano	92
6.3.2.	Cabeza y cara	94
6.3.3.	Brazos y cuerpo	96
6.4.	Avatares definidos	97
6.5.	Resumen	99
7.	Base de datos relacional	101
7.1.	Estructura	101
7.2.	Descripciones paralelas	105
7.2.1.	Variaciones de idioma y dialectales	106
7.2.2.	Variaciones emocionales	107
7.2.3.	Variaciones genéricas	108
7.3.	Resumen	109
8.	HLSML. Notación de entrada	111
8.1.	Descripción del HLSML	111
8.1.1.	Descripción de una secuencia de signos <i>establecidos</i>	112
8.1.2.	Descripción de un signo mediante sus PFQ	112
8.1.3.	Descripción del diccionario dactilológico	115
8.1.4.	Descripción de los <i>clasificadores</i> en HLSML	115
8.1.4.1.	Clasificadores nominales	116
8.1.4.2.	Clasificadores introflexivos	116
8.1.4.3.	Clasificadores icónicos	117
8.1.4.4.	Clasificadores afijales	119
8.1.5.	Definición de variaciones y elementos prosódicos	121
8.2.	Comparación de HLSML con SiGML y SWML	123
8.3.	Resumen	126
9.	Módulo de síntesis de gestos	127
9.1.	Proceso de síntesis y dependencias entre PFQ	127
9.2.	Cinemática inversa	129
9.2.1.	Definiciones	130
9.2.2.	Cálculo cinemático	132
9.2.2.1.	Cálculo del ángulo del codo	133
9.2.2.2.	Cálculo de la rotación del hombro	133
9.2.3.	Relaciones	133
9.2.3.1.	Hombro: rango válido y rango óptimo	134
9.2.3.2.	Muñeca	135
9.2.3.3.	Fórmula final	137
9.3.	Resolución de colisiones	137
9.3.1.	Trabajo relacionado	137

9.3.2. Volumen de colisión	138
9.3.3. Algoritmo aplicado	139
9.3.4. Resolución de colisiones durante la cinemática inversa	141
9.4. Resumen	142

III Experimentación y Evaluación

10. Evaluaciones objetivas	145
10.1. Evaluación de la animación en tiempo real	145
10.1.1. Tasa de refresco según el número del polígonos de la malla	146
10.1.2. Tasa de refresco según el tamaño de la ventana	146
10.2. Tiempos de ejecución de cada sección del proceso de síntesis	148
11. Evaluaciones subjetivas	151
11.1. Capacidad expresiva y comunicativa de la cara del avatar	151
11.1.1. Comprensión de la lectura de labios	152
11.1.2. Naturalidad en el movimiento de labios	153
11.1.3. Expresiones faciales	154
11.2. Representación de la tensión	155
11.2.1. Tensión en la <i>configuración</i>	155
11.2.2. Tensión en el hombro	157
11.3. Evaluación signos aislados	158
11.3.1. Respuesta abierta	159
11.3.1.1. Resultados del avatar <i>Yuli</i>	159
11.3.1.2. Resultados del avatar <i>Daniel</i>	160
11.3.2. Respuesta cerrada	161
11.4. Evaluación frases	162
11.4.1. Frases compuestas por secuencias de signos <i>establecidos</i>	163
11.4.2. Clasificadores	164
11.4.3. Variaciones emocionales	164
11.5. Informe de la Fundación CNSE	167
11.6. Resumen de los resultados	168

IV Herramientas y Aplicaciones Desarrolladas

12. Aplicaciones para la generación de contenido	173
12.1. Contenido de niveles tercero y cuarto	173
12.2. Contenido de niveles primero y segundo	175
13. Herramientas auxiliares	179
13.1. Aplicación IK y control de colisiones	179
13.2. Definición de transiciones en el PFQ <i>orientación</i>	180

14. Aplicaciones de usuario	183
14.1. Aplicación para PC	183
14.2. Aplicación para PocketPC	185
14.3. Visualización en Nintendo DS Lite	185

V Conclusiones y Trabajo Futuro

15. Conclusiones	189
15.1. Resumen de aportaciones	189
15.1.1. Diseño modular del sintetizador	189
15.1.2. Estructura del avatar	190
15.1.3. Base de datos relacional	190
15.1.4. Notación de entrada: HLSML	191
15.1.5. Síntesis de emociones	191
15.1.6. Herramientas desarrolladas	191
15.1.7. Resultados obtenidos	192
15.2. Publicaciones	192
15.2.1. Publicaciones nacionales	192
15.2.2. Publicaciones internacionales	193
16. Trabajo futuro	195

Apéndices

A. Glosario de abreviaturas	201
B. Quaternions	203
C. Diagrama detallado de la base de datos relacional	205
D. Jerarquía de los elementos definidos en HLSML	209
E. Informe de la Fundación CNSE	211
Bibliografía	231

Índice de figuras

2.1.	Algunas <i>configuraciones</i> utilizadas en la LSE.	9
2.2.	Diccionario dactilológico en LSE.	11
2.3.	Ejemplo de <i>signos reconocidos o establecidos</i>	12
2.4.	Ejemplo de clasificador afijal.	17
2.5.	Ejemplos de la notación Stokoe.	19
2.6.	Ejemplos de la notación SignWriting.	19
2.7.	Ejemplos de la notación HamNoSys.	20
2.8.	Estructura de SEA.	22
2.9.	Ejemplos de la notación SEA.	23
2.10.	Ejemplos de la notación Szczepankowski.	24
3.1.	LSE en un fichero GIF.	28
3.2.	Fotogramas de una película animada de LSE.	31
3.3.	Dispositivo de captura.	32
3.4.	TESSA y VISIA. Avatares animados por captura de movimientos.	35
3.5.	Arquitectura básica de un sintetizador paramétrico.	36
3.6.	Definición unificada de esqueleto y malla.	39
3.7.	Animación esquelética.	40
3.8.	Estructuras esqueléticas utilizadas para la animación de avatares.	42
3.9.	Distintos avatares utilizados para la síntesis de signos.	45
3.10.	Simulación de una escena escalada al espacio de signado.	50
4.1.	Variación emocional en el PFQ <i>configuración</i>	59
4.2.	Ejemplo de desvinculación jerárquica.	60
4.3.	Sistema de coordenadas de la escena tridimensional.	61
4.4.	Distintas variaciones de un movimiento rectilíneo.	63
4.5.	Variaciones emocionales de un movimiento.	64
4.6.	Definición del espacio de signado.	68
5.1.	Arquitectura del Sintetizador de LSE propuesto.	74
5.2.	Detalle del módulo de síntesis de gestos.	76
5.3.	Reparto de tareas entre cliente y servidor.	80
5.4.	Distribución de cargas de proceso en las cuatro alternativas.	81
6.1.	Comparación del diseño esquelético aplicado a las manos del avatar.	87
6.2.	Malla y Esqueleto del Avatar.	91

6.3.	Diagrama de los huesos que componen la mano.	93
6.4.	Diagrama de las partes de la cara deformables.	95
6.5.	Secciones dependientes de cada hueso del contorno de la boca.	96
6.6.	Comparación de los esqueletos de los avatares.	98
6.7.	Comparación de las mallas de los avatares.	99
7.1.	Diagrama simplificado de la base de datos.	102
7.2.	Diagrama detallado de la base de datos.	106
7.3.	Definición de variaciones dialécticas.	107
7.4.	Definición de variaciones emocionales.	108
8.1.	Ejemplo HLSML: Secuencia de signos <i>establecidos</i>	113
8.2.	Ejemplo HLSML: Definición dinámica de un signo.	114
8.3.	Ejemplo HLSML: Deletreo.	115
8.4.	Ejemplo HLSML: Clasificador Nominal.	117
8.5.	Ejemplo HLSML: Clasificador Introflexivo.	118
8.6.	Ejemplo HLSML: Clasificador Icónico.	120
8.7.	Ejemplo HLSML: Clasificador Afijal.	122
9.1.	Diagrama de la estructura esquelética del brazo.	131
9.2.	Gaussianas de tensión del hombro.	135
9.3.	Tensión del hombro en función de Φ y S_2	136
9.4.	Poliedros de colisión.	139
9.5.	Vector de tendencia.	140
9.6.	Vector de salida.	141
10.1.	Evaluación de la tasa de refresco según los polígonos del avatar.	147
10.2.	Evaluación de la tasa de refresco según el tamaño de la ventana.	148
11.1.	Algunas de las expresiones utilizadas en la evaluación 11.1.3.	154
11.2.	Imagen utilizada en la pregunta 6 de la evaluación 11.2.1.	157
11.3.	Fotogramas de los vídeos utilizados en la evaluación 11.2.2.	158
12.1.	Aplicación de captura para los PFQ.	174
12.2.	Aplicación de generación de contenido: Vista general.	176
12.3.	Aplicación de generación de contenido: Definición de PFQ.	178
13.1.	Aplicación auxiliar: Desarrollo de IK.	180
13.2.	Aplicación auxiliar: Transición de Orientaciones.	181
14.1.	Aplicación de síntesis para PC.	184
14.2.	Aplicación para PDA.	185
14.3.	Visualización de la síntesis en una Nintendo DS Lite.	186
B.1.	Resultado de la interpolación de quaternions.	204
C.1.	Diagrama detallado de la base de datos relacional.	206

Índice de tablas

2.1.	Acciones construidas durante la representación de un diálogo.	15
2.2.	Combinaciones de las unidades expresivas en HamNoSys.	21
2.3.	Resumen de las distintas notaciones para la descripción de los signos.	25
3.1.	Principales zonas de atención cuando se observa un mensaje signado.	32
3.2.	Relación entre los fonemas y los visemas del español.	38
5.1.	Escenarios distribuidos: Detalle de las cargas de proceso.	82
6.1.	Desglose de los huesos utilizados en el avatar por ubicación.	84
6.2.	Desglose de los huesos utilizados en el avatar según función.	89
6.3.	Descripción de la malla del avatar <i>Daniel</i>	90
6.4.	Comparación entre los esqueletos de los avatares <i>Yuli</i> y <i>Daniel</i>	97
6.5.	Comparación entre las mallas de los avatares <i>Yuli</i> y <i>Daniel</i>	97
10.1.	Descripción del ordenador utilizado para la evaluación 10.1.1.	146
10.2.	Descripción de los ordenadores utilizados para la evaluación 10.1.2.	146
10.3.	Tiempo medio de síntesis en distintas unidades de la LSE.	149
11.1.	Tasa de acierto en el reconocimiento del movimiento labial.	152
11.2.	Resultados de calidad para el movimiento labial <i>con</i> sonido.	153
11.3.	Resultados de calidad para el movimiento labial <i>sin</i> sonido.	154
11.4.	Porcentaje de respuestas para la evaluación de expresiones faciales.	156
11.5.	Resultados de la evaluación de tensión en la mano.	157
11.6.	Resultado de la evaluación de tensión en el brazo.	159
11.7.	Resultados test de respuesta abierta con <i>Yuli</i>	160
11.8.	Resultados test de respuesta abierta con <i>Daniel</i>	161
11.9.	Resultados test de respuesta cerrada con el <i>Daniel</i>	162
11.10.	Resultados del test de frases en LSE.	163
11.11.	Resultados de los test de los clasificadores.	165
11.12.	Variaciones temporales en la síntesis emocional.	166
11.13.	Resultados test de síntesis emocional para la frase 1.	166
11.14.	Resultados test de síntesis emocional para la frase 2.	166
11.15.	Resumen de las evaluaciones subjetivas.	169
11.16.	Resumen de las evaluaciones sobre síntesis de signos.	170

Sintetizador Paramétrico Multidispositivo de LSE

12.1.	Tiempo medio de definición de un signo.	176
13.1.	Transiciones entre orientaciones.	181
E.1.	Correspondencia de los tests y las secciones de este documento. . . .	211

Capítulo I

Introducción y Antecedentes

Apartado 1

Introducción y motivación

La legislación española reconoce la Lengua de Signos Española (LSE) y establece el derecho de las personas sordas al acceso a la información según sus preferencias, incluida la LSE [38, 39]. Como consecuencia de esta normativa se establece, entre otros, la presencia de un intérprete en los actos públicos, subtítulo en LSE y que los contenidos de todos los documentos oficiales sean accesibles en LSE. Sin embargo, actualmente el único espacio televisivo subtítulo en LSE es el debate sobre el estado de la nación, además de los programas especialmente creados para personas sordas, cuyos presentadores utilizan LSE. Tampoco las páginas web pertenecientes a diversas instituciones del estado español cumplen esta normativa¹. La causa de esta falta de adaptación se debe a que el único sistema que permite generar contenido signado con la suficiente calidad para ser aceptado por la comunidad sorda está basado en vídeos de personas, presentando además el inconveniente de un elevado coste económico de producción y falta de flexibilidad. Supongamos que se genera un vídeo por cada una de las webs oficiales del estado, comunidades, ayuntamientos, etc.; además de los nuevos contenidos generados cada día, cualquier modificación realizada sobre alguna de estas páginas, implicaría una nueva grabación en vídeo, a lo que se añade la problemática relacionada con la gestión, almacenamiento y transmisión de contenidos en vídeo. La solución ideal para este problema implica el uso de un módulo de traducción entre la lengua oral y la lengua de signos, junto con un sintetizador. Si no se parte de texto, es necesario incluir, además, un reconocedor de voz. El desarrollo de un módulo de traducción se ve afectado por la ausencia de una teoría que recoja en su totalidad la casuística de la lengua de signos. Una posibilidad para afrontar la

¹Aunque las personas sordas pueden leer, su grado de comprensión no es equivalente a una persona oyente. Esto es debido a que el texto en español no sigue la estructura gramatical, ni organiza la información del mismo modo que la LSE. El efecto es similar a tomar una oración en español y escribirla con palabras inglesas, un inglés reconocería las palabras pero no entendería el significado.

traducción automática, reduciendo la dependencia de esta formalización, se basaría en el uso de técnicas de aprendizaje automático, que tienen el inconveniente de que requieren gran cantidad de información recogida en ambas lenguas, material que no está disponible. Junto a estos problemas, que pueden darse inclusive al afrontar la traducción entre dos lenguas orales, hay que añadir que la lengua de signos es una lengua visual, por lo que, no puede aplicarse exactamente el mismo paradigma que en la traducción entre lenguas orales. Los sistemas de traducción actuales sólo presentan soluciones parciales, restringiendo el proceso a dominios muy concretos.

La síntesis de lengua de signos es un proceso mediante el cual se genera un contenido visual que representa un determinado mensaje siguiendo las reglas de dicha lengua de signos. Se puede considerar que la síntesis de signos es una parte específica de la síntesis de gestos. No todos los gestos son necesariamente signos; un signo es un gesto que representa un concepto y que es comúnmente aceptado en una determinada comunidad de personas sordas. La síntesis automática de lengua de signos implica que la generación de dicho contenido visual se realice sin la intervención directa de una persona. Este campo de investigación ha experimentado un gran auge en la última década, siendo prácticamente inexistente con anterioridad.

Afrontar la síntesis automática de lengua de signos requiere el manejo de disciplinas tan dispares como son la animación por ordenador y el procesamiento del lenguaje natural. A la dificultad de tener que afrontar dos disciplinas tan diferentes como las anteriores habría que añadir que los estudios lingüísticos sobre la lengua de signos no tienen la solidez presente en los estudios de las lenguas orales, encontrando teorías dispares para definir la morfología, sintaxis y gramática de las lenguas de signos, lo que complica la realización de una formalización de la síntesis de signos, requisito previo en un proceso de automatización. Debido a estas dificultades, puede justificarse el hecho de que hayan sido muy pocos los grupos que hayan afrontado en profundidad este proceso.

Dado que, el proceso completo no está disponible debido a la dificultad de realizar una traducción automática, valoremos una solución intermedia. Supongamos que se realiza una traducción manual más una síntesis automática, el proceso sería más barato y flexible que el uso de vídeo, sin embargo, esta solución, que permitiría generar contenidos en LSE, tampoco es utilizada. La razón que hemos encontrado para ello es que los sintetizadores se han desarrollado principalmente como módulos en sistemas completos utilizando una descripción paramétrica de los signos y, por lo tanto, la definición del mensaje a sintetizar no es trivial. Además, estos sistemas no permiten definir modificaciones en la generación de los signos, variaciones prosódicas ni mejoras

extraídas de las opiniones manifestadas por personas sordas, quienes realmente son a los que va dirigida la síntesis.

Otro obstáculo a superar es el rechazo entre la comunidad sorda a los avatares no ya por el hecho de que sea un personaje digital el que presente el mensaje, ya que hemos observado que vídeos de avatares animados manualmente tienen mayor grado de aceptación, sino porque el modo de animación es diferente en ambos casos: la animación paramétrica genera un resultado con un aspecto demasiado artificial, lo que penaliza la percepción que los usuarios tienen de la síntesis.

Por último, los sistemas de síntesis actuales se han diseñado como aplicaciones para PC. La actual tendencia de la sociedad se decanta por el uso de dispositivos portátiles, ya sean netbooks, UMPC o PDA. Estos dispositivos permiten acceso a la información en cualquier lugar, pero la mayoría no disponen de los recursos gráficos y de procesamiento necesarios para gestionar todo el proceso de síntesis de lengua de signos. Este hecho impide el acceso a la información presentada como mensajes signados. La gran diversidad de dispositivos complica la posible adaptación de los sistemas de síntesis actuales a muchos de estos dispositivos.

En resumen, el siguiente paso en la síntesis de signos y motivación de esta tesis, consiste en subsanar las deficiencias existentes en los sistemas actuales.

En este trabajo proponemos un sintetizador basado en un planteamiento mixto: la descripción de los signos es paramétrica pero se realiza de forma manual. Para que ajustes tales como la definición de la evolución temporal de los signos (no permitida en las notaciones actuales) puedan ser realizadas por expertos en LSE. Dado que estos expertos no tienen generalmente grandes conocimientos informáticos, se ha diseñado un conjunto de aplicaciones que permiten realizar estas tareas a nivel de usuario básico. Por otra parte, se ha trabajado en la definición de una notación de entrada para describir de forma sencilla mensajes en LSE, siendo ésta una de las carencias de los sistemas actuales. Además, la arquitectura del sintetizador se ha desarrollado teniendo en cuenta la adaptación del proceso de síntesis a dispositivos con distinta capacidad de proceso, lo que facilitará disponer de mensajes sintéticos en lengua de signos en dispositivos móviles.

Otros aspectos que se han abordado en esta tesis han sido el desarrollo de un avatar con características específicas para la síntesis de signos, la definición de las modificaciones necesarias para generar variaciones prosódicas y la representación de emociones en el mensaje signado.

Apartado 2

Fundamentos teóricos

A lo largo del periodo de elaboración de esta tesis, ha quedado de manifiesto la gran variedad de aproximaciones con las que se aborda el estudio lingüístico de las lenguas de signos (LS). El estudio de los distintos niveles de las LS ha sido abordado desde varios puntos de vista dando lugar a diferentes teorías para explicar la morfología de los signos, las unidades sintácticas e incluso la manera de representar los signos. Parece necesario realizar una revisión a la componente teórica de la LSE para ofrecer al lector una visión general de la problemática referente a la creación de un sintetizador de signos.

En primer lugar se discutirá la estructura interna de un signo mediante la descripción de los parámetros que conforman los signos gestuales. A continuación se enumerarán las distintas unidades presentes en el discurso signado. Dicha clasificación se realizará atendiendo a las propiedades sintáctico-semánticas de estas unidades. Por último se repasará brevemente las principales notaciones existentes para la descripción de los signos.

2.1. Parámetros formativos quinésicos

Se denominan Parámetros Formativos Quinésicos (PFQ) a los elementos articulatorios que configuran el signo gestual. La literatura hace referencia en ocasiones a estos parámetros utilizando el término «fonemas», pero este uso es incorrecto y proviene de las aproximaciones al estudio de las lenguas de signos desde la perspectiva del estudio de las lenguas orales. La evolución en el estudio de la lengua de signos permite diferenciar tres etapas en el nivel de detalle con la que se ha abordado el estudio de las LS.

La primera de ellas es la etapa previa a cualquier estudio lingüístico sobre las LS. Los signos eran considerados como una unidad indivisible y no se estudiaban elementos básicos que los formaran. Es interesante resaltar el trabajo de Fernández Villalibre [34], quien presenta en 1851 el primer diccionario de signos para la LSE. En dicha obra simplemente se realiza una descripción sobre la manera de generar los signos sin atender en ningún momento a una definición formal de su estructura.

La siguiente etapa comienza en 1960 tras los estudios de William Stokoe. En su obra *Sign Language Structure* realiza el primer estudio de la LS [97]. Los PFQ que se identifican en esta etapa son el lugar de articulación del signo, la configuración de la mano y el movimiento de las manos en el desarrollo del signo. Posteriormente, en 1973, Battison [9] introduce un nuevo parámetro, la orientación de la mano.

Por último, la tercera etapa a partir de 1990, en la que nos encontramos en la actualidad. Aquí, el número de PFQ se amplía hasta siete, añadiendo el plano de articulación, el punto de contacto y los componentes no manuales. Así mismo, se estudian las relaciones de simultaneidad y secuencialidad de los PFQ que componen la lengua de signos. A esta etapa pertenecen los estudios realizados por Muñoz [80, 81].

Por lo tanto, si deseamos afrontar la síntesis de cualquier lengua de signos, se deberá tener en cuenta estos siete parámetros:

- *Queirema, forma o configuración* de la mano o manos que intervienen en la articulación del signo (Figura 2.1). Este parámetro está definido por la flexión de las articulaciones de los dedos. Cada LS posee sus propias *configuraciones*, constituyendo un subconjunto de entre todas las posibles. En el estudio propuesto por Muñoz [80] se presentan setenta y una *configuraciones* distintas en la LSE. Este número ha resultado ser insuficiente, ya que durante dicho estudio se omitieron un conjunto de *configuraciones* que pueden ser consideradas parte de la gesticulación no lingüística, como son las formas de la mano relativas a la representación de los números. Además, dicho estudio no incluye ciertas *configuraciones* definidas en el diccionario dactilológico internacional.

Se observa que en las LS ciertas *configuraciones* sólo aparecen al comienzo de los signos y otras sólo al final. También se definen pares de *configuraciones* que se ejecutan de manera consecutiva. Además el número de *configuraciones* utilizadas en la mano dominante es muy superior al utilizado en la mano base o no dominante. Como veremos posteriormente, estos últimos factores no son relevantes para la síntesis de la LSE.

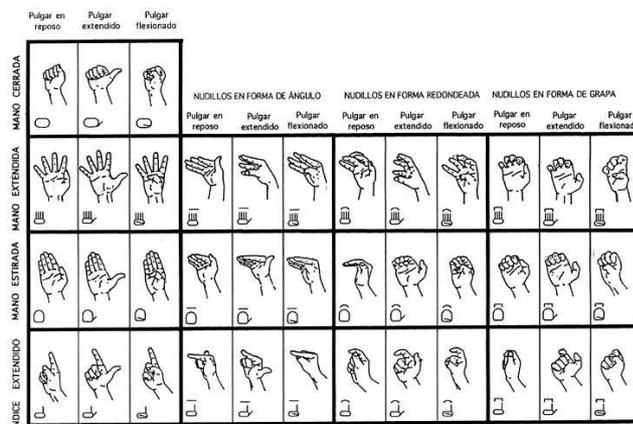


Figura 2.1: Algunas configuraciones utilizadas en la LSE.

- *Queirotoponema u orientación* de la mano o manos que intervienen en la articulación de los signos. Siempre tomando como referencia el cuerpo del signante. Este es un parámetro que no ha sido estudiado con detalle en la LSE. Las rotaciones de muñeca se realizan en los tres ejes del espacio, en ángulos de aproximadamente 45° , generando un total de 208 posibles orientaciones.
- *Toponema, localización o lugar de articulación* referido a la posición donde se articulan los signos. La definición de este parámetro utiliza un punto anatómico del cuerpo del signante como referencia para establecer la posición donde se articula el signo, aunque no es preciso que haya contacto físico. El espacio en el que se articula el signo se encuentra frente al cuerpo del signante y ocupa aproximadamente desde la cintura hasta la parte superior de la cabeza, siendo escasos los signos que se salen de este espacio. Es importante señalar que un cambio en la localización, incluso entre zonas no muy distantes, implica un cambio de significado importante. Podemos definir cuatro áreas principales en la que se encuentran estos puntos de referencia: cabeza y cuello, tronco, brazo no dominante y mano no dominante.
- *Esquedema o plano*, hace referencia a la distancia del punto de articulación al cuerpo del signante. Por lo tanto, tiene relación directa con la extensión de los brazos al articular los signos. Se consideran cuatro planos principales: el primero constituye el plano que ocupa el cuerpo del signante; el segundo se sitúa aproximadamente a la distancia que se encuentra el codo; el tercero hacia la mitad del antebrazo y el cuarto, el que se alcanza con el brazo completamente extendido.

- *Haptonema o punto de contacto* entre la mano dominante del intérprete y otras partes del cuerpo. No aparece en todos los signos y tiene gran variación entre distintas LS. Este PFQ también interviene en la definición espacial de la posición de las manos en la ejecución del signo.
- *Kinema o movimiento* de las manos del signante en la realización de los signos. Este parámetro es el más complejo de describir, debido a que es necesario representar tanto la trayectoria que toman las manos como la velocidad y aceleración del movimiento. Normalmente se define la trayectoria de manera aproximada así como algunos valores preestablecidos para la velocidad y la tensión muscular.
- *Componente no manual*. Este PFQ hace referencia a la expresión facial, movimientos de cabeza, hombros y tronco. Encontramos signos en los que aparece más de un componente no manual, tanto de forma simultánea como de forma secuencial. La utilización del componente no manual aporta variaciones al significado del signo, además de naturalidad y expresividad al mensaje.

2.2. Clasificación de las unidades de las LS

Las unidades que componen las LS pueden ser clasificadas en tres grupos principales atendiendo a un análisis semántico de la información que representa esa unidad. Esta segregación deja de manifiesto, para la mayoría de los casos, una relación directa entre el contenido semántico y la complejidad en la descripción del signo. Estos tres grupos, enumerados de menor a mayor complejidad son:

- El *diccionario dactilológico*. Este elemento dentro de las LS es la representación del alfabeto de las lenguas orales mediante signos gestuales monomanuales. La generación de cada una de las letras implica la definición de una *configuración* y una *orientación*, recibiendo los PFQ *localización*, *plano* y *punto de contacto*, valores preestablecidos y comunes a todas las letras. Algunas de las letras requieren la definición de *movimiento* o variaciones en la *orientación* de la mano (Figura 2.2).

La ejecución de dicho diccionario se realiza con la mano principal. Durante la ejecución de una secuencia de letras utilizando este diccionario influye, en gran medida, la interacción con el interlocutor que puede dar lugar a una interrupción en la secuencia de letras, si el concepto que se está deletreando ha sido entendido por el otro interlocutor.

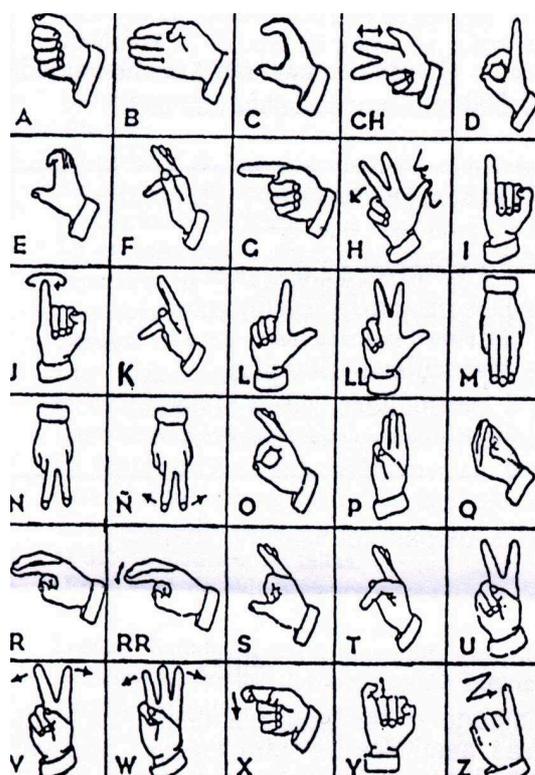


Figura 2.2: Diccionario dactilológico en LSE.

- Los *signos establecidos*. Estos elementos poseen tanto una representación como un significado determinado y aceptados entre la comunidad sorda que los utiliza. Cada uno de estos elementos hace referencia a un concepto o idea dentro del mensaje signado. Aunque el proceso por el que un nuevo signo se establece como signo reconocido suele ser progresivo, las personas sordas son capaces de generar signos nuevos dentro de una conversación para hacer referencia a personas o conceptos que no tienen un signo relacionado.

Como se ha mencionado, la ejecución del signo relacionado a un concepto sólo se mantiene constante dentro de un núcleo de personas sordas. Es por esto que los signos asociados a un concepto son diferentes entre distintos países, e incluso entre distintas regiones del mismo país. En este trabajo doctoral se ha optado por utilizar las representaciones recogidas en el diccionario editado por la Fundación CNSE [36], ya que refleja un esfuerzo unificador de toda la LSE (Figura 2.3).

- Los *clasificadores*. Este grupo recoge construcciones semánticas de gran complejidad en el discurso signado. Estos elementos se utilizan para añadir significado y riqueza a la frase signada. Su ejecución se nutre de la capacidad de descripción inherente a la lengua de signos. Dada la complejidad de estas unidades, hemos

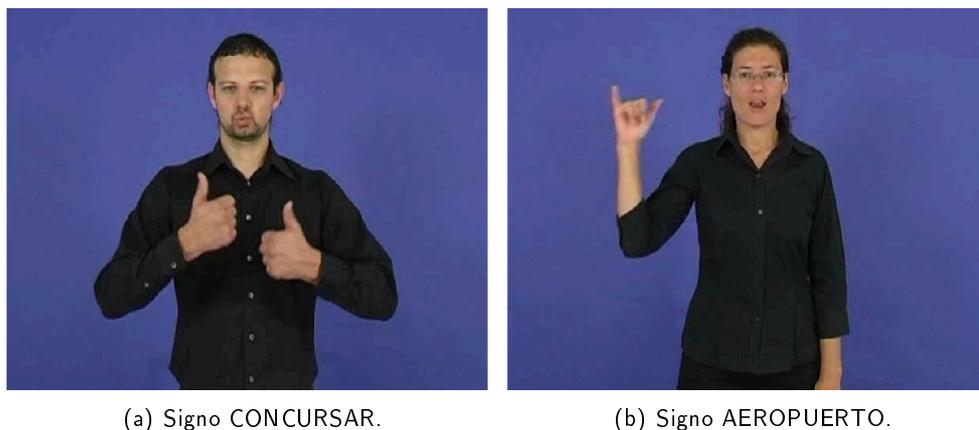


Figura 2.3: Ejemplo de *signos reconocidos o establecidos* pertenecientes al diccionario normativo [36].

reservado la Subsección 2.2.1 para un estudio más detallado del fenómeno de los *clasificadores*.

2.2.1. Clasificadores o construcciones clasificatorias

El trabajo de Herrero Blanco [46] ha sido clave para el planteamiento seguido en esta tesis en lo referente a los clasificadores. La gran importancia de este trabajo es que centra su estudio en la LSE frente a otros autores que han estudiado otras LS y por lo tanto es el trabajo más relevante para esta tesis doctoral. Además, el análisis de los *clasificadores* se aborda desde una perspectiva morfológica, más cercana a la aproximación que debe ser utilizada para la creación de un sintetizador. Herrero Blanco presenta en su trabajo los siguientes tipos de clasificadores:

2.2.1.1. Clasificadores nominales

Esta categoría de clasificadores, también denominada como «nombres clasificatorios», describe la capacidad manifiesta de la LSE para generar nuevos significados o conceptos, utilizando una pareja de signos. Así, por ejemplo, las construcciones AGUA CAUCE (río) y AGUA POLEA (pozo), donde el signo AGUA se comporta como nombre clasificatorio. Este signo puede aparecer en una construcción de manera aislada aportando su significado, como por ejemplo YO AGUA BEBER. Además, se debe tener en cuenta que un nombre clasificatorio puede aparecer acompañando a otros signos, sin que la fusión de los significados genere un nuevo concepto lógico (AGUA MONTAÑA no se traduce como *montaña de agua* sino como *agua de la montaña*).

Una de las mayores complicaciones presentadas por Herrero Blanco es la manera que se tiene de poder diferenciar los nombres clasificatorios de otras construcciones morfológicamente iguales, pero que no se deben considerar como tales. Éste es el caso de los nombres compuestos (AGUA TIERRA significa «barro») o construcciones derivativas que permiten crear adjetivos («beato» es PERSONA PIEDAD), gentilicios (PERSONA MADRID) o nombres a partir de verbos (el término «anfitrión» se forma como PERSONA INVITAR).

Como se ha manifestado, cualquier análisis que se lleve a cabo debe ser considerado únicamente desde el punto de vista de un sintetizador y, por tanto, el factor relevante del mismo es la morfología de estos elementos para poder realizar la síntesis. Que la construcción sea un nombre clasificatorio, un nombre con un núcleo, un nombre compuesto o un derivativo resulta irrelevante para la síntesis. Todas estas construcciones se abstraen en una construcción de la forma «signo + signo», donde ambos signos pertenecen a la categoría de *signos reconocidos o establecidos*.

2.2.1.2. Clasificadores introflexivos

Los clasificadores introflexivos recogen la capacidad que tiene la lengua de signos para matizar o añadir significado en un signo mediante la modificación de alguno de los PFQ que lo componen. Varios de los autores que han estudiado este proceso en las LS han encontrado analogías en distintas lenguas orales. En estos casos, ciertos fonemas de la palabra son sustituidos por otros distintos o pronunciados de otra manera variando el significado de dichos términos.

Según Supalla [100] este hecho se pone de manifiesto en variaciones del PFQ *configuración* durante la ejecución de determinados signos. Tal es el caso del signo LIBRO en el que la *configuración* con la que se realice dicho signo permite describir su grosor. De igual manera cuando se aplica la configuración del signo VOLCADO al signo TELEVISIÓN se añade al concepto de televisión una descripción sobre su situación espacial. Esta característica de la LSE sólo se aplica si se cumple la siguiente restricción semántica: tanto el signo clasificador como el clasificado deben ser propiedades tangibles, como por ejemplo la que hace referencia a la forma de un libro, pero no al contenido de dicho libro.

De igual manera se puede describir a los clasificadores numerales. En la LSE existe una variedad del signo HORA que se signa sobre la muñeca izquierda (que simularía llevar el reloj) con un movimiento circular de la muñeca que emula el movimiento de las manillas del reloj. La mano derecha utiliza la *configuración* de índice extendido

(equivalente a signar «1»). Se puede observar que para indicar un determinado número de horas, en vez de generar una construcción del tipo («número» + HORA) la ejecución del signo HORA se altera modificando la configuración de la mano dominante. Según el número de horas que se desee indicar, la mano dominante toma esta configuración en la ejecución de HORA. Si bien ciertos autores no consideran esta construcción como un clasificador [46], morfológicamente son idénticas.

Muñoz [80] presenta variaciones flexivas del *movimiento* en la ejecución de un signo para generar la variación reflexiva del verbo. Tal es el caso del signo DAR que se ejecuta con un movimiento horizontal alejándose del cuerpo del signante. Cuando dicho movimiento se dirige hacia el cuerpo del propio signante, se genera la idea de «darme». Posteriormente Herrero Blanco [46] formaliza esta modificación en los predicados de movimiento.

La conclusión que aporta el estudio de los clasificadores introflexivos es la capacidad y posibilidad que se demuestra en las LS para modificar alguno de los PFQ en la ejecución de un signo para añadir un mayor significado. El gran número de combinaciones posibles descarta cualquier aproximación basada en su descripción manual. Por tanto se trata de una alteración sobre la representación normal de un signo que debe ser definida en la traducción y contemplada por el sintetizador para realizar las modificaciones pertinentes en dicha representación.

2.2.1.3. Clasificadores icónicos

Esta categoría recoge la capacidad de representación de un mensaje a través de la mímica. El signante es capaz de utilizar el espacio de signado tanto para representar distintas características de un objeto, como para crear un pequeño escenario en el que situar distintos elementos de la conversación e interactuar con ellos.

La literatura se refiere a este tipo de clasificadores como «clasificadores predicativos» o simplemente «clasificadores» [46]. Liddell [68] recurre a la teoría de espacios mentales para la definición de estos clasificadores. La descripción mímica que está haciendo el signante sólo adquiere sentido cuando se establece una relación con la realidad descrita, en lo que Liddell denomina un «espacio mental fusionado». El signante utiliza sus manos para completar descripciones de objetos y situarlos en el espacio de signado. El planteamiento de Liddell [69] defiende las características lingüísticas de este tipo de clasificadores, dado que aunque habla de un proceso de flexibilización de los PFQ sigue considerándolos como tal (morfemas).

Cogill-Koez [20] plantea que estos «clasificadores predicativos» pueden responder a un conjunto de plantillas preestablecidas. Durante su disertación, define cómo para cada uno de los PFQ cómo puede adaptarse a estas plantillas prefijadas. El parámetro *configuración* parece ser el más regulado, y como se verá en el trabajo de Huenerfauth [48] es un hecho bastante aceptado. Según esto, ciertas *configuraciones* se relacionan de manera constante con la representación de ciertos elementos, como pueden ser personas, animales u objetos. El parámetro que no consigue establecerse dentro de una plantilla preestablecida es el parámetro de *localización*. Esto adquiere sentido si se aplica la teoría de Liddell, ya que dicha construcción responde a la representación del espacio mental del signante. Este hecho hace que Cogill-Koez se planteen la naturaleza de estos clasificadores como entidades lingüísticas o no.

Dentro de esta representación se incluye el conjunto de modificaciones que es capaz de incluir el signante en el discurso para emular un diálogo entre dos o más personas. Lillo-Martin [71] defiende que el signante (mediante movimientos de los hombros y la cabeza, cambiando la dirección en la que mira y/o variando la expresión de la cara) tenderá a interpretar al personaje activo dentro de la presentación. Aunque según rebaten Liddell y Metzger [70], esta acción no suele estar presente a lo largo de toda una representación, por lo que no debe ser considerada como el único elemento en la definición de papeles. De manera paralela, Engberg-Pedersen [30] añade, a la variación en la expresión del signante para cambiar de papel, un cambio de referencias. Este cambio se aplica, tanto al espacio de signado donde ha colocado de manera imaginaria objetos, como a la realización de pronombres (en este caso, un pronombre de primera persona no hace referencia al narrador).

Por último resaltar la aportación que hacen Liddell y Metzger planteando el concepto de «acción construida» definido como la capacidad que tiene el narrador (signante) para representar la actitud, posición y acciones de los personajes sobre los que está narrando, ejecutando dichas acciones en el mundo real. En particular, la acción más frecuente se basa en la dirección en la que los ojos y la cabeza apuntan. La Tabla 2.1 muestra algunos ejemplos de acciones que pueden aparecer de manera independiente unas de otras.

Tipo de acción	Lo que indica
Articulación de signos.	Lo que el personaje dice o piensa.
Dirección de la cabeza o los ojos.	Dirección en la que mira el personaje.
Expresión facial del signante.	Lo que el personaje siente.
Gestos de las manos y los brazos.	Gestos producidos por el personaje.

Tabla 2.1: Acciones construidas durante la representación de un diálogo.

Independientemente de si los clasificadores icónicos son gestos lingüísticos o no, su morfología difiere en cierta medida de los *signos establecidos*, lo que es relevante para la creación de un sintetizador de lengua de signos. Si bien PFQ como la *configuración* o la *orientación* no requieren una aproximación diferente, la posición de las manos (PFQ *localización* y *plano*) requiere el uso de un sistema de referencia continuo e independiente de referencias anatómicas. De igual manera, los *movimientos* de las manos son capaces de mimetizar movimientos de objetos en el espacio real. Es por esto, por lo que las trayectorias de las manos deben poder ser definidas arbitrariamente (ej: La descripción del trayecto de una montaña rusa).

La definición de expresiones faciales es un elemento incluido dentro del PFQ *no manual* de los *signos establecidos*, por lo tanto no implica una modificación adicional sobre las consideraciones necesarias para la definición de este tipo de elementos de una LS. Por el contrario, muchos autores coinciden en que el punto al que se orienta la vista o la cabeza del signante es muy importante. Entre todos, se debe destacar el «análisis de aceptación» planteado por Baham [7] por el cual el movimiento de la cabeza del signante genera una asociación entre el verbo, el signante (o personaje representado) y el objeto sobre el que recae la atención, y que se convierte en el complemento directo del verbo. Por consiguiente es necesario establecer un método para definir el punto hacia el cual el avatar dirige la mirada y la cabeza.

2.2.1.4. Clasificadores afijales

Como manifiesta Herrero Blanco [46], “Una de las características fonológicas de las LSs es que cuentan con dos articuladores manuales, las dos manos”. Esta característica permite generar construcciones en las que hay dos elementos presentes, representados por cada una de las manos. Así por ejemplo, la mano pasiva puede representar mediante el uso de un clasificador icónico el objeto de la acción verbal representada por la mano activa. También puede darse el caso en el que se aplique una variante de clasificador introflexivo, ya que la ejecución del signo con la mano activa se modifica para que la realización del signo que representa la acción verbal termine orientada hacia la mano pasiva (modificando el PFQ *orientación*). Incluso la posición final de dicha mano no sea la que dicta el signo, sino que también se posicionará en relación con la mano pasiva (modificando los PFQ *localización*, *plano* y *movimiento*).

Resulta muy ilustrador el ejemplo siguiente: El signo MIRAR-DETENIDAMENTE tiene un movimiento de la mano hacia el frente, dicho signo mantiene la orientación de la mano activa constante (Figura 2.4(a)). En una construcción que haga referencia



(a) Signo MIRAR-DETENIDAMENTE. (b) Signo MIRAR-DETENIDAMENTE modificado por un clasificador afijal.

Figura 2.4: Ejemplo de las modificaciones que implican en la ejecución de un signo la inclusión de un clasificador afijal.

a «mirar una naranja» se presentará la naranja mediante su signo. A continuación, utilizando un clasificador icónico, se establece su situación en el espacio de signado. Inmediatamente después, utilizando la mano activa, ejecutará el signo MIRAR-DETENIDAMENTE pero modificado mediante un clasificador afijal. De esta manera el final del signo se verá modificado para que la mano activa apunte hacia la mano pasiva que representa el objeto mirado (en este caso una naranja, Figura 2.4(a)).

Desde el punto de vista morfológico, se determina que estos clasificadores afijales tienen similitudes morfológicas con los clasificadores introflexivos, dado que hay una modificación de los PFQ del signo. Igualmente, se puede considerar que los clasificadores afijales requieren la ejecución de algún tipo de clasificador icónico, concluyendo que los clasificadores afijales son construcciones paralelas de los dos tipos de clasificadores antes descritos.

Una diferencia importante es que, en este caso, los nuevos PFQ que definen un clasificador introflexivo no se obtienen exclusivamente de la definición de otros signos. Estas nuevas modificaciones, utilizan la posición de otros elementos del cuerpo para definir la variación introflexiva del signo. Por esta razón es necesario dotar al sintetizador de un sistema que permita modificar el signo utilizando el sistema de definición de coordenadas utilizado para el clasificador icónico.

2.3. Notaciones para la descripción de los signos

Las lenguas orales han desarrollado un sistema de representación gráfica que permite plasmar de manera perdurable la información que se expresa con las mismas: la escritura. La aproximación a la definición de la escritura no es única, ya que puede estar basada en un sistema fonético, ideográfico o iconográfico. Las LS también han desarrollado varios sistemas que permiten recoger la descripción de los signos, y al igual que ha pasado con las lenguas orales, no existe una única aproximación. En esta sección repasaremos las notaciones más importantes.

2.3.1. Notación Stokoe

La notación Stokoe (1965) recibe el nombre de su autor William Stokoe y fue presentada por primera vez en su diccionario de Lengua de Signos Americana (ASL) [98] que recoge la definición paramétrica de los signos de esta lengua (Figura 2.5). Esta notación es la primera aproximación a la descripción paramétrica (o fonética) de las LS. Esta notación utiliza caracteres alfanuméricos así como un conjunto reducido de símbolos para definir la posición, la *configuración* y el *movimiento* de las manos.

La estructura de esta notación comienza con la descripción de la posición de las manos (Tab o *tabula*) que intervienen en el signo. Esta posición puede definirse de manera absoluta utilizando como referencia una parte del cuerpo o bien de manera relativa; generalmente la mano activa se posiciona dependiendo de la pasiva.

La *configuración* de la mano (Dez o *designator*) se describe utilizando la letra o número correspondiente al diccionario dactilológico de la ASL. Sobre este conjunto de *configuraciones* se pueden añadir un conjunto de modificadores que permiten indicar la flexión de los dedos relevantes en dicha *configuración*. En el momento de establecer esta notación, el PFQ *orientación* no se consideraba desvinculado como un PFQ independiente (ver Sección 2.1). Por esta razón la definición de la orientación de la mano se describe supeditada a la descripción de la *configuración*. Esto se realiza mediante un conjunto de símbolos presentados como subíndices del carácter que define la configuración.

Por último, el *movimiento* de la mano (Sig o *signification*) se describe utilizando un superíndice de la configuración. Esta notación permite definir los movimientos tanto de traslación de las manos como de variaciones de otros PFQ, en especial de la *orientación*, mediante un conjunto de símbolos. La notación permite describir movimientos que deben realizarse tanto en paralelo como de manera secuencial según la disposición de los superíndices.

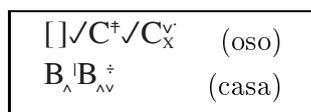


Figura 2.5: Ejemplos de la notación Stokoe para las definiciones de los signos OSO y CASA en ASL.

Esta notación presenta una serie de limitaciones lógicas como consecuencia del estado de los estudios sobre las LS en el momento de su definición. El parámetro PFQ *orientación* está incluido dentro de la definición de la *configuración*, por lo que no es posible establecer una variación en este parámetro de manera independiente. Otro de los PFQ que no se contemplan es el *punto de contacto*. Por último, todas las referencias a la componente *no manual* son omitidas de la definición. Su uso no se ha extendido debido, en gran medida, a la complejidad y al alto grado de abstracción de esta notación.

2.3.2. Notación SignWriting

La notación SignWriting [102] fue desarrollada por Sutton en 1974 como adaptación de la notación DanceWriting [101]. La gran diferencia de esta notación frente a la notación Stokoe, es que se trata de un sistema de representación iconográfico de los signos. Además, esta notación omite todo análisis paramétrico de las componentes del signo (Figura 2.6).

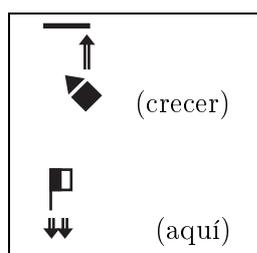


Figura 2.6: Ejemplos de la notación SignWriting para las definiciones de los signos CRECER y AQUÍ en LSE.

La *configuración* se representa en esta notación mediante el uso de un conjunto de iconos que representa de manera esquemática la forma de la mano. Sobre este icono, se define la *orientación* mediante tres alteraciones del icono de *configuración*: la primera es la inserción de una discontinuidad en el dibujo de la configuración que permite diferenciar entre los planos frontal y horizontal. La segunda se define mediante la rotación de dicho icono, lo que define la dirección de los dedos en el

plano de rotación indicado. Por último el relleno del fondo del icono permite ajustar la orientación de la palma de la mano en la dirección previamente definida.

La *localización* de la mano se identifica mediante un icono que muestra la referencia anatómica necesaria para el posicionamiento de la mano. Se pueden definir seis maneras para el caso de que la mano dominante entre en contacto con otra parte del cuerpo. Los movimientos se definen fácilmente utilizando flechas para describir su trayectoria y el grado de aceleración o tensión con que deben representarse. Por último, define el PFQ *no manual* mediante dibujos que indican expresiones faciales o movimientos corporales.

El manejo de esta notación resulta muy intuitivo y fácil de entender por lo que su uso se ha extendido en la comunidad sorda. Por el contrario, su tratamiento en aplicaciones informáticas resulta muy complejo debido precisamente a su componente gráfica. De igual manera, existe cierta libertad a la hora de definir los signos, dado que no se basa en una estructura fonética. También hay que destacar que esta notación no define el PFQ *plano*, ni especifica la parte de la mano activa o *punto de contacto*.

2.3.3. Notación HamNoSys

La notación HamNoSys (1989) presentada por Prillwitz *et al.* [86] utiliza la aproximación de la notación Stokoe realizando una descripción de los signos a nivel de los PFQ. Este nivel descriptivo añade la idea de iconicidad presentada en la notación SignWriting lo que aumenta su grado de usabilidad. Otra característica en común con la notación SignWriting es su independencia del idioma, al contrario que la notación Stokoe, ya que se describen los PFQ utilizando iconos representativos de la forma. La última de las novedades importantes de esta notación HamNoSys y que no posee SignWriting es la definición de una estructura sintáctica rígida que permite la formalización de este lenguaje, especialmente con vistas a su uso en el ordenador [43] (Figura 2.7).

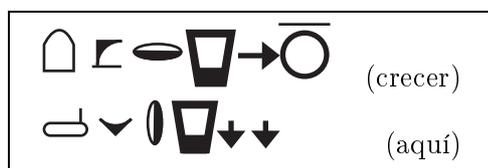


Figura 2.7: Ejemplos de la notación HamNoSys para las definiciones de los signos CRECER y AQUÍ en LSE.

La última versión disponible de la notación HamNoSys es la 4.0 [42]. En esta versión de la notación, los PFQ se describen de la siguiente manera:

La *configuración* se expresa mediante una combinación de iconos y números. Los iconos, de manera similar a la notación SignWriting, describen de manera esquemática la forma que adquiere la mano. Junto a estos iconos se especifica, mediante números, los dedos afectados en la *configuración* y, mediante símbolos, el grado de flexión de dichos dedos.

Esta notación desvincula la definición de la *orientación* de la *configuración*. Este PFQ se describe utilizando la combinación de dos iconos. El primero indica la dirección en el espacio y el segundo la orientación de la muñeca en esa dirección.

La posición de las manos queda definida pues por los PFQ *localización* y *plano*, siendo la única notación existente que incluye la descripción del PFQ *plano*. La definición de la *localización* se hace utilizando referencias anatómicas descritas igualmente por iconos. El *plano* se distingue como un conjunto discreto de cinco posibles valores, especificando el valor que debe tomarse por defecto si dicho PFQ se omite de la descripción. La descripción del *movimiento* se realiza de manera detallada mediante iconos que definen la trayectoria básica, la escala de dicho movimiento, la velocidad y los niveles de aceleración.

La componente *no manual* se aborda describiendo las expresiones faciales con cuatro elementos, las cejas, los párpados, la nariz y la boca (que incluye labios, mejillas y lengua). El número de posibilidades para cada uno de estos cuatro elementos puede verse en la Tabla 2.2. La definición de este PFQ también incluye la posibilidad de representar movimientos corporales. Para ello aprovecha la definición previa de movimientos que, en este caso, son aplicados a la parte del cuerpo para la que desea indicar el movimiento y que han sido definidos dentro de la notación mediante un conjunto discreto de posibilidades.

Componente de la expresión	Unidades definidas
Cejas	4
Párpados	13
Nariz	3
Boca	59 + visemas ^a

^aUn visema es el patrón de referencia visual de un fonema.

Tabla 2.2: Número de posibles estados para cada componente de la expresión facial definidos para la notación HamNoSys.

HamNoSys es la notación más completa para la descripción de los signos debido a su desarrollo conjunto con determinados proyectos destinados a la síntesis virtual de signos. Como inconvenientes de esta notación podemos resaltar que carece de la definición del PFQ *punto de contacto*. Aunque esta notación define un parámetro de

punto de contacto referido a las articulaciones de la mano, dichos elementos se usan para la definición de la *configuración* y no para el PFQ *punto de contacto*.

Sin embargo, el hecho de agrupar las posiciones de la lengua, los labios y las mejillas en un sólo parámetro expresivo, penaliza el número final de posibles expresiones, cuando sería más adecuada su definición independiente. Otro inconveniente de esta notación, referida por el propio Hanke [43], es el elevado número de símbolos de la misma lo que hace difícil su aprendizaje y da lugar a múltiples descripciones equivalentes para un signo.

2.3.4. Notación SEA

El Sistema de Escritura Alfabética (SEA) es una notación alfabética desarrollada en 2003 por Herrero Blanco [44, 45] para la descripción de los signos de la LSE. La notación SEA define los parámetros del signo (el autor se refiere a ellos como clases de fonemas) siguiendo la organización presentada en la Figura 2.8. Esta aproximación a la descripción parametrizada (o fonológica) de los PFQ de un signo mediante un sistema alfanumérico resulta muy similar a la notación Stokoe. La organización de los parámetros responde al orden en el que se procesan los parámetros en la generación del signo. Una diferencia importante entre las notaciones Stokoe y SEA es que esta última se representa linealmente, sin la necesidad de uso de subíndices ni superíndices, lo que simplifica su uso. Los parámetros que considera para dicha notación son:

- *Q*: la *configuración* de la mano
- *L*: el lugar de signado
- *C*: la existencia de *contacto* entre *Q* y *L* y su naturaleza
- *O*: la *orientación* de la mano
- *D*: el *movimiento* direccional
- *F*: el *movimiento* interno
- *S*: la bimanualidad

S LCQODF

Figura 2.8: Estructura de la notación SEA.

Esta, relativamente reciente, notación alfabética presenta una ventaja importante frente a sistemas de representación icónicos como son SignWriting o HamNoSys, y es

su posible manejo mediante ordenadores. Esta característica tiene su lado negativo en lo que a su manejo por personas se refiere, ya que carece de intuitividad en su uso, algo inherente a las notaciones icónicas (Figura 2.9).

re-V	(sordo)
sòaméaha àezpeyb	(escribir)

Figura 2.9: Ejemplos de la notación SEA obtenidos de [36].

En lo referente a los PFQ que se pueden describir con esta notación, se destaca que el *plano* no se incluye como elemento independiente y va ligado a la definición del lugar de signado mediante la definición de un sistema de coordenadas. En esta notación no se incluye la definición del PFQ *no manual*. Herrero Blanco argumenta que este PFQ sólo tiene función prosódica. Sin embargo se admite que la forma de la boca, la posición de la lengua y las expresiones faciales son relevantes en para el significado de un signo. Igualmente el parámetro *F* hace referencia a movimientos internos, que indican cambios en la *configuración* u *orientación* de la mano. Se demuestra así la falta de un criterio común, que nos plantea la siguiente cuestión: el hecho de que la mano pase de una *configuración* de dedos abiertos a otra de puño cerrado dentro de un signo, ¿debería definirse como una variación de *configuración* o un movimiento interno de los dedos de la mano?

2.3.5. Notación Szczepankowski

Se trata de una notación basada en caracteres ASCII desarrollada en 1994 por Szczepankowski y Rona [104], Szczepankowski [103], con el objetivo inicial de dar apoyo a la docencia, demostrando ser fácilmente adaptable para su uso en el ordenador [33]. Esta notación describe los signos mediante los siguientes campos:

- la *configuración* de la mano, mediante un conjunto discreto de 52 posibilidades
- la *orientación* codificada en base a un número de dos dígitos
- la *localización* descrita en relación a partes del cuerpo
- la relación entre las manos del signante, en términos de posición y distancia
- una dirección de *movimiento* entre un conjunto de 21 posibilidades
- parámetros adicionales al *movimiento* como son velocidad o distancia

Los cuatro primeros puntos describen componentes estáticos del signo, mientras que los dos últimos describen la parte dinámica del mismo.

PUm:35tpp # VI<+	(sordo)
PE:23k }/LBk:13k # P:III\V<-"	(escribir)

Figura 2.10: Ejemplos de la notación Szczepankowski. Las descripciones presentadas corresponden al modo de ejecución de los signos en la Lengua de Signos Polaca.

Aunque esta notación ha resultado ser pionera en el uso de un sistema alfanumérico y lineal para la descripción de los PFQ de un signo, presenta un número considerable de inconvenientes que dificultan su uso. Inicialmente, puede observarse que no todos los PFQ están presentes en la notación: el *plano*, *punto de contacto* y el componente *no manual* no están descritos. Otra limitación de esta notación es la capacidad para definir los PFQ que se deben utilizar. Como se explicó en la Sección 2.1, el número de orientaciones posibles es de 208, luego utilizando solamente dos dígitos no se pueden abarcar todas las posibilidades. Lo mismo ocurre cuando se trata de definir el *movimiento*, la limitación de todos los posibles a veintiuno resulta insuficiente. Al igual que pasa con el sistema SEA, elimina la iconicidad de la representación, lo que conlleva mayor grado de dificultad en su aprendizaje. Por último, en términos prácticos, un gran inconveniente para su uso como descripción para la LSE es que ha sido desarrollado para la Lengua de Signos Polaca, y toda la documentación y notación están en polaco (Figura 2.10).

2.3.6. Resumen de las notaciones

En la Tabla 2.3 presentamos un resumen con las principales características de las notaciones presentadas:

	Stokoe (2.3.1)	SignWriting (2.3.2)	HamNoSys (2.3.3)	SEA (2.3.4)	Szczepankowski (2.3.5)
Representación	Texto	Gráfica	Gráfica	Texto	Texto
Usabilidad	Media	Muy Alta	Alta	Media	Media
Uso en ordenador	Si ^a	No	No	Si	Si
Dependiente del idioma	Si	No	No	No	No ^b
PFQ definidos	3	6	6	5	4

^aRequiere cierta adaptación.

^bEsta notación está desarrollada para para la LS polaca.

Tabla 2.3: Resumen de las distintas notaciones para la descripción de los signos.

2.4. Conclusión

En este apartado hemos descrito tres elementos teóricos de la LS:

- Los distintos PFQ que intervienen en la descripción de un signo. Así mismo, hemos presentado cómo distintos autores que han descrito y añadido progresivamente PFQ diferentes hasta alcanzar los siete actuales.
- Las distintas unidades semánticas que aparecen en un mensaje de LS: el *diccionario dactilológico*, los *signos reconocidos o establecidos* y los *clasificadores*. Estas tres unidades establecen distintos niveles de complejidad semántica en el mensaje signado. En el caso de los *clasificadores* hemos visto cómo distintos autores presentan teorías alternativas para justificar estas construcciones en distintos idiomas. Nosotros nos hemos centrado en la descripción de estas unidades realizada para la LSE y utilizando una aproximación morfológica.
- Las distintas notaciones utilizadas para la descripción de los PFQ de los signos. En esta variedad de notaciones queda patente las distintas aproximaciones existentes según el número de PFQ que se consideran necesarios y por tanto incluidos en la notación.

Apartado 3

Estado del arte

3.1. Uso de la LS en los medios

Esta sección pretende realizar un repaso sobre las distintas técnicas mediante las cuales se presentan las lenguas de signos (LS) en los medios de comunicación. Aunque la mayoría de los trabajos versan sobre las aplicaciones que utilizan LS, en esta sección se ha tratado de ahondar en las técnicas y no en las aplicaciones. Las tres aproximaciones que se han utilizado para representar LS han sido el uso de notaciones, grabaciones en vídeo de personas y animaciones infográficas a partir de avatares virtuales.

3.1.1. Uso de notaciones

Parece lógico pensar que dado que tradicionalmente se ha utilizado el texto para comunicar y almacenar las lenguas orales, una primera aproximación para realizar la misma tarea pudiera ser utilizar una de las notaciones escritas de las LS. Ahmed y Seong [3] se basaron en esta aproximación para permitir la comunicación bidireccional entre personas sordas mediante teléfonos de texto.

La aplicación propuesta por Ahmed y Seong realiza una traducción de texto a notación SignWriting. Como se explica en la Sección 2.3, se trata de la notación que mayor aceptación ha tenido entre la comunidad sorda, debido principalmente a que resulta muy intuitivo y no requiere prácticamente aprendizaje. Obviamente esta aproximación para la comunicación de LS establece la misma relación de pérdida de naturalidad con el lenguaje signado que el lenguaje escrito establece con el lenguaje hablado.

Esta aproximación plantea como ventaja un número bajo de recursos de almacenamiento y de tiempo de cálculo en comparación con las soluciones que utilizan vídeo o elementos 3D. Además, permite una comunicación bidireccional entre personas sordas utilizando una manera de comunicarse más cercana a su lengua nativa (aunque en esta aplicación la entrada de información sigue siendo en texto).

3.1.2. Vídeos de personas

Dado que la LS es un lenguaje visual, la manera más natural de almacenar o transmitir este lenguaje generado por personas es en forma de vídeos. En primer lugar hay que mencionar los sistemas que utilizan vídeos pre-grabados para presentar mensajes largos. Tal es el caso de la página SignWriting¹, en la que disponen de lecciones sobre la lengua de signos, poemas y cuentos infantiles, entre otros. La mayoría de los vídeos que pueden encontrarse en esta web están en ASL. Dentro de los recursos en LSE encontramos la Biblioteca Virtual Miguel de Cervantes² que dispone de diversos recursos en forma de vídeos signados, entre ellos lecciones de gramática de la LSE. Además, en esta página web se utilizan imágenes dinámicas, en formato gif, con pequeñas animaciones de signos (Figura 3.1). Estas imágenes se utilizan para marcar e identificar los enlaces a otras páginas utilizando lengua de signos.

Dentro de esta categoría se incluye la subtitulación de programas de televisión. En este caso la imagen de un intérprete de signos se compone con la imagen del programa original. Dependiendo de si el programa se emite en directo o en diferido, la grabación del intérprete deberá ser realizada de manera simultánea o no.



Figura 3.1: Secuencia de imágenes que tiene un fichero gif para representar el signo LINGÜÍSTICA.

Esta aproximación para presentar contenido signado mediante vídeos de personas tiene la ventaja de presentar el mensaje manteniendo toda la naturalidad e inteligibilidad que tendría un intérprete en directo. Lo que se convierte también en uno de sus inconvenientes, puesto que se trata de un proceso manual. La generación de contenidos es un proceso costoso, cuya capacidad está limitada por el factor humano. Además, los contenidos generados no pueden ser reutilizados en distintos mensajes. Por último, hay que señalar que los requisitos de almacenamiento necesarios para esta

¹www.signwriting.org

²www.cervantesvirtual.com

solución son mayores en comparación con las otras aproximaciones que se plantean a lo largo de esta sección.

El uso de videoconferencia es una variante del uso de vídeos que permite la interacción entre dos personas. En una situación ideal, este método permitiría a dos personas la comunicación a distancia usando lengua de signos. Esta aproximación tiene el inconveniente de exigir a la red de comunicación un ancho de banda relativamente grande para permitir un vídeo fluido y un retardo relativamente pequeño. Las condiciones deseables no siempre están disponibles, en especial si se utilizan dispositivos móviles. Por esta razón, Cherniavsky *et al.* [17] plantean el uso de una tasa variable de imágenes por segundo unido a una variación de la velocidad de refresco de diferentes secciones de la imagen. Las secciones que mayor velocidad de refresco deberán tener son aquellas en las que el signante oyente más se fija. Aunque esta técnica ya se ha aplicado a la videoconferencia, la misma aproximación puede ser aplicada al *streaming* de cualquier contenido signado. Especialmente si la capacidad de la red a la que está conectado el dispositivo del cliente no proporciona suficiente ancho de banda.

3.1.2.1. Síntesis de vídeo por concatenación de unidades

El principal inconveniente del uso de vídeos signados para la representación de contenidos se debe al hecho de que la creación de los mismos es un proceso manual. Es por esto que los vídeos sólo representan un mensaje preestablecido, es decir, no permiten adaptarse dinámicamente a cualquier contenido. Para conseguir esta funcionalidad deberemos sintetizar automáticamente la LS.

La técnica de síntesis de LS que se utiliza para poder generar mensajes de manera automática a partir de vídeos signados es equivalente a un sintetizador de voz por concatenación de palabras. El trabajo presentado por Solina *et al.* [96] se basa en la utilización de secuencias de vídeo pre-grabadas. Cada una de esas secuencias contiene un signo completo. Un sistema de estas características posee la inteligibilidad idónea en la representación de cada signo, pero la transición entre distintos signos concatenados puede mermar este aspecto. Dicho trabajo utiliza grabaciones con unos parámetros bien definidos como son que el intérprete comience y termine la articulación de todos los signos en la misma posición, que el intérprete sea siempre el mismo, no cambie su indumentaria, se mantengan las condiciones lumínicas del fondo y el ángulo y posición de la cámara. Para realizar una concatenación suave

entre los distintos signos se realiza previamente un procesado de imagen para determinar la posición de los brazos seguido de un *morphing de vídeo*³ entre las imágenes final e inicial de dos signos correlativos, siendo en este punto donde se produce una pérdida de calidad en el mensaje final. Estas técnicas de procesado de la señal de vídeo también son utilizadas por Neidle *et al.* [82], aunque el objetivo de Neidle *et al.* es el etiquetado de secuencias de vídeo para el desarrollo de una base de datos que permita el estudio de las LS (en particular en este proyecto se estudia la ASL).

Este sistema disminuye los requisitos de almacenamiento respecto a los mensajes grabados en vídeo, manteniendo prácticamente el nivel de inteligibilidad de un intérprete humano. La penalización en la naturalidad se debe a la invariabilidad en la realización de los signos: de igual manera que una persona no firma dos veces igual, tampoco realiza exactamente igual el mismo signo. Como bien manifiesta Solina *et al.*, la adición de nuevas unidades de signos es un proceso complejo si se quiere evitar fallos de continuidad. Para solventar este problema realizó todas las grabaciones en el mismo día. Dicha planificación de trabajo resulta inviable si se pretende grabar un número elevado de signos.

La reducción de requisitos de almacenamiento, aunque sustancial con respecto a un vídeo completo, no alcanza los niveles de los sistemas basados en animación de personajes virtuales, que pueden reutilizar información intra-signo. Por último, hay que señalar que el procesado de la señal de vídeo y los cálculos para la transición entre secuencias tienen un alto coste de procesamiento.

3.1.3. Animación de personajes virtuales

La tercera aproximación a la representación de mensajes signados es el uso de un avatar⁴ en 2 ó 3 dimensiones. Los avatares que realizan lengua de signos poseen la misma estructura básica que los personajes que se utilizan para interfaces de usuario o videojuegos. Una malla de polígonos forma la apariencia del avatar, mientras que la animación de la malla se realiza mediante una estructura esquelética o mediante técnicas de *morphing de malla*⁵. La estructura de los avatares se estudiará con más detalle en la Sección 3.2.1. En esta sección únicamente nos vamos a centrar en las

³Se denomina *morphing de vídeo* a un tipo de transición entre dos imágenes en la que se tienen en cuenta las formas representadas. Dicha técnica realiza una interpolación entre la posición de dos elementos que definen un mismo objeto. Dicha posición puede establecerse manualmente o mediante el procesado de las imágenes.

⁴Personaje virtual diseñado para interactuar con el usuario.

⁵La técnica de *morphing de malla* es conceptualmente similar al *morphing de vídeo*. La diferencia está en que no es necesario ningún procesamiento o la definición de puntos equivalentes en distintas imágenes. La transición se realiza interpolando las posiciones de un vértice entre los distintos instantes de la animación.

aproximaciones que se utilizan para definir dicha animación y así poder compararlo con sus equivalentes que usan el vídeo.

3.1.3.1. Técnicas de animación de personajes

La definición de la animación del avatar puede realizarse mediante tres técnicas posibles o una combinación de ellas. Dichas técnicas son: animación manual, captura de movimientos y descripción paramétrica del signo.

En la animación manual de un mensaje signado se sigue la misma técnica utilizada para la generación de películas de animación. Un animador profesional define las rotaciones de todas las articulaciones del avatar. Esta definición puede realizarse en todos los fotogramas de la animación final o sólo en algunos de ellos (fotogramas clave). Se trata de un proceso lento pero que permite generar resultados de alta calidad, ya que la animación puede ser retocada para hacer que la expresividad del avatar resulte lo más realista posible (Figura 3.2).

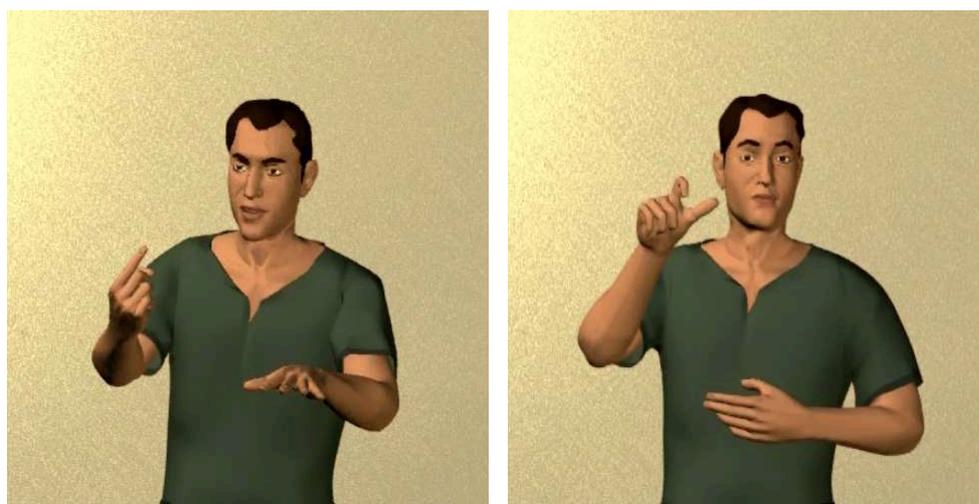


Figura 3.2: Esta figura contiene distintos fotogramas de una película en LSE. La animación ha sido creada manualmente [95].

La captura de movimiento (Figura 3.3) implica el uso de un conjunto de dispositivos destinados a registrar las rotaciones que realizan las articulaciones de una persona durante el proceso de signando. Las técnicas son muy variadas, como exponen Sagawa y Takeuchi [88], pueden usarse desde sistemas mecánicos como los guantes de captura utilizados por Loomis *et al.* [72], sistemas de procesamiento de imagen utilizados por el propio Sagawa y Takeuchi o con marcadores especiales colocados en el cuerpo del signante y grabación posterior con cámaras en el espectro infrarrojo. Esta última



Figura 3.3: Dispositivo utilizado para captura de movimientos y expresión facial.

técnica fue utilizada por Bangham *et al.* [8] para grabar las expresiones del signante durante la ejecución de los signos.

Aunque esta técnica se presenta aquí como parte de aplicaciones de síntesis (Figura 3.4), está más extendida en el campo del reconocimiento de signos, en particular queremos destacar el sistema de diálogo presentado por Lu *et al.* [73] para la Lengua de Signos Japonesa. Esta aplicación requiere el uso de dos cámaras y la utilización por parte del usuario de un guante de color para facilitar el procesado de la señal. El trabajo presentado incluye un experimento para medir las zonas del signante sobre las que se presta una mayor atención, que se ha resumido en la Tabla 3.1.

Sección de la imagen	Porcentaje de atención
Boca	60,3 %
Cara	16,8 %
Manos	4,1 %
Subtítulos	15,7 %

Tabla 3.1: Principales zonas de atención cuando se observa un mensaje signado.

Esta técnica permite convertir directamente los movimientos realizados por una persona en los movimientos del avatar. De este modo se obtiene una animación muy natural y expresiva. Aunque también presenta ciertos inconvenientes. En primer lugar es el coste y la complejidad en el uso del equipo de captura. Por otra parte, los valores obtenidos para las rotaciones articulares restringen en gran medida la anatomía del avatar, que debe ser proporcional a la de la persona grabada. La segmentación de unidades semánticas, para una posible reutilización, es un proceso costoso ya que debe ser realizada manualmente, dado que la animación capturada es un flujo continuo de información. De este flujo no todos los PFQ pueden ser obtenidos directamente, como el *plano* o la *localización*, y requieren un proceso complejo si se desea obtener dicha descripción. Por último, hay que destacar que la captura de movimientos suele

realizarse muestreando a una frecuencia de entre 30Hz y 60Hz . [23]. Los requisitos de almacenamiento son equivalentes a los de una animación manual e inferiores en comparación con la aproximación basada en vídeo. No obstante, no son lo suficientemente reducidos como para equipararse a la aproximación paramétrica.

La tercera técnica que se utiliza está basada en el estudio de los PFQ. La síntesis del avatar recibe la definición de los signos utilizando alguna de las notaciones paramétricas descritas en la Sección 2.3 y planteadas por Irving y Foulds [54]. Esta información es utilizada para la definición de las articulaciones del avatar en instantes de tiempo concretos (fotogramas clave). Las orientaciones de dichas articulaciones son definidas en instantes intermedios de tiempo mediante técnicas de interpolación. Un ejemplo de esta aproximación fue propuesto por Zwiterslood *et al.* [110].

Como se presenta en la Sección 2.2, la gran variabilidad en la realización de las distintas unidades que componen el mensaje signado requiere que la síntesis de signos sea lo suficientemente flexible como para acomodarse a todas esas variaciones. La animación paramétrica es la que posee mayor facilidad para cubrir dicho requisito. Dado el número de combinaciones posibles, el uso de otras aproximaciones no sería viable. La aproximación paramétrica reduce en gran medida los requisitos de almacenamiento. Kennaway [61] defiende esta técnica frente al uso de información capturada previamente puesto que permite una fácil adaptación a las diferentes anatomías que puedan presentar los avatares. A esto último hay que añadir la flexibilidad para generar movimientos complementarios a la lengua de signos que den mayor naturalidad y variabilidad al mensaje sintético. Por último esta aproximación no requiere el uso de equipos especiales para la captura de movimientos o la dependencia de una persona que realice las capturas de movimiento, lo que facilita extender el diccionario de términos.

La contrapartida que presenta el sistema de animación paramétrico es una disminución de la naturalidad de movimientos del avatar. Implica además, un aumento en los procesos de cálculo necesarios para generar el mensaje signado, como por ejemplo, la posición de las manos debe ser definida mediante cálculos de cinemática inversa (IK) [106]. Incluso, las transiciones entre dos posiciones definidas pueden dar lugar a colisiones entre distintas partes del cuerpo. Estos dos procesos serán más estudiados en las secciones 9.2 y 9.3.

Aunque se han presentado estas tres técnicas por separado, no quiere decir que sean excluyentes unas con otras. Se puede abordar el proceso de síntesis utilizando combinaciones de varias de ellas. Se podría plantear una síntesis paramétrica en la que las distintas *configuraciones* estuvieran definidas por un animador humano u obtenidas de un sistema de captura de movimientos. De igual manera, se podría

tomar la información capturada de una persona y posteriormente ser retocada por un animador profesional.

3.1.3.2. Alternativas en la síntesis basada en avatares

Se ha expuesto cómo la presentación de mensajes signados utilizando vídeos grabados puede abordarse mediante vídeos de mensajes completos o a través de la composición de secuencias que representan unidades aisladas. Los mensajes signados basados en animación de avatares permiten estas dos alternativas y además añaden una tercera.

La primera aproximación de los mensajes en vídeo presenta la posibilidad de que todo el mensaje se haya generado de manera continua. Cuando se describió la animación manual de avatares, se hizo referencia a que el método es equivalente a los utilizados en las películas cinematográficas de animación. Por lo tanto, el resultado final de dicha labor da lugar a una secuencia animada equivalente a un vídeo [95]. De igual manera se puede tomar la captura de movimiento de una persona y aplicarla en conjunto a un avatar. Esta aproximación permite adaptarse a un entorno en el que no se dispone de suficiente ancho de banda para el envío de un vídeo. Además hace que el mensaje resulte más impersonal al no estar asociado a una persona.

Como inconveniente, el uso de mensajes completos no permite la generación de contenidos de manera dinámica. Para suplir esta deficiencia, Bangham *et al.* [8] aporta una solución equivalente a la composición de secuencias de vídeo, la composición de secuencias de animación⁶. Esta técnica concatena las pistas de animación capturadas para un determinado hueso, definiendo una transición entre los valores final e inicial de dos secuencias consecutivas. Lo que presenta una ventaja frente a su homóloga aplicada a secuencias de vídeo, la generación de transiciones entre secuencias consecutivas es mejor y no se observa variación en la calidad de imagen. Esto es así porque se puede generar una transición suave entre los valores de las rotaciones de cada articulación. Aun así, la calidad y naturalidad de la animación durante estas transiciones no alcanza la calidad obtenida en la animación capturada.

Las dos aproximaciones anteriores adaptan en cierta medida las técnicas que se utilizan en el vídeo: la definición de un mensaje completo y la concatenación de unidades. Pero como hemos visto, la animación de avatares permite la definición paramétrica de los signos. Esta última alternativa permite que no sólo las transiciones entre signos sean sintéticas, sino que toda la animación lo sea. Esta técnica usada

⁶Grieve-Smith [40] los llama "sintetizadores 3D basados en glosas por concatenación de unidades".

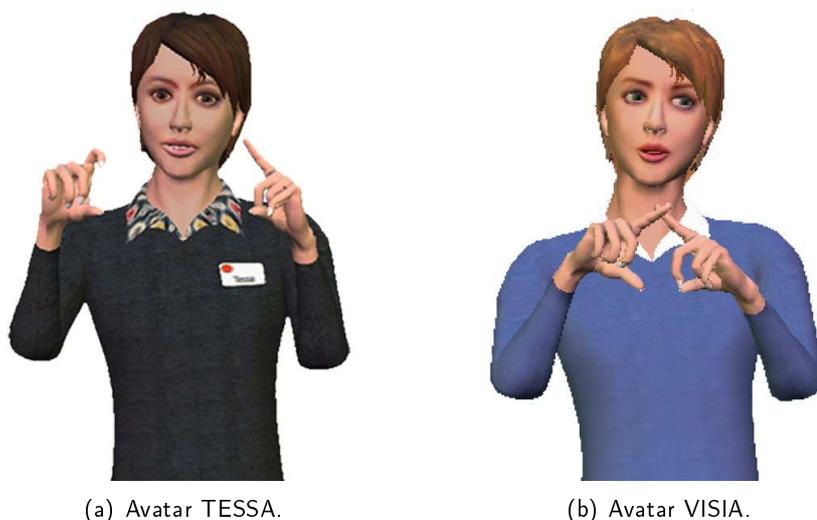


Figura 3.4: TESSA y VISSIA. Estos avatares permiten la animación tanto mediante el uso de secuencias completas de signos como por concatenación de secuencias individuales.

por Zwiterslood *et al.* [110], permite modificar dinámicamente la síntesis de signos manteniendo constante la calidad en la animación del avatar.

3.2. Síntesis paramétrica de signos

La técnica más extendida para realizar la síntesis de signos es la aproximación paramétrica. Como hemos visto en la sección anterior, no es una técnica que permita obtener resultados con el nivel de naturalidad que ofrece el vídeo. Incluso es la aproximación que requiere un mayor tiempo de cómputo debido a la necesidad de generar la animación del avatar a partir de la descripción del signo.

El hecho de que ésta haya sido la técnica que más se ha utilizado se justifica con el hecho de ser la opción que requiere menos recursos para ampliar la cobertura del diccionario, dado que es el método que mayor grado de automatización permite. De igual modo, permite mantener una calidad de animación constante a lo largo de todo el mensaje sintetizado. Por último, esta aproximación permite cubrir toda la variabilidad existente en la LS que se presentó en el Apartado 2.

La Figura 3.5 muestra los principales elementos que constituyen un sintetizador paramétrico de LS. En esta sección estudiaremos las distintas definiciones de avatares que pueden encontrarse en la literatura, haciendo mención a la base matemática que existe tras la animación esquelética. De igual modo revisaremos las posibles maneras que existen de representar la información necesaria para la síntesis de signos.

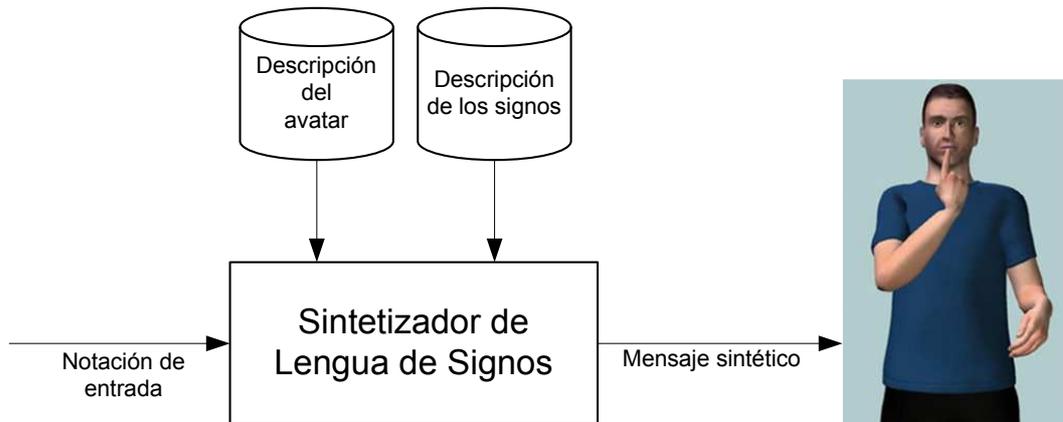


Figura 3.5: Arquitectura básica de un sintetizador paramétrico de Lengua de Signos. En algunas arquitecturas la descripción de los signos está incluida en la propia notación de entrada.

3.2.1. Avatares

Un avatar es una representación virtual de una persona. Esta representación puede hacerse en 2D o 3D. Para el estudio de los avatares que se han utilizado para la síntesis de LS se ha optado por realizar una primera clasificación en función de la anatomía definida. Posteriormente se repasan las estructuras definidas para la animación del avatar y se presentan algunos ejemplos de los avatares más representativos en este área. La gran mayoría de los avatares utilizados son tridimensionales debido al realismo superior que se puede alcanzar, aunque en determinados trabajos, más centrados en la traducción, se ha optado por utilizar un avatar bidimensional [90].

3.2.1.1. Avatares parciales

La primera aproximación a la síntesis paramétrica de LS la componen las aplicaciones que sólo se centran en la representación de las manos. Los únicos PFQ que pueden representarse con sólo las manos son la *configuración* y la *orientación*. Por lo tanto estas aplicaciones omiten toda representación de los signos y se centran en la dactilología.

Teniendo en cuenta el número limitado y reducido de las unidades dactilológicas, es factible una aproximación mediante dibujos estáticos. Este es el caso de la aplicación propuesta por Agravat [2], desarrollada en Flash presenta una secuencia de imágenes que corresponden a la traducción dactilológica de la palabra introducida. Esta aplicación no contempla transiciones entre distintas letras, lo que penaliza el resultado final.

El uso de una aproximación en 3D permite que la generación de transiciones entre dos unidades consecutivas pueda crearse dinámicamente, lo que mejora el resultado obtenido. La primera aproximación en este aspecto fue propuesta por Su y Furuta [99] al definir un modelo tridimensional de una mano. El modelo propuesto resultaba muy sencillo al estar compuesto simplemente por cilindros y esferas. Su y Furuta se decantan en este estudio por la animación sintética al considerarla más flexible que la captura de movimientos. El diseño de la mano fue mejorado por McDonald *et al.* [77] mejorando la apariencia de la mano propuesta por Su y Furuta al modelar los dedos mediante elipsoides. Aun así, el uso de una estructura no deformable para representar la mano no es la que ofrece resultados más realistas.

Los *talking heads* o cabezas parlantes no pueden definirse por si mismas como síntesis de signos, pero consideramos que este tipo de avatares deben ser mencionados. Esta opinión no se debe a la aportación que realizan por si mismas, sino a las posibilidades que surgen cuando se integran en un avatar completo. Dado que nos estamos centrando en la síntesis paramétrica de signos, de igual modo sólo son relevantes las aproximaciones que usen esta técnica para la animación facial, como por ejemplo las aproximaciones puramente paramétricas planteadas por Beskow [10, 11] o basadas en la captura de movimientos y posterior segmentación de unidades [74]. El principal objetivo de estas cabezas es la de proporcionar una interfaz más natural a ciertas aplicaciones ya que presentan un movimiento de labios sincronizado con el audio. Esto permite además la generación de aplicaciones para asistencia a personas con limitaciones parciales de audición [12] u otro tipo de discapacidad psicológica [21]. Pero esta técnica no puede ser utilizada por si misma para la comunicación con personas sordas que se comunican mediante LS.

Dado que las formas que adquiere la boca para realizar la simulación del habla son un subconjunto de las requeridas para la síntesis de signos, estos proyectos deben ser tenidos en cuenta en la síntesis de signos; sin olvidar que el resto de la animación facial, usada para la simulación de expresiones, compone gran parte del PFQ *no manual*.

Como se ha comentado, una de las principales posibilidades que presentan las *cabezas parlantes* es la simulación de movimiento labial emulando el habla. El trabajo de Martínez Lazalde *et al.* [75, 76] presenta una importante aproximación para la síntesis de movimiento labial en español. De este trabajo se ha obtenido la Tabla 3.2 que ha sido utilizada en esta tesis para generar la animación relacionada con la simulación del habla.

La mayoría de los sistemas de animación facial utilizan una malla definida por polígonos. Esta malla es animada siguiendo una técnica denominada *morphing de*

Grafema	Visema	Fonemas
silencio, h	NEUTRO	
a	a	a
e	e	e
i	i	i
o	o	o
u	u	u
f	f	f
g, w	g	g
j	j	x
l	l	l
b, m, p, v	bmp	b, m, p
ch, ll, y, x	chy	tʃ, ʎ, j, x
d, s, t, z, c	dst	d, s, t, θ
k, q, c	k	k
n, ñ	n	n, ɲ
r, rr	r	ɾ, r

Tabla 3.2: Relación entre los fonemas y los visemas del español.

mall, que hemos nombrado anteriormente. El resultado de esta manera de modelar el avatar le proporciona una apariencia aceptable. Gracias a las técnicas de iluminación la superficie de la cara no muestra aristas, sin embargo, estas aristas si aparecen en la silueta del modelo. Los autores van Zijl y Barker [108] proponen el uso de superficies NURBS [85] para mejorar la apariencia del avatar. El problema es que dicha técnica requiere un alto procesamiento y por tanto no se obtienen animaciones suficientemente fluidas cuando la generación de fotogramas se realiza en tiempo real.

El último de los avatares parciales que se van a tratar es el propuesto por Adamo-Villani y Beni [1]: el *semantroid*. Este avatar sólo representa la cabeza y las manos para generar lengua de signos. El objetivo de este tipo de representación es poder incluir el avatar cuando existe una restricción vertical del espacio que impide la representación del avatar completo. Permite además, centrarse en la cara y manos, que son los elementos que reciben mayor atención.

Aunque el aspecto visual de esta aproximación pueda presentar ciertas ventajas frente a un avatar completo, existen ciertas implicaciones que desaconsejan su uso. El cuerpo del signante se toma como referencia para la *localización* lo que implica que, al desaparecer éste, determinados signos presentarán ambigüedad al no conocerse este PFQ. Lo mismo ocurre con el PFQ *plano*: la distancia de las manos al cuerpo no puede ser medida y por lo tanto el resultado final no es exacto. De igual manera el *movimiento* de las manos carece de un punto de referencia que permita medir su amplitud y velocidad.

3.2.1.2. Avatares completos

Los avatares completos representan todos los elementos necesarios para la síntesis de signos. En ocasiones, dado que el encuadre de la cámara sólo muestra al avatar de cintura para arriba, la parte inferior del mismo no suele modelarse.

Todo avatar virtual requiere dos de elementos: una estructura visible que defina la apariencia del avatar (malla) y una estructura de control que permita modificar la forma de la malla para generar distintas posturas (esqueleto). Esta aproximación es válida para emular la apariencia del cuerpo humano y los movimientos articulatorios. Para las variaciones en la apariencia del avatar no generadas por una rotación articular se utiliza la técnica denominada *morphing de malla*. La animación de las expresiones faciales se realiza mediante la definición de diversas expresiones alternativas, denominadas objetivos de transformación. La posición de los vértices visibles se interpola entre los valores definidos en los objetivos de transformación para simular dichas expresiones. Los objetivos de animación pueden no cubrir la totalidad de los vértices de la malla, lo que permite que las modificaciones afecten únicamente a una parte de la cara.

En ocasiones, ciertos autores han optado por integrar ambas estructuras en una sola. Como puede observarse en la Figura 3.6, la malla está compuesta por un conjunto de objetos no deformables que crean la apariencia del avatar, los mismos objetos crean la estructura de control necesaria para definir la postura.



Figura 3.6: Avatar definido unificando la estructura visible y la estructura de control.

Descripción del esqueleto y la transformación sobre la malla

Una estructura esquelética es una cadena jerárquica (cadena cinemática) en la que cada elemento se define como una transformación sobre el sistema de referencia del antecesor en dicha cadena. Esta transformación se compone de una traslación (T), una orientación (R) y un escalado (S). Por lo tanto un punto descrito en coordenadas homogéneas $p = (x, y, z, 1)$ en el sistema de coordenadas de un cierto elemento, se

describe como $p' = (x', y', z', 1)$ en el sistema de coordenadas del elemento antecesor por la transformación Ecuación 3.1.

$$p' = TRSp \quad (3.1)$$

La transformación de la malla en función de las modificaciones del esqueleto se calcula siguiendo la Ecuación 3.2.

$$v' = \sum w_i M_i B_i v \quad (3.2)$$

La posición v' de un vértice tras la transformación equivale al sumatorio de las transformaciones aplicadas por cada hueso ($M_i B_i$) a la posición inicial del vértice v ponderada por el peso o influencia que tiene dicho hueso sobre el vértice en cuestión (w_i). Donde, B_i es la transformación del sistema de referencia hueso i a un sistema global cuando el hueso está en reposo (posición por defecto). Así mismo, se define M_i como la transformación del sistema de referencia del hueso i al sistema global de referencia en el instante dado. La Figura 3.7 muestra de manera esquemática esta operación.

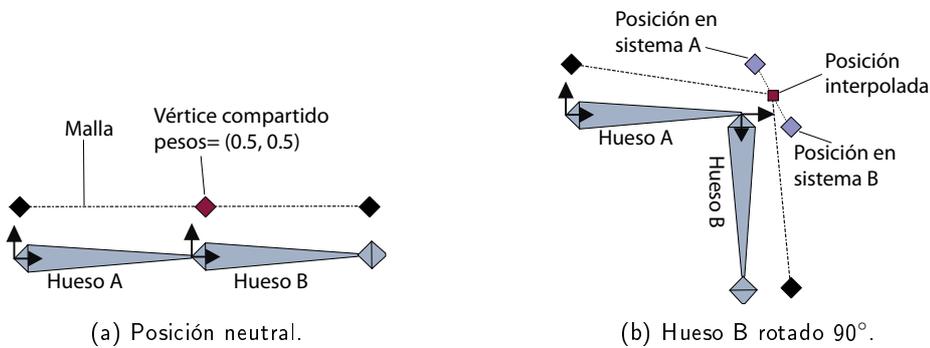


Figura 3.7: Ejemplo del funcionamiento de la deformación de la malla debido a una transformación en el esqueleto.

Estructuras esqueléticas más utilizadas

Una vez explicado el fundamento matemático de la animación esquelética, se debe realizar un recorrido por los diversos esqueletos utilizados para la síntesis de signos. Todos los diseños parten de un estudio anatómico del cuerpo humano sobre el que realizan una serie de simplificaciones para reducir el número de elementos en la cadena cinemática. La principal simplificación se produce al definir la estructura que simula la columna vertebral, lo que disminuye en gran medida el número de articulaciones.

Uno de los diseños más extendidos entre los proyectos de lengua de signos es H-ANIM [58]. Este estándar presentado por Roehl permite la definición de una estructura esquelética para un humanoide. La definición de este humanoide debe cumplir ciertos niveles de articulación y establecer ciertos límites a la libertad de acción de las articulaciones para cumplir con el estándar. Sin embargo, el estándar no especifica ningún tipo de relación entre las dimensiones anatómicas del modelo, que deben ser tenidas en cuenta al aplicar animaciones complejas.

El estándar H-ANIM incluye la definición de los FAPs⁷ organizados en un vector que permite establecer la expresión facial. Cada elemento de ese vector define el peso que se debe aplicar a cada uno de los objetivos de transformación. Estos objetivos sólo afectan a determinadas partes de la cara, por lo que la composición de expresiones es más sencilla. Estas definiciones han sido tomadas del estándar MPEG-4 [55]. De igual manera, se toma de dicho estándar la definición de los BAPS⁸ que permiten definir las rotaciones de cada una de las articulaciones que componen el cuerpo. Por último, este estándar define también un conjunto de posiciones del cuerpo (Apéndice C del estándar H-ANIM) utilizando como referencia ciertos vértices de la malla que deben ser definidos por el diseñador. El conjunto de puntos definidos por H-ANIM no cubre la totalidad de puntos requeridos para la definición de la LS. La ventaja de esta estructura estándar es que al estar contemplada dentro del estándar VRML [56, 57] puede ser visualizada mediante un visor estándar. La Figura 3.8 muestra distintas estructuras esqueléticas, en particular una estructura definida mediante H-ANIM en la Figura 3.8(a) y las estructuras utilizadas en el proyecto ViSiCAST (Figura 3.8(b)) y eSIGN (Figura 3.8(c)).

Papadogiorgaki *et al.* [84, 83] plantea el paso de SWML (ver Punto 3.2.2.1) a BAPS y FAPs para generar la animación de LS. Según Papadogiorgaki *et al.* uno de los problemas que surgen al combinar los FAPs y los BAPs se produce al sincronizar ambos elementos, dado que los FAPs requieren mayor tiempo de proceso. La solución propuesta es la de supeditar los BAPs a los FAPs. Esto implica generar una relación de dependencia entre distintos PFQ en base únicamente a la definición de la librería 3D que se está utilizando, lo que no es justificable. Caridakis y Karpouzis [15] optan por el uso de una notación intermedia previa a la definición de los BAPS, el lenguaje STEP [47]. STEP permite la definición de acciones del avatar a un nivel de abstracción superior a de los BAPs, y permite definir de manera sencilla acciones paralelas.

Como hemos visto, H-ANIM se presenta como una herramienta con la que se pueden generar animaciones de un personaje humanoide. Esto es especialmente útil

⁷Facial Animation Parameters

⁸Body Animation Parameters

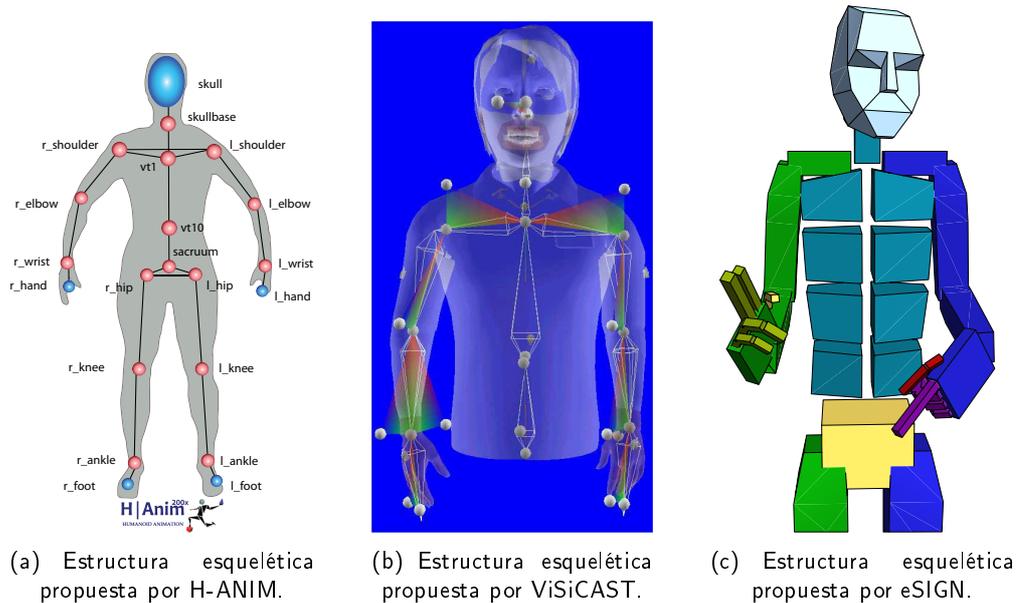


Figura 3.8: Estructuras esqueléticas utilizadas para la animación de avatares.

cuando el objetivo de la investigación que se lleva a cabo no pretende lidiar con el avatar. No obstante hemos visto cómo se mezcla la definición de la malla con la definición del esqueleto cuando se definen los puntos del cuerpo. Este hecho puede provocar que una malla con bajo nivel de detalle no permita definir todos los puntos necesarios para la lengua de signos. La animación que permite realizar H-ANIM utiliza cinemática directa, luego todos los cálculos necesarios que requieran cinemática inversa deberán realizarse aparte. Por último, Yang *et al.* [107] defiende que una implementación hecha en Java3D funcionará más rápido y será más fácil de implementar que una hecha con H-ANIM (Figura 3.8(a)).

Otro diseño esquelético que ha tenido gran repercusión en la síntesis de signos es la estructura creada por los proyectos ViSiCAST (Figura 3.8(b))y eSIGN (Figura 3.8(c)). La estructura básica de ambos esqueletos es similar entre sí aun cuando el modo de definir las animaciones es distinta. Recordemos que ViSiCAST genera la animación a partir de información capturada de un intérprete y el proyecto eSIGN mediante la aproximación paramétrica. De igual modo, este diseño tiene ciertas similitudes con el diseño H-ANIM. La mayor diferencia entre H-ANIM y la aproximación seguida en eSIGN es en relación al número de puntos que se utilizan de la malla para definir el PFQ *localización*, en el caso del proyecto eSIGN, los puntos de la malla que se escogen están en relación con la definición de la notación HamNoSys.

La animación de los avatares se logra mediante modificaciones en los valores de rotación de las articulaciones a lo largo del tiempo. Según se elije la librería gráfica, se

plantean dos maneras para dicha animación. La primera de ellas requiere la definición de orientaciones clave para las articulaciones con una marca de tiempo. Será la propia librería gráfica la encargada de interpolar dichos datos para obtener los valores de rotación en instantes de tiempo intermedios. La segunda opción requiere que estos valores se definan fuera de la librería, por lo que se debe generar una rutina que calcule dichos valores.

Aunque el método de definición de la animación anterior suele ser el más extendido, otros autores como Lebourque y Gibet [65, 66] plantean la adaptación de sistemas de control robóticos para la animación del esqueleto del avatar. Esta aproximación ha recibido el nombre de GESSYCA, aunque no aporta una mejora de la calidad final del sistema. Además, no permite la definición de las expresiones faciales, lo que justifica que no haya tenido una mayor aceptación.

Avatares en la síntesis de LS

El último punto a considerar sobre el uso de los avatares pasa por la revisión de los propios avatares. Como puede verse en la Figura 3.9, el acabado visual de los modelos utilizados en distintos proyectos es muy variado. En función del planteamiento teórico en que se basa y de la tecnología que se utiliza, se deben establecer restricciones en el aspecto final del avatar para obtener una determinada tasa de imágenes por segundo. Los factores que influyen en dicha tasa son el número de polígonos de la malla, el uso de texturas, la implementación de la librería gráfica utilizada, etc.

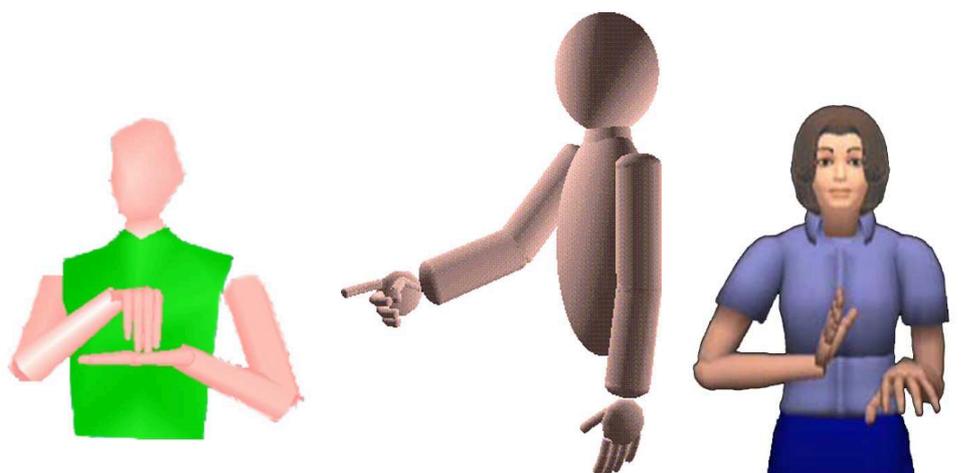
Los avatares más sencillos que se pueden encontrar en la literatura utilizan un conjunto de volúmenes básicos para su definición. Dos ejemplos claros de esta aproximación son los utilizados por Francik y Fabian [35] (Figura 3.9(a)) y por Lebourque y Gibet [65] (Figura 3.9(b)). Ambos avatares carecen de la capacidad para representar expresiones faciales. Dada su simplicidad (especialmente el presentado por Francik y Fabian) el resultado son imágenes confusas con distintas posiciones de las manos, lo que penaliza la inteligibilidad.

El siguiente nivel de complejidad en los avatares queda representado por el modelo utilizado por Huenerfauth [49] (Figura 3.9(c)). Aunque utiliza estructuras no deformables, el resultado gana en inteligibilidad con respecto a los anteriores gracias a un diseño más elaborado y permite además, la simulación de expresiones faciales, con lo que se logra avanzar en naturalidad.

Aunque el resultado obtenido con el avatar de Huenerfauth ha mejorado en inteligibilidad, la falta de continuidad en la malla del avatar penaliza en cierto grado el

resultado final. La obtención de un avatar óptimo se consigue en un primer paso con el uso de mallas continuas deformables, tal es el caso de los proyectos VSIGN [32] (Figura 3.9(d)) y VSIGNS [53] (Figura 3.9(e)).

El siguiente grado de mejora en a la apariencia del avatar es la aplicación de texturas para dotar de mayor realismo visual al avatar. Esta es la aproximación que se ha aplicado a los avatares de los proyectos ViSiCAST (Figura 3.9(f)) y eSIGN (Figura 3.9(g)) [63]. El buen resultado obtenido con el avatar de eSIGN ha dado lugar a su adopción por otros proyectos para realizar la síntesis, tal como el propuesto por San Segundo *et al.* [89].



(a) Avatar "Francik y Fabian".

(b) GeSSyCa.

(c) Avatar "Huenerfauth".



(d) VSIGN.



(e) VSIGNs.



(f) ViSiCAST.



(g) eSIGN.

Figura 3.9: Distintos avatares utilizados para la síntesis de signos.

3.2.2. Almacenamiento de signos

Tras describir la estructura de los avatares que se utilizan para la síntesis de signos, debemos centrarnos en el modo en que se describen los signos para ser sintetizados. Como vimos en la Sección 2.3, existen numerosas notaciones para la descripción de los signos. Estas notaciones, con ciertas modificaciones, han sido adaptadas para realizar la descripción de los signos mediante ordenador. Aunque el uso de una notación es el modo más común de describirlos, existen autores que proponen el uso de una base de datos relacional para almacenar la descripción del signo.

3.2.2.1. Notaciones

El primer criterio que se plantea en el uso de una determinada notación para la síntesis de signos es la compatibilidad con el manejo automático por ordenador. Lógicamente, debido a su componente gráfico, notaciones como HamNoSys o SignWriting complican este uso. Las notaciones que pueden ser utilizadas directamente en un sintetizador automático de signos son ASCII-Stokoe, SEA, y Szczepankowski.

Resulta paradójico que las notaciones que puede utilizarse directamente para la síntesis de signos, sin necesidad alguna de modificación, sean las que menos se hayan extendido. Una de las posibles razones para esto es el hecho de que no contemplen la definición del PFQ *no manual*. Otra posible razón es que, en el momento en que se presentaron las notaciones SEA y Szczepankowski, existieran otras notaciones más extendidas y con mayor capacidad representativa. En particular, el uso de Szczepankowski a nivel internacional se ve penalizado por el hecho de que toda su documentación esté en polaco.

El sistema de síntesis *SignSynth* desarrollado por Grieve-Smith [40] utiliza la notación ASCII-Stokoe para la definición de signos. Dado que el primer paso a realizar, si se utiliza una notación alfanumérica, es el procesado de la cadena de texto, la elección que realiza, al utilizar el lenguaje de programación PERL para procesar esta cadena es la más acertada. Posteriormente, como resultado de dicho proceso se genera la animación utilizando VRML. Sin embargo, es el propio Grieve-Smith quien admite que dicha notación no es la más idónea para la descripción de la lengua de signos.

El uso de la notación Szczepankowski por Francik y Fabian [35] para la síntesis de signos es definida por el autor como un punto intermedio entre la comodidad (aunque parca en detalles) de la notación SignWriting y la excesiva rigurosidad de HamNoSys. Independientemente de la sencillez del avatar utilizado, la síntesis se ve penalizada por las limitaciones de la notación al no definir el PFQ *no manual*.

Como se ha manifestado anteriormente, el uso de una notación gráfica en el ordenador presenta una dificultad añadida debido a su manejo. Independientemente de esto y debido a la capacidad representativa de HamNoSys, Caridakis *et al.* [16] plantean el uso de esta notación como método para describir los signos. Estas descripciones en HamNoSys son convertidas a secuencias en lenguaje STEP, lo que permite definir la animación de un avatar en H-ANIM.

La notación HamNoSys permite una descripción precisa de los signos, aunque su carácter gráfico dificulta su manejo con el ordenador. La primera adaptación formal de dicha notación para ser utilizada en el ordenador fue descrita por Elliott *et al.* [29] y denominada HML. Esta notación se planteó únicamente como paso previo a la notación SiGML, dado que se trata de una simple codificación de HamNoSys en XML manteniendo el árbol sintáctico de HamNoSys. Independientemente de su carácter de solución parcial ha sido utilizada por Sapountzaki *et al.* [91] para la definición de signos. Al igual que Caridakis *et al.* [16], se utiliza el lenguaje STEP para generar la animación del avatar, aunque en este caso a partir de la definición obtenida en HML.

La notación SiGML es uno de los resultados [42, 29] más extendidos del proyecto ViSiCAST. Esta notación se basa en la notación HamNoSys, pero no constituye una simple recodificación de dicha notación, como es el caso de HML; SiGML plantea una serie de modificaciones sobre HamNoSys. La principal de estas modificaciones permite definir el PFQ *no manual*, que no estaba descrito en la versión de HamNoSys existente en ese momento. SiGML introduce además la posibilidad de definir la sincronización de este último PFQ con el resto de PFQ contemplados en HamNoSys. SiGML es una notación que permite una descripción morfológica exhaustiva de un signo. Además de toda la información relacionada con la descripción de los signos, SiGML también incluye información relativa a la representación 3D, como por ejemplo el avatar que se desea utilizar o la descripción de la cámara [62, 28].

La última versión disponible de SiGML se basa en HamNoSys3. Elliot *et al.* [27] justifica el uso de esta versión basándose en que las principales modificaciones que introduce HamNoSys4 se centran en una mejor definición del PFQ *no manual* que, como se comentó antes, SiGML ya había implementado.

La misma estrategia aplicada para el HamNoSys-SiGML se aplicó para el SignWriting. Rocha [87] diseñó un método para representar los mensajes escritos en SignWriting en formato XML. La principal diferencia entre el SWML y el SiGML radica en la formalización que posee HamNoSys y de la que carece SignWriting. Es por esto que la notación SWML no sólo representa los elementos iconográficos presentes en el mensaje, sino que además describe la disposición espacial de dichos iconos, que es la parte más importante de la descripción del signo. El único proyecto que utiliza

esta notación como descripción de los signos ha sido presentada por Papadogiorgaki *et al.* [84]. Por lo tanto, aunque SignWriting es una notación que ha alcanzado una gran difusión entre personas sordas, SWML, su adaptación a los ordenadores, no ha seguido dicha tendencia. Junto con la falta de formalización, otro de los factores que ha podido influir en este hecho es que cuando fue desarrollado, la notación SiGML ya se venía utilizando desde algunos años atrás.

Todas las notaciones que hemos contemplado en este punto comparten una serie de deficiencias. La primera de ellas es que no contienen información de carácter temporal que indique la duración total del signo o durante cuánto tiempo debe permanecer una posición estática. Se podría argumentar que el texto tampoco contiene este tipo de información y, aun así, se han creado sintetizadores de voz. La diferencia entre la LS y la voz, es que la última puede ser estudiada con mayor facilidad y además lo ha sido durante más tiempo. Además existe otra gran diferencia entre la voz y la LS: el número de parámetros que se deben contemplar. La voz contiene un único parámetro⁹, los fonemas; definiendo la voz como una construcción secuencial de fonemas. Como se ha presentado en la Sección 2.1, la LS consta de siete parámetros. Es más, seis de los parámetros deben duplicarse en caso de los signos bimanuales. Por tanto, no sólo es conveniente tener información temporal para la realización de un signo, sino que además es importante definir la sincronización de los distintos PFQ.

3.2.2.2. Bases de datos relacionales

En el punto anterior se ha hecho un recorrido por las distintas notaciones que se utilizan para describir la estructura interna de los signos. Si bien ésta es la opción más extendida, no por ello es la única posible. La otra aproximación que se ha estudiado es el uso de una base de datos relacional para la descripción ya sea de los signos o de la animación. Ciertos autores no distinguen entre una base de datos relacional, que posee una estructura interna compleja, y una base de datos descrita como una lista de entradas en las que se almacenan los signos.

La única aproximación que utiliza una base de datos relacional orientada a la síntesis de signos fue propuesta por Furst *et al.* [37]. La estructura de la base de datos relacional propuesta por Furst *et al.* se divide en dos tablas: la primera de ellas recoge la información necesaria para la descripción de las *configuraciones* del signo; la segunda, almacena toda la información considerada necesaria para la generación de los signos: *orientación*, *posición*, *movimiento*. El contenido de estas dos tablas recoge tanto información geométrica como lingüística relativa a los signos.

⁹Si no consideramos modificaciones prosódicas o la síntesis de un idioma tonal.

Aunque el tema que se trata en esta tesis doctoral se centra en el proceso de síntesis de LS, no se puede dejar sin mencionar la base de datos *SignPhon* presentada por Crasborn *et al.* [25, 26]. Esta base de datos no se centra en la síntesis, sino en el estudio de las propias LS. Dicho estudio abarca desde niveles internos tales como la estructura del signo, hasta niveles superiores en los que se analiza la semántica de la información recogida. Dentro de las nueve tablas que componen la base de datos SignPhon se almacena también información no relacionada con la estructura de la LS. Esta información relaciona los datos obtenidos con las fuentes originales de información de las que se extrajo. Dado que estas fuentes son a menudo vídeos grabados, SignPhon se convierte en una fuente de datos etiquetados.

Antes de finalizar esta sección es preciso destacar el trabajo realizado por Chittaro *et al.* [18], donde se plantea la definición de animaciones mediante una base de datos relacional que se organiza en tres niveles lógicos: El nivel inferior recoge la definición de una postura estática del avatar, el nivel intermedio recoge animaciones descritas como secuencias de elementos del nivel inferior; por último, el nivel superior describe animaciones más complejas como concatenación de animaciones sencillas del nivel intermedio. La definición del contenido de los tres niveles se realiza mediante de la herramienta H-Animator [14]. Aunque esta aproximación de Chittaro *et al.* utiliza una base de datos relacional para la definición de la animación, la estructura de la misma no está fundamentada en ningún estudio lingüístico de las LS.

3.3. Síntesis de clasificadores

En el punto 2.2.1.3 se ha expuesto cómo las personas sordas pueden describir escenas sin necesidad de utilizar signos. Los clasificadores icónicos (o predicados clasificadores) utilizan un gran componente mímico para dicha tarea. La dificultad que presenta la gestión automática de estos elementos radica en el proceso de abstracción que realiza un signante al utilizar estos clasificadores. Dicha abstracción tiene una importante dependencia semántica, intrínsecamente ligada a los procesos mentales del signante. Por último, cabe destacar que este aspecto de las LS no ha sido suficientemente estudiado para postular una teoría que cubra todos los posibles aspectos de los clasificadores icónicos.

El único autor que ha trabajado en la traducción y síntesis automáticas de estos clasificadores es Huenerfauth [48, 49]. Para ello ha utilizado la teoría formulada por Liddell [69], en la que el signante mantiene una representación mental de la escena que está describiendo. La representación de la escena puede adaptarse a un conjunto de plantillas que, aunque extenso, puede considerarse finito. Esto llevó a Huenerfauth



Figura 3.10: Simulación de una escena escalada al espacio de signado.

a combinar dos aproximaciones para lograr la representación de los predicados clasificadores.

La primera de estas aproximaciones se basa en el uso de la información obtenida de la aplicación AnimNL [6, 13, 92]. Esta aplicación tiene por objeto la generación de escenas tridimensionales a partir de descripciones en lenguaje natural. El modo de funcionamiento es equivalente a otras aplicaciones similares [24, 19, 93] que se basan en el análisis del texto para obtener las relaciones espaciales de los objetos descritos para generar posteriormente la escena 3D utilizando una librería de modelos tridimensionales. La aproximación de Huenerfauth consiste en utilizar estos elementos para guiar las manos del avatar. El primer paso implica el escalado de la escena generada por la aplicación al espacio de signado (Figura 3.10). Lógicamente, para realizar la representación del clasificador icónico no es necesario visualizar los objetos de la escena, salvo el propio avatar. Por esta razón se establece que todos los objetos sean invisibles y utiliza la situación de los mismos para definir las posiciones que deben adoptar las manos. Como puede observarse resulta una aproximación literal al modelo de Liddell ya que existe una escena invisible que es representada por el avatar.

Según Huenerfauth, el uso de este modelo es sobre-generativo, por lo tanto mejorable. Para lo cual añade el uso de un conjunto de diferentes plantillas abstractas

basándose igualmente en lo propuesto por Liddell [69], lo que constituye la 2ª aproximación. Estas plantillas recogen información sobre PFQ como la *configuración* o la *orientación* que deben utilizarse según la naturaleza del elemento representado.

3.4. Evaluación de mensajes sintéticos

Al igual que sucede con los sintetizadores de voz, no existe un método objetivo para evaluar la calidad e inteligibilidad de la síntesis. Dicha evaluación debe ser llevada a cabo por expertos en la LS que se está evaluando, preferiblemente signantes nativos.

La síntesis de signos es una técnica relativamente reciente y no se ha establecido un protocolo estándar para llevar a cabo su evaluación. Por esta razón, distintos grupos han planteado diversos escenarios para la evaluación de los signos. Para su mejor comparación la descripción de las diferentes aportaciones a la evaluación de las LS se han organizado siguiendo un orden ascendiente en la complejidad semántica de las unidades evaluadas.

San Segundo *et al.* [90] propone, para medir la comprensibilidad de un avatar bidimensional, la evaluación del diccionario dactilológico. Esta evaluación se realiza presentando unidades independientes, con lo que no es posible obtener información de contexto. La evaluación, realizada por diez personas, consta de dos etapas distintas entre las cuales se informa a los usuarios de los resultados obtenidos en la primera de ellas. Las tasas de acierto obtenidas son de un 70 % en la primera y un 100 % en la segunda. Los resultados obtenidos ponen de manifiesto la mejora producida en la comprensión cuando existe un periodo de adaptación al modo de signar que utiliza el avatar.

El siguiente nivel de evaluaciones, siguiendo el criterio de complejidad semántica, se centra en la medición del grado de comprensión de signos aislados. Kennaway *et al.* [63] comienza la evaluación del avatar vGuido (ver 3.2.1.2) comparando las tasas de reconocimiento de signos aislados entre los resultados obtenidos en Gran Bretaña utilizando Lengua de Signos Británica (BSL) y los resultados obtenidos en Holanda utilizando Lengua de Signos Holandesa (SLN). Para este conjunto de evaluaciones contó con un grupo de entre seis y diez personas sordas para los que la LS es su principal modo de comunicación. Los resultados de la evaluación inicial aportaron un 70 % de aciertos en BSL y un 75 % en SLN. En una evaluación final, tras mejorar sustancialmente la componente no manual de la representación, la tasa de aciertos del experimento repetido en Holanda, obtuvo una porcentaje de aciertos del 95 %. Una disminución en la tasa de reconocimiento no es siempre atribuible a un signado

deficiente, sino que puede deberse a la gran variabilidad regional en la definición de los signos.

Una vez evaluados los signos como elementos aislados, el siguiente paso en complejidad es la evaluación de frases completas y grupos de frases. Sobre los experimentos anteriores, en los que únicamente se mide la calidad de la síntesis, la evaluación de frases añade una evaluación gramatical implícita, que debe ser considerada ya que no es dependiente en exclusiva de la síntesis. Kennaway *et al.* [63] plantea el uso de un avatar animado mediante captura de movimientos (avatar Visia2) como referencia de comparación para el avatar animado por definición paramétrica (vGuido). La tasa de reconocimiento que se alcanzó para frases en BSL fue del 46 % y del 35 % para el SLN. Estos resultados se mejoraron tras incorporar las modificaciones descritas en el párrafo anterior, alcanzando una tasa del 58 % para SLN y del 71 % en Lengua de Signos Alemana (DGS) permitiendo repetición de las frases.

Tras la evaluación de frases y párrafos generados por el sintetizador, la siguiente evaluación mide la complejidad que encuentran las personas sordas al interactuar con estas herramientas. Cox *et al.* [23] evaluó el uso de un avatar basado en captura de movimientos como asistente en una oficina de correos. Los resultados de reconocimiento de signos individuales en las frases alcanzan una tasa de reconocimiento del 81 % y del 61 % para frases completas. La gran varianza entre los resultados obtenidos por los seis participantes en el proceso, se atribuye (igual que hizo Kennaway) a la gran varianza dialectal de las LS. Estos experimentos permitieron comparar la facilidad con la que una persona sorda realizaba una transacción según le atendiera un dependiente o un avatar. Los resultados indicaron que la duración de la operación fue el doble cuando se utilizó el avatar y el grado de satisfacción de los clientes sordos disminuyó. En una reflexión más detallada de los resultados, Cox *et al.* [22] anota que la aceptación del sistema fue mayor entre los dependientes al permitir una comunicación con personas sordas, que antes no era posible. Por el contrario la opinión de los clientes sordos reflejó una preferencia por la interacción con los dependientes en vez de con el avatar. La mayor penalización sobre el avatar fue la baja calidad de las expresiones faciales. Es de esperar que si se mantiene la tendencia observada en los experimentos de Kennaway *et al.* [63] tras la mejora del componente no manual los resultados mejoren.

Por último, el mayor nivel de complejidad semántica es presentado por los clasificadores. Como se ha visto en la Sección 3.3, la única aproximación a la síntesis automática de estas unidades es la realizada por Huenerfauth *et al.*, quien realiza una síntesis de clasificadores icónicos. Además, son la única referencia que ha presentado un modo de evaluar la calidad y comprensibilidad de estas unidades.

El planteamiento experimental de Huenerfauth *et al.* [50, 51] y de Huenerfauth [49] debe permitir la evaluación de una unidad semántica utilizada para la descripción de una escena. La evaluación de los clasificadores sintetizados se centra en un conjunto de animaciones que presentan escenas en las que hay movimiento de algún objeto de la escena o en las que se sitúan espacialmente una serie de objetos. Para realizar la evaluación de la síntesis de clasificadores icónicos se utiliza un conjunto de vídeos generados por ordenador que representan tanto la escena descrita por el avatar, como escenas similares que actúan de distractores. Los tests fueron realizados por quince personas utilizando diez frases para las cuales se ofrecían tres opciones, de las cuales una era correcta. Los resultados mostraron una mejoría del 20 % con respecto al uso de inglés signado¹⁰.

¹⁰Se denomina inglés signado, o español signado, a una secuencia signada que sigue la estructura gramatical del idioma oral. Por lo tanto no se debe confundir el concepto de idioma signado con lengua de signos.

Capítulo II

Descripción del Sintetizador Paramétrico

Apartado 4

Formalización

Tras la lectura del capítulo anterior, el lector habrá observado que no existe una teoría única y asentada que describa la LS. Esta diversidad se hace extensiva tanto a los sistemas de notaciones para la representación de los signos como a las aproximaciones técnicas para llevar a cabo la síntesis de LS. En este apartado se realiza una revisión de todo lo visto anteriormente y se presenta la formalización necesaria para plantear un sintetizador de lengua de signos que abarque, con la suficiente flexibilidad, los aspectos más relevantes de la LS.

La estructura de este apartado es semejante a la del Apartado 2. Veremos cómo se deben definir y relacionar los PFQ para describir los signos *establecidos* y el diccionario dactilológico. Una vez formalizadas las dos unidades de las LS independientes del mensaje, nos centraremos en el tercer tipo de unidades, los *clasificadores*. Por último, se presenta la aproximación que se ha seguido en este trabajo para la síntesis de variaciones emocionales.

4.1. Formalización de los PFQ

Un sintetizador de voz utiliza una secuencia de fonemas, los elementos básicos del habla, para componer las palabra del idioma con el que se trabaja. Si optásemos por abordar la síntesis de un idioma tonal como es el caso del chino, esta secuencia debería complementarse con una segunda destinada a definir el tono.

Como se ha visto en la Sección 2.1, la LS contempla siete PFQ independientes. Esto implica siete secuencias independientes. Además, seis de estas secuencias se duplican al describir las dos manos. Algunos de estos parámetros sólo contemplan un número

reducido de unidades, como es el caso del *plano*, pero otros parámetros corresponden a elementos formantes con un número elevado de posibilidades: la *configuración* presenta más de setenta unidades sólo para la LSE; la *orientación* de la mano puede definirse por doscientos ocho valores.

Las secuencias de PFQ deben definirse de tal manera que sea fácil establecer la sincronización entre unidades de las distintas secuencias. Para facilitar esta tarea se ha optado por utilizar una escala porcentual relativa a la duración del signo para la definición de las secuencias de PFQ. Las aproximaciones existentes consideran el signo dividido en etapa inicial, media y final. Utilizar una escala de cero a cien permite definir las secuencias que componen un signo con mayor grado de detalle. Así mismo, esta definición separa la descripción de las secuencias de PFQ de la duración total del signo, lo que permite modificar la duración del signo, con fines prosódicos, de manera sencilla.

El planteamiento que se ha seguido se basa en la independencia de cada uno de estos PFQ, dado que afectan a distintas partes de la anatomía del intérprete. La única herramienta o modo de trabajo que permite trasladar este hecho a un avatar virtual es una estructura jerárquica, un esqueleto. A continuación se presenta la formalización que se ha realizado con cada uno de los PFQ que definen la LS:

- *Configuración* de la mano. Las distintas formas que adopta la mano, según la flexión de los dedos, es el primero de los PFQ que se define en la lengua de signos. La definición de la *configuración* requiere establecer la rotación para cada una de las articulaciones de los cinco dedos. Este requisito supone almacenar quince rotaciones distintas para cada una de dichas *configuraciones*. La ventaja que nos aporta el uso de un esqueleto en animación es que, si se genera el conjunto de huesos correspondientes a una mano como una copia especular de la otra, la aplicación de idénticas rotaciones a ambas manos generará la misma *configuración* en las mismas.

Éste es uno de los PFQ que pueden definir variaciones en función del estado emocional del signante y de la prosodia aplicada al mensaje. Debido a esto se contempla el hecho de que cada una de las *configuraciones* definidas sea representada por conjuntos diferentes de rotaciones articulares, lo que se traduce en que la mano del avatar puede presentar una mayor tensión muscular o, por el contrario, relajación en la forma de los dedos. También hay que señalar que la capacidad de rotación en las articulaciones humanas varía con la edad. Si se deseara emular esta cualidad humana, ésta es la aproximación que debería utilizarse. En la Figura 4.1 se presenta una *configuración* en las tres variaciones definidas.

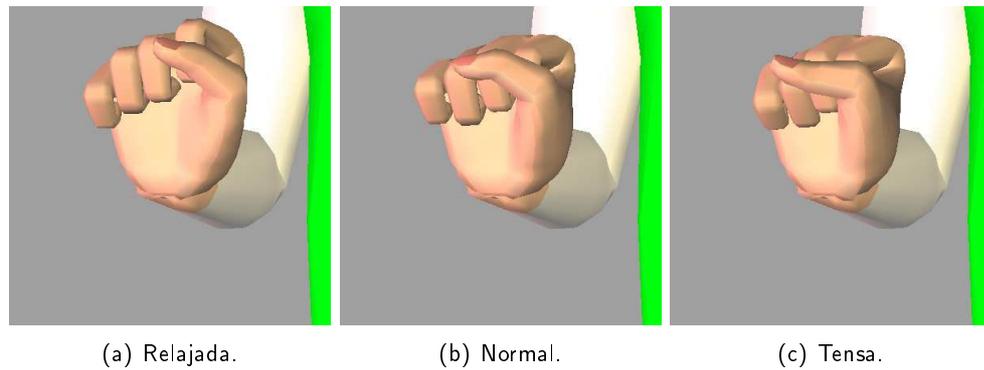


Figura 4.1: Ejemplo de definiciones paralelas de una misma *configuración*, en este caso de puño cerrado, en tres estados distintos de tensión muscular.

Puede considerarse que la forma que adquiere un dedo es independiente de la de los otros. Para cuantificar adecuadamente el espacio de almacenamiento requerido, se puede definir cada uno de los dedos por separado. Siguiendo esta aproximación, en un nivel intermedio entre la definición de una *configuración* y la descripción de las formas de los dedos, se establece una relación para determinar qué formas adquiere cada dedo en dicha configuración.

Aunque la aproximación anterior permite reducir el espacio de almacenamiento, en este trabajo se ha optado por no utilizarlo. Esta decisión se ha tomado con el objetivo de permitir un mayor grado de variación entre las formas de las manos, modelando de manera independiente todos los *Queiremas* y de este modo ganar en naturalidad. Por esta razón, para cada una de las *configuraciones* definidas, se han almacenado las rotaciones de todos los dedos en conjunto.

- *Orientación* de la muñeca. La *orientación* se define como la dirección que toman la palma y el dedo índice de las manos. Aparentemente, este elemento, sólo implica definir la rotación para el hueso del esqueleto que simula la muñeca, pero existe un factor importante derivado de la jerarquía natural de movimientos del brazo. La orientación absoluta de la muñeca se modifica debido a variaciones en la orientación del hueso del antebrazo. Éste, a su vez, se ve alterado por variaciones en la orientación del brazo y así sucesivamente. Por lo tanto, al definir el PFQ *orientación* se debe compensar la influencia que tienen brazo y antebrazo sobre la orientación de la muñeca.

La aproximación seguida por el resto de autores es incluir una restricción a la orientación de la muñeca en los cálculos de Cinemática Inversa (IK). Esta restricción adicional requiere una modificación en el algoritmo de IK, que implica mayor tiempo de cálculo y complejidad. La compensación de la muñeca se debe realizar cada vez que la orientación del brazo o del antebrazo varía.

Esto se produce cada vez que cambian los valores de los PFQ *localización*, *plano* y *movimiento*. Por consiguiente, se está supeditando la gestión del PFQ *orientación* a la gestión de estos otros tres PFQ. Este cálculo puede omitirse mediante el uso de una desvinculación en la jerarquía que elimine la herencia de las transformaciones de rotación de los nodos superiores.

Para ver un ejemplo de este planteamiento se presentan dos imágenes, que representan un sistema esquelético de tres huesos. La orientación del segundo hueso se altera al rotar el primero (Figura 4.2(a)). Para evitar esta herencia en la transformación de rotación, establecemos que la orientación del segundo hueso sea independiente de la de su antecesor (Figura 4.2(b)). Si aplicamos esto a la muñeca del avatar (ver Punto 6.1.2.1) podremos definir el PFQ *orientación* con total libertad y sin tener que modificar su gestión por una restricción del modo de implementación.

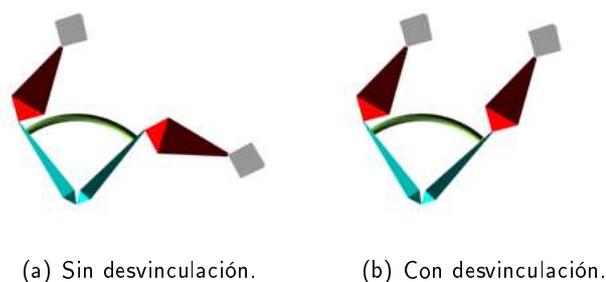


Figura 4.2: Ejemplo de desvinculación en la jerarquía de los huesos para la transformación de rotación.

Como se verá en el Punto 6.1.2.1, la estructura de la muñeca utilizada es distinta de la de otros sistemas y de la humana, lo que ha permitido que el hueso de la muñeca sólo herede la posición de su hueso padre. De esta manera, la orientación que tengan ambas manos no se verá influida por las rotaciones del hombro y del codo, con lo que puede considerarse como un elemento completamente independiente.

Esta definición del PFQ *orientación* presenta un inconveniente: las transiciones entre distintas *orientaciones* no siempre van a cumplir las restricciones anatómicas de la muñeca ya que, por defecto, se realizan por el arco esférico más corto. Para evitar este efecto se ha descrito cómo deben ser las transiciones entre dos *orientaciones* dadas, presentándose tres posibles soluciones: a través del arco esférico más corto, del arco más largo o pasando a través de una *orientación* intermedia dada. En este trabajo se han definido, hasta el momento, la mitad de todas las *orientaciones*; para cada dirección del espacio, sólo se han definido cuatro de las ocho orientaciones posibles.

- *Localización y Plano.* El tratamiento para la formalización de estos dos PFQ, se ha realizado de manera conjunta debido a que existe una relación directa entre ambos. Como se ha visto, el PFQ *localización* hace referencia a la posición de la mano en el plano frontal del intérprete y el PFQ *plano* especifica la distancia horizontal entre la mano del signante y su cuerpo. Por lo tanto, el PFQ *localización* aporta las componentes (x, y) y el *plano* aporta la componente (z) , logrando de esta manera obtener una coordenada en el espacio tridimensional (ver Figura 4.3). El hecho de que en este trabajo se especifique la formalización de manera conjunta no debe conducir al error de pensar que no son PFQ independientes. Debe quedar claro que son dos PFQ que pueden ser manejados y definidos de manera independiente pero que resultan complementarios a la hora de definir la posición en el espacio.

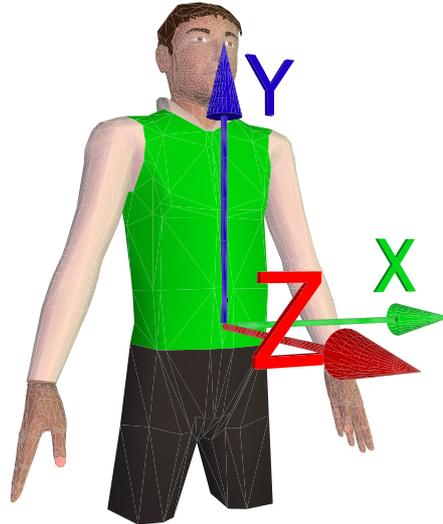


Figura 4.3: Sistema de coordenadas de la escena tridimensional.

- *Localización.* En el momento de plantear la formalización de este PFQ debemos recordar el modo en que se define en la LS. Para definir la posición en la que se articula un signo, se hace siempre referencia a una parte del cuerpo, brazos o rostro del intérprete. Esto implica que las referencias son relativas y por lo tanto no se puede utilizar un valor estático para definir una determinada *localización*. Por ejemplo, si definimos un signo que se articula en la nariz podríamos obtener a priori las coordenadas de la nariz y almacenarlas. Pero, si en el instante en el que la mano se sitúa frente a la nariz el cuerpo está inclinado hacia un lado o la cabeza no mira al frente, las coordenadas que tenemos almacenadas no reflejarían la ubicación correcta de la nariz en ese momento.

La solución planteada por otros autores define las posiciones anatómicas mediante un conjunto de vértices de la malla. Para obtener la posición de una parte del cuerpo del avatar se toman las coordenadas de estos vértices. Aunque esta aproximación es válida (ya que la precisión con la que se obtiene las coordenadas de la referencia anatómica es máxima), consideramos que presenta algunos inconvenientes.

Por una parte la malla del avatar es la parte visual del mismo y el proceso de cálculo de las deformaciones definidas por el esqueleto es costoso. Por lo tanto, si se depende de estas deformaciones para la síntesis (se necesita obtener las coordenadas de los puntos de referencia) se está utilizando un alto porcentaje del proceso en una serie de costosas operaciones sin resultado visual para el usuario. Por otra, se hace depender el proceso de síntesis del diseño artístico del avatar (normalmente habrá sido realizado por un diseñador gráfico sin conocimiento de los requisitos de la síntesis de signos). Supongamos que se desea diseñar un avatar lo más sencillo posible para permitir su animación en tiempo real en un dispositivo con recursos gráficos limitados. Si el conjunto de vértices con los que se ha definido una parte del avatar es menor al de las referencias anatómicas de esa parte, la síntesis no será correcta. Por ejemplo, si se define el tronco del avatar mediante un prisma de cuatro caras verticales, compuesto por ocho vértices, no permitirá definir todas las localizaciones que se definen en el torax del signante.

Este conjunto de inconvenientes nos llevó a plantear el uso del esqueleto para esta tarea eliminando así toda dependencia de la definición de la malla. Utilizando el esqueleto para definir las referencias anatómicas, ya no es necesario describirlas mediante el identificador de un vértice, sino por el nombre de un hueso¹. Debido a que, los huesos que definen la animación del avatar son insuficientes para cubrir todas las referencias necesarias, se ha incluido un nuevo tipo de huesos denominado «huesos de localización» (ver Punto 6.1.2.3). Esta aproximación permite llevar a cabo el proceso de síntesis con total independencia de la definición de la malla, lo que evitará los inconvenientes antes mencionados. Presenta sin embargo, el inconveniente de una precisión ligeramente inferior al uso de la malla en la definición de la posición de las referencias anatómicas.

Adicionalmente, hemos planteado la definición del PFQ *localización* mediante una coordenada y un hueso. Esto permite la posibilidad de añadir

¹Por definición, la posición de un hueso es la posición de su punto de pivote.

un desplazamiento adicional a partir de las coordenadas en las que se localiza un hueso. De esta manera se contemplarían situaciones tales como ‘bajo la barbilla’, ‘sobre la cabeza’ o ‘a un lado del hombro’.

- *Plano*. El plano se ha formalizado mediante valores estáticos y dependientes de la escala del modelo. Además de los cuatro planos que se definen en la teoría se han añadido planos adicionales para flexibilizar la síntesis. Entre estos nuevos planos destaca la definición de un plano anterior al cuerpo del signante y un plano nulo.

La existencia de un plano anterior al cuerpo del signante se utiliza en signos en cuya articulación alguna de las manos termina o comienza en esa posición.

El concepto de plano nulo se ha desarrollado para definir la circunstancia en la que se desea que haya contacto entre dos puntos del cuerpo, lo que formalmente consiste en definir que la distancia horizontal entre la mano y la *localización* es cero. Para ello la coordenada z debe ser obtenida del «hueso de localización» que se utiliza para describir el PFQ *localización*.

- *Movimiento*. Los *movimientos* en las LS han sido definidos como variaciones de la posición de las manos a lo largo del tiempo. Cada movimiento está definido por una secuencia de variaciones de posición que deben alcanzarse en determinadas fracciones de la duración del movimiento. Este modo de definición del mismo, permite establecer, para una misma trayectoria, distintos niveles de aceleración (Figura 4.4).

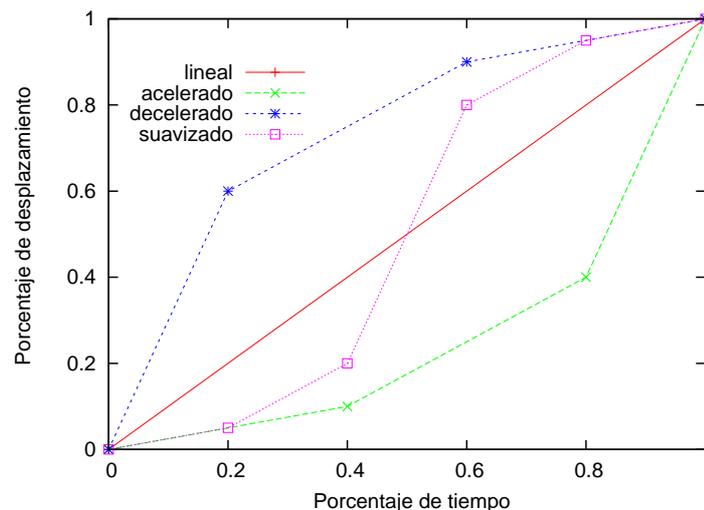


Figura 4.4: Estas gráficas muestran distintas definiciones de un mismo movimiento rectilíneo. La duración y amplitud del movimiento es igual en todas ellas, lo que varía es la aceleración con la que se ejecutan cada una de las etapas intermedias en cada alternativa.

La duración de un movimiento se establece por dos instantes definidos como porcentajes de la duración del signo. Esta aproximación implica determinadas ventajas con respecto a lo definición de una duración absoluta predefinida para cada movimiento. La primera de ellas es la facilidad con la que es posible modificar la duración de los signos de manera dinámica. La segunda es que no es necesario establecer una duración mínima para un signo en función de los movimientos de los que consta disponiendo, por tanto, de mayor flexibilidad a la hora de describirlo.

El *movimiento* es también uno de los PFQ susceptibles de modificación debido a la prosodia y a las emociones, por lo que se ha contemplado la definición de variaciones para un movimiento atendiendo a factores prosódicos. El hecho de que la LS define distintos grados de tensión en la realización de un *movimiento* como parte de los factores diferenciadores del significado de un signo, complica la definición de alteraciones prosódicas. Por lo tanto, es necesario plantear dicha modificación emocional manteniendo dichas variaciones (Figura 4.5).

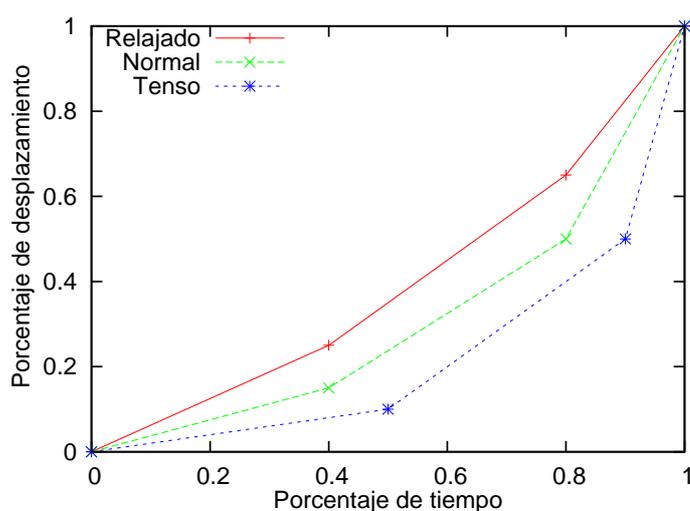


Figura 4.5: Distintas variaciones propuestas para un movimiento rectilíneo y acelerado, en función de una modificación emocional.

- *Punto de Contacto*. Este PFQ se define como el punto de la mano, normalmente la articulación de algún dedo, que va a entrar en contacto con otra parte del cuerpo o bien debe alcanzar una determinada posición. Para describir este PFQ se utiliza el nombre del hueso relevante de la mano, ya sea un «hueso de animación» o de «localización».
- *Componente no manual*. El último de los PFQ de la LS hace referencia a la postura del cuerpo, los hombros, la cabeza y todas las expresiones de la cara.

Como se señaló anteriormente, la tendencia más común es definir las expresiones faciales utilizando la técnica de *morphing de malla*. En nuestro caso, para simplificar el proceso y unificar el método de animación, tanto los movimientos articulares como las expresiones faciales se han diseñado para ser definidas por rotaciones de los huesos correspondientes en el esqueleto definido, por lo tanto, son tratados igual que las rotaciones de muñeca o las configuraciones. La definición de los huesos que controlan la animación facial, se explicaría en detalle en la Subsección 6.3.2.

Para facilitar la composición de una mayor variedad de expresiones faciales y optimizar los recursos de almacenamiento se ha optado por mantener la independencia entre los distintos elementos de la cara en su gestión. A modo de ejemplo, podemos decir que el hecho de que las cejas estén elevadas o fruncidas es independiente de que la boca o los párpados estén cerrados o abiertos. Pero, por el contrario, la forma que adquiere la boca es única y debe ser tratada como una unidad, sin segregar la definición de los labios. Cada uno de estos grupos independientes han sido denominados en este trabajo como ‘expresiones simples’ y la agrupación de éstas, ‘expresiones complejas’. Los distintos grupos que se tienen en cuenta para el PFQ *no manual* son:

- cabeza
- pecho
- cintura
- ceja derecha
- ceja izquierda
- ojos (ambos ojos miran al mismo sitio)
- párpados del ojo derecho
- párpados del ojo izquierdo
- nariz
- mejilla superior derecha
- mejilla superior izquierda
- mejilla inferior derecha
- mejilla inferior izquierda
- lengua
- boca

Este es el último de los PFQ dependientes de información prosódica y síntesis de emociones. Pero, en contraposición a la *configuración* y al *movimiento*, no se pueden definir variaciones en las expresiones, simplemente se definen expresiones distintas.

Antes de finalizar esta sección, se debe señalar la existencia de cierto grado de ambigüedad en la definición de un signo. En particular, nos referimos a la posibilidad de definir determinados movimientos de las manos ya sea mediante el propio PFQ *movimiento* o bien mediante distintas entradas para los PFQ *plano* y *localización*. El resultado visual de ambas aproximaciones es el mismo, pero no la definición lingüística. Esta ambivalencia no influye en la síntesis de un signo aislado o no modificado.

Sin embargo, la aplicación de un *clasificador* dará lugar a distintas modificaciones según la definición que se haya hecho del signo.

4.1.1. Definición de los *signos reconocidos* y de la *dactilología*

La formalización anterior, realizada con los PFQ de la LS, permite su tratamiento independientemente unos de otros y de la implementación utilizada. Esto da lugar a la descripción de los *signos reconocidos* definiendo secuencias de PFQ de manera sencilla y sin ninguna restricción. De igual manera, se puede acceder fácilmente a sus descripciones en caso de que fuera necesario modificar alguno de los PFQ. El modo más sencillo de definir un mensaje en LSE es mediante una secuencia de glosas. Este es el objetivo que ha primado en la definición del formato de entrada al sintetizador (ver Apartado 8).

Los elementos del diccionario dactilológico se equiparan a signos monomanuales; la única diferencia entre estos elementos y los *signos reconocidos* es que no se tiene en cuenta el PFQ *punto de contacto* durante la síntesis. La definición de esta unidad y su almacenamiento no presentan diferencias con respecto a lo requerido por los *signos reconocidos*. Para facilitar la descripción de una secuencia dactilológica se ha optado por permitir la definición de la secuencia de letras que se desea deletrear el mensaje de entrada. Si se hubiera seguido el sistema de los *signos reconocidos*, tendríamos que definir también las glosas de estos elementos.

Los signos *establecidos* y los elementos del diccionario dactilológico se definen mediante PFQ. Este hecho implica que su definición pueda realizarse y almacenarse previamente a la síntesis. La síntesis de estos dos elementos sigue el mismo proceso: obtención de la definición almacenada de la unidad, gestión de los PFQ y generación de las pistas de animación. Esta secuencia de operaciones todavía es independiente del mensaje.

4.2. Formalización de los *clasificadores*

En la Subsección 2.2.1 se ha presentado la clasificación morfológica que Herrero Blanco [46] hace de los *clasificadores* en la LSE. Dado que estas unidades de la LSE poseen una sólida base semántica, no pueden ser descritos con anterioridad al proceso del mensaje. Los cuatro tipos de *clasificadores* descritos, afectan en distinto grado a la actual propuesta de definición del mensaje signado (secuencias de glosas de los

signos reconocidos). A continuación se discute la repercusión que cada construcción clasificatoria implica para la síntesis de la LSE.

- Los *nombres clasificatorios* son construcciones compuestas por dos *signos reconocidos*, signados de manera consecutiva con el objetivo de definir un nuevo concepto. Se definió que esta construcción es morfológicamente equivalente a una secuencia de dos signos. Esto implica que su síntesis no se distingue de la síntesis de cualquier otra pareja de signos, con lo que no altera el planteamiento actual de síntesis y definición del mensaje.

En el caso de que se decidiera disminuir el tiempo de transición entre estos dos signos para enfatizar esta construcción, debe contemplarse la posibilidad de definir específicamente dicho tiempo. Esta definición del tiempo sólo puede hacerse en la descripción del mensaje a sintetizar.

- Las construcciones clasificatorias *introflexivas* implican el uso de la definición de alguno de los PFQ de un signo en la ejecución de otro. Esta modificación implica la alteración del proceso por el que se recupera la información que describe un signo. Dicha alteración consiste en obtener la descripción del PFQ del signo clasificador y sustituirla por la del signo modificado. Una vez sustituida la nueva secuencia del PFQ, el proceso de síntesis continua de manera transparente gracias a que el manejo de todos los PFQ se hace de manera independiente.

La presencia de un clasificador *introflexivo* sólo puede definirse en la traducción. Por lo tanto, la existencia de este tipo de clasificación, la definición del signo modificado y del clasificador, deberán ser definidos en el mensaje de entrada al sintetizador.

- Los *clasificadores icónicos* son acciones mímicas estructuradas dirigidas a la representación de una escena mediante gestos. Independientemente de su estructura lingüística, la forma que adquieren las manos y su orientación pueden describirse mediante el uso de *configuraciones* y *orientaciones*. La diferencia estriba en que la posición en que se sitúan las manos debe definirse de manera continua, en contraposición al sistema discreto usado en la *localización* y el *plano*. Esta restricción también se aplica a la definición del punto al que el avatar debe dirigir la cara y la mirada. Además, se debe poder describir los movimientos de las manos con mayor flexibilidad que las unidades definidas en la síntesis de signos.

Para cubrir los requisitos establecidos para la definición de la posición de las manos y puntos a los que el avatar dirige su mirada, debe establecerse un nuevo

método de posicionamiento, así como, el modo en que se describen los movimientos en este nuevo sistema de referencia. Por último, las nuevas definiciones deberán procesarse en el momento de la síntesis, debido a que el movimiento del avatar dependerá de la información extraída del procesado del mensaje de entrada.

Para la descripción de estos clasificadores hemos definido un sistema de coordenadas, cuya definición debe ser independiente de la posición del avatar en la escena, luego el origen de coordenadas debe ser relativo al avatar y debe mantenerse independiente del tamaño del avatar. Con estas condiciones hemos impuesto que el origen de coordenadas del sistema sea la ubicación del punto anatómico del centro del pecho. Y que, las coordenadas en este nuevo sistema se especifiquen como fracciones de la longitud del húmero del avatar. La Figura 4.6 muestra la disposición en el sistema de coordenadas definido.

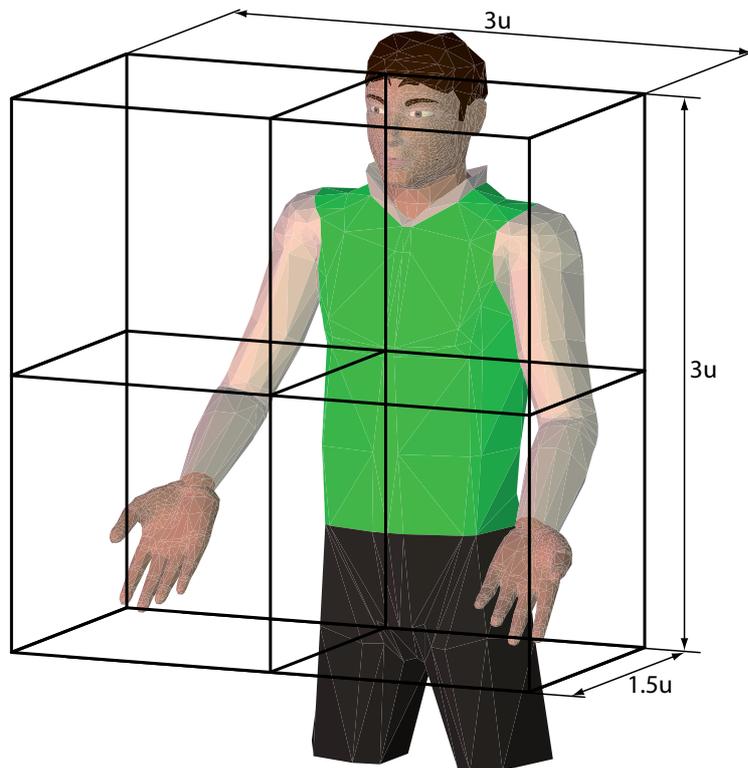


Figura 4.6: Espacio de coordenadas utilizado para la descripción de clasificadores icónicos y afijales. El valor de u es la longitud del antebrazo del avatar.

La definición de este sistema de coordenadas basado en fracciones de la longitud del húmero facilita así mismo el proceso de traducción. Se trata de que este modo de definir posiciones sea una representación directa del modelo que se utilice durante el procesamiento del lenguaje natural. Como hemos mencionado

anteriormente, un *clasificador icónico* representa una escena, por lo que los elementos que componen dicha escena deben ser situados en la misma. Un método sencillo para conseguirlo es utilizar una tabla tridimensional, cuyos índices son valores numéricos enteros. Mediante esta aproximación para la definición de los elementos de la escena, las relaciones espaciales existentes entre ellos puede establecerse fácilmente.

- El último tipo de *clasificadores* propuestos son los *clasificadores afijales*. En el momento de su definición se estableció que se podían considerar como una mezcla de los *clasificadores introflexivos* y los *icónicos*. Mientras la mano no dominante está ejecutando un *clasificador icónico*, la mano dominante está realizando un signo reconocido. Sin embargo, la mano dominante no realiza la ejecución normal del signo sino que modifica la parte final del signo para hacer que la mano dominante finalice el movimiento orientada y colocada en relación con la mano pasiva.

La consecuencia inmediata de los *clasificadores afijales* para la síntesis de signos es que necesitan definir la ejecución de un signo *establecido* de manera simultánea a la de un *clasificador icónico*. Este hecho se debe plasmar en la definición de la notación de entrada del sintetizador. Otro punto importante, que deriva de los *clasificadores afijales*, es el conjunto de modificaciones que se deben aplicar a la síntesis de un signo. A diferencia de los *clasificadores introflexivos*, la nueva definición no se obtiene de la descripción de otro signo *establecido*, sino que debe generarse dinámicamente en función de la definición de *clasificador icónico*. Además, esta modificación no se aplica a toda la secuencia de los PFQ afectados, sino sólo a la fracción final. Según se haya definido el signo, esto podría implicar añadir un elemento extra a la secuencia de los PFQ o sustituir alguno de los últimos elementos de la secuencia. En la aproximación seguida sólo se modifica la *orientación* de la mano dominante y su posición final (afectando a los PFQ *localización, plano y movimiento*).

4.3. Formalización de la síntesis emocional

A continuación se presentan los elementos de la LS que se ven alterados en función del estado emocional del signante y de la prosodia.

En la Sección 4.1 se ha expuesto el modo en el que los PFQ *configuración* y *movimiento* sufren variaciones dependiendo del componente emocional. De igual modo, el PFQ *no manual* se define para representar dicho estado emocional. Esto implica que se debe contemplar la definición de versiones paralelas de estos PFQ.

Al igual que la aceleración de los *movimientos* puede matizarse (ver Figura 4.5), también la aceleración con la que varían los distintos PFQ en un signo puede verse alterada. Esto implica no sólo distintas variaciones de los PFQ, sino también en la definición de secuencias de los mismos

La duración de un signo varía en función de la tensión con la que se ejecuta. En estados emocionales de mayor relajación, la duración se alarga, mientras que por el contrario, cuando existe una mayor tensión en la ejecución del signo, se acorta. Esta variación en los tiempos también se aplica a la transición entre signos consecutivos. Una frase tensa dejará menos tiempo entre signos y la representación relajada de la misma frase realizará la transición de un signo al siguiente de manera más pausada.

Las dos primeras modificaciones sólo pueden ser definidas con anterioridad a la síntesis, en la definición de los PFQ o en la de los signos *establecidos*. La modificación que afecta a la duración de un signo puede establecerse tanto en el momento de definir la variación emocional del signo como en tiempo de síntesis a través del mensaje de entrada. Ésta última se lleva a cabo mediante la definición de un modificador que altere la duración total del signo. Ya que la definición de las secuencias de PFQ se realiza mediante porcentajes de la duración total del signo, este escalado temporal no introduce un mayor nivel de complejidad. La modificación referente al tiempo de transición entre dos signos sólo puede ser definida en el mensaje de entrada del sintetizador, por lo que deberá procesarse en tiempo de síntesis.

4.4. Conclusión

Antes de continuar con los siguientes apartados de este capítulo, resulta interesante hacer un resumen de los requisitos que debe cumplir un sintetizador, desde nuestro punto de vista, a partir de lo estudiado en este apartado de formalización:

- Gestión independiente de los PFQ que permita el acceso a la definición de un signo *establecido* fácilmente, aplicado tanto al almacenamiento de información como a la definición del esqueleto.
- Definiciones alternativas de los signos *establecidos*, de las secuencias de PFQ que los forman y de los PFQ *configuración* y *movimiento*.
- Modificación de la estructura del avatar para eliminar la dependencia existente entre la definición de la malla y el proceso de síntesis lo que requiere incluir nuevos elementos en la estructura esquelética del avatar independientes de la estructura de animación.

- Una notación de entrada al sintetizador que permita:
 - Definir una frase o mensaje signado mediante las glosas de los signos *establecidos* que lo conforman.
 - Definir el idioma de una frase o de un signo en particular.
 - Definir una palabra para ser deletreada.
 - Definir la modificación en la síntesis de un signo posibilitando la sustitución de alguno de sus PFQ por la definición del mismo PFQ en otro signo.
 - Definir la modificación en la velocidad con la que se debe ejecutar un signo.
 - Definir el tiempo que debe emplearse en la transición entre unidades consecutivas, ya sea a nivel global o de manera particular entre dos signos.
 - Definir una postura estática del avatar sin utilizar referencias anatómicas y a partir de esta definición generar animaciones del mismo.
 - Definir ejecuciones paralelas de un signo y un clasificador icónico, indicando además que la realización del signo debe ser modificada en su parte final.

Los requisitos aquí definidos son la base que justifica el diseño e implementación del avatar (Apartado 6), la base de datos relacional (Apartado 7) y la notación de entrada a la que se le ha dado el nombre de HLSML (Apartado 8). La solución que proponemos para la adaptación del sintetizador a distintos dispositivos móviles es el uso de un diseño modular que permita asignar a la máquina del cliente sólo los módulos que pueda gestionar. Esta propuesta se recoge en el Apartado 5.

Apartado 5

Arquitectura del sintetizador

En este apartado se presenta la arquitectura diseñada para el sintetizador de LSE. Tras estudiar en la Sección 5.1 los distintos módulos que componen dicho sintetizador, en la Sección 5.2 veremos cómo distribuir la ejecución de cada uno de los módulos entre un servidor y un cliente de síntesis para adaptarse a los recursos de los dispositivos que puedan ser utilizados por el usuario.

5.1. Arquitectura del sintetizador

Como puede verse en la Figura 5.1, la arquitectura se divide en un conjunto de elementos independientes: «Parser de HLSML», «Base de Datos Relacional», «Servidor Web», «Síntesis de Gestos», «Render» y «Visualización». A continuación, se describen cada uno de estos modelos, haciendo especial hincapié en la labor del módulo de «Síntesis de Gestos» dentro del proceso global.

5.1.1. Parser de HLSML

El mensaje a sintetizar debe definirse en la notación HLSML, hemos desarrollado esta notación para describir, de manera sencilla, los elementos de un mensaje en LSE (las características de la notación HLSML se presentan con mayor detenimiento en el Apartado 8). El módulo «Parser de HLSML» verifica que el mensaje recibido cumple la estructura de un fichero XML y la definida en el documento DTD del HLSML¹. Una vez realizada dicha validación, la información contenida en el mensaje se desglosa en los siguientes elementos:

¹La implementación de este módulo utiliza un analizador de XML SAX incluido en Java.

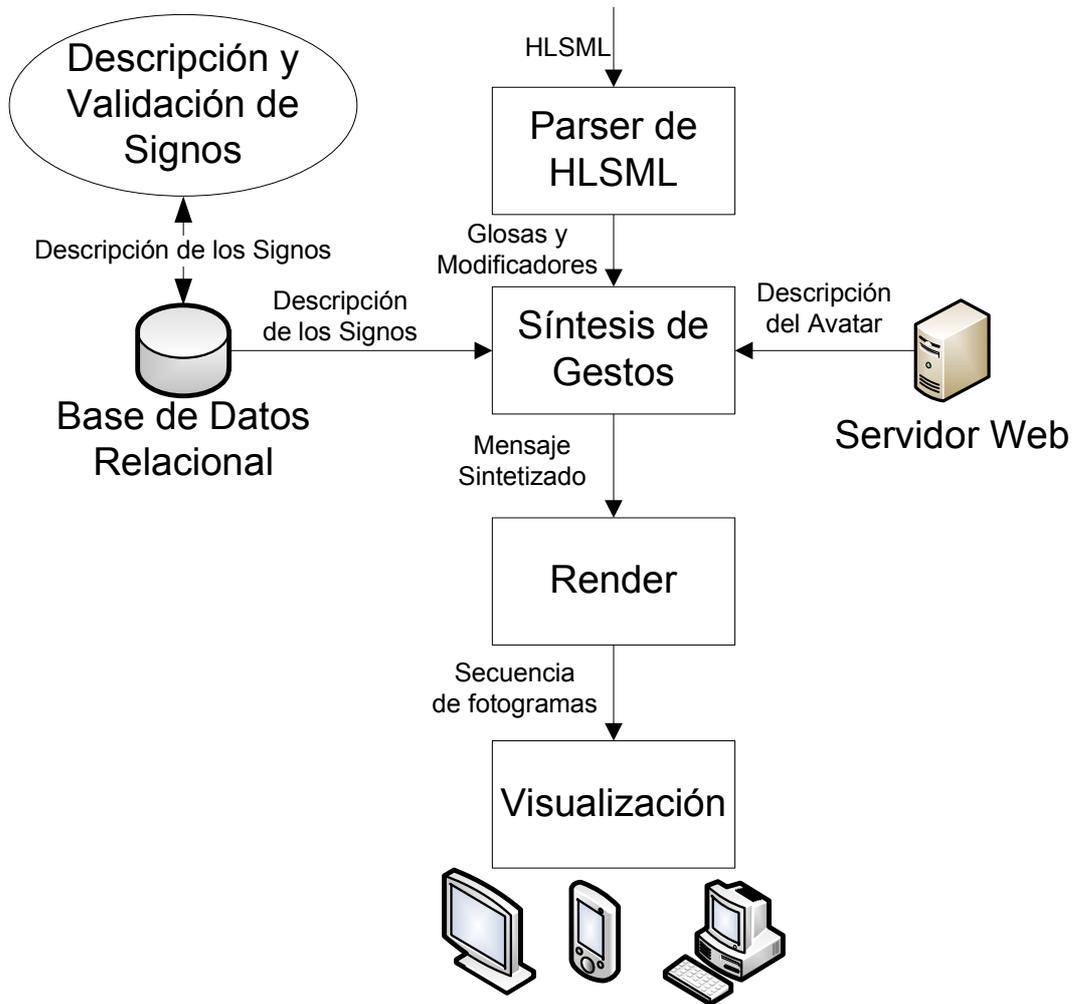


Figura 5.1: Arquitectura del Sintetizador de LSE propuesto.

- Secuencias de signos *establecidos* y palabras a deletrear.
- Clasificadores introflexivos que alteran el proceso de recuperación de información desde la base de datos relacional.
- Descripción de clasificadores icónicos.
- Información prosódica.
- Modificaciones sobre la animación.

5.1.2. Base de datos relacional

En este punto del trabajo es necesario presentar la función de la base de datos relacional utilizada en la arquitectura del sistema de síntesis (su estructura se describe

detalladamente en el Apartado 7). Se ha utilizado una base de datos relacional única y centralizada que almacena la descripción de todos los signos *establecidos*, así como el diccionario *dactilológico*. Para ello se establecen relaciones entre estas unidades y los PFQ que componen la LS. Los *clasificadores* también utilizan para su síntesis una parte de la información almacenada en la base de datos, en particular la definición de algunos PFQ.

El uso de una base de datos relacional con las características mencionadas presenta una serie de ventajas:

- Incluir la definición de los signos en la base de datos relacional permite la simplificación y reducción de tamaño en el mensaje de entrada.
- La información contenida en la base de datos permite un ajuste más preciso en la definición de la síntesis; este ajuste es realizado por un experto en LSE.
- Una base de datos centralizada permite supervisar la definición de los signos de manera única, ya que las alternativas basadas en la descripción en la notación de entrada no aseguran que la definición de los signos sea válida.
- La actualización del contenido de la base de datos relacional, ya sea al introducir nuevos elementos o mejorar los existentes, está disponible de manera instantánea para todos los usuarios.

5.1.3. Servidor web

En este trabajo se ha optado por utilizar la definición JSR-184 [59] como tecnología 3D. Este estándar define el formato de fichero «m3g» para la descripción de todos los elementos de la escena tales como la geometría, las luces, las cámaras, pistas de animación, los materiales de la geometría y las estructuras jerárquicas de huesos. Al inicio de la ejecución, la aplicación de síntesis descarga este fichero «m3g» desde el «Servidor Web» utilizando el protocolo HTTP. Este servidor contiene los ficheros de configuración necesarios para el manejo del esqueleto, la gestión del control de colisiones, el manejo de elementos auxiliares, etc.

5.1.4. Síntesis de gestos

El módulo de «Síntesis de Gestos» es el más relevante dentro del proceso de síntesis. Su cometido es crear el mensaje sintetizado, el cual se compone de la definición de la escena, que incluye la descripción del avatar y de la animación para todos los

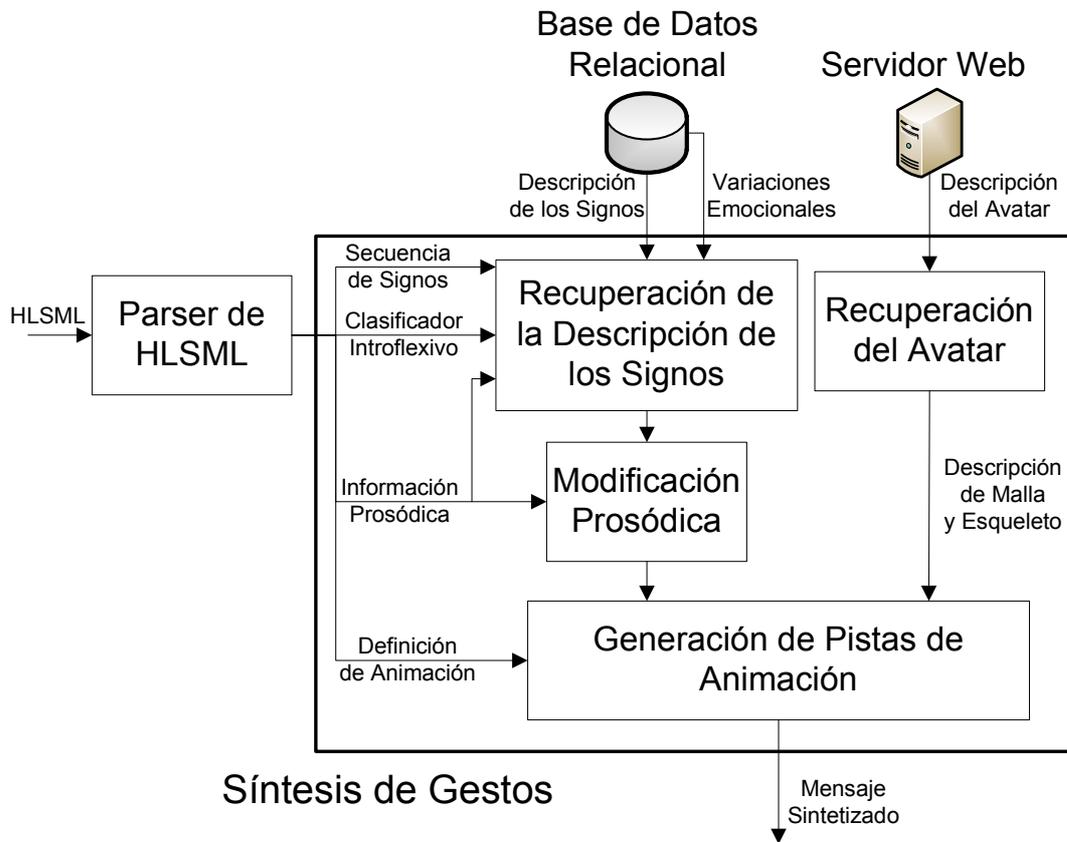


Figura 5.2: Detalle del módulo de síntesis de gestos.

huesos del esqueleto que requieran ser animados. Para la descripción de este módulo seguiremos el diagrama presentado en la Figura 5.2. A continuación se describen brevemente los elementos más importantes de este módulo, que serán tratados con mayor profundidad en el Apartado 9.

Recuperación del avatar

De manera paralela a la recuperación de la información de la base de datos relacional se obtiene la descripción del avatar desde el servidor web (junto con el resto de elementos de la escena). Este proceso sólo es necesario realizarlo al inicio de la ejecución del sintetizador; en posteriores generaciones de mensajes únicamente será preciso eliminar la animación anteriormente definida.

Recuperación de la descripción de los signos

El mensaje HLSML contiene la secuencia de signos que componen el mensaje a sintetizar. El primer paso en la síntesis de gestos es la recuperación de la descripción de estos signos de la base de datos relacional. La generación de las consultas a la base de datos relacional utiliza como clave principal la glosa de los signos especificada en el mensaje. Como indicamos en la Sección 4.2, ciertos clasificadores alteran alguno de los PFQ de la descripción de un signo *establecido*, siendo en este punto de la síntesis donde se aplica esa clasificación. Las consultas a la base de datos se adaptan a la existencia de un *clasificador introflexivo* para recuperar la información correcta. Por último, también se ha mencionado cómo las modificaciones prosódicas alteran la definición de ciertos PFQ, lo que conlleva una nueva modificación en las consultas a la base de datos acorde con estas alteraciones.

Modificación prosódica

Antes de generar la animación, la información obtenida de la base de datos debe alterarse en función de los modificadores prosódicos contenidos en el mensaje de entrada. Estas modificaciones pueden alterar la velocidad de reproducción de un signo o el tiempo de transición entre los mismos. En este punto se establecen otras directivas que van a modificar el proceso que se realiza en el módulo «Render», en particular, definiendo el modo de interpolación.

Generación de pistas de animación

Una pista de animación es una secuencia de claves que definen una característica de un objeto de la escena y llevan asociada una marca de tiempo. Esta característica puede ser alguna de las tres propiedades básicas, como son la posición, la orientación o el escalado, o bien relativa a alguna otra característica del objeto, como por ejemplo, su color. La definición de los signos recuperada de la base de datos relacional junto con la descripción de las animaciones establecidas en el HLSML para los *clasificadores* se procesan para generar las pistas de animación de la orientación de los huesos del esqueleto y la posición de los objetos auxiliares.

La descripción contenida en la base de datos relacional de algunos PFQ consiste en un conjunto de orientaciones para una serie de articulaciones. Esta información puede ser utilizada directamente en la creación de las pistas de animación. Los PFQ que se definen mediante referencias anatómicas o coordenadas espaciales necesitan

ser procesados, junto con la descripción del avatar, para obtener las rotaciones que se aplicarán a las articulaciones implicadas. La descripción del avatar es relevante para este proceso y será estudiada con mayor detalle en el Apartado 6.

5.1.5. Render

El proceso de «Render» (o *rendering*) consiste en la generación de una imagen 2D a partir de la definición de una escena 3D, creada combinando una serie de elementos extraídos del «Servidor Web» y las pistas de animación generadas en el módulo de «Síntesis de Gestos». Para obtener cada imagen, hay que definir la orientación de todos y cada uno de los huesos del esqueleto establecida a partir de los valores de las pistas de animación. En estas pistas no está definida la orientación de los huesos para todos los fotogramas, sino de manera discreta. Para obtener la rotación de cada hueso en el instante de la generación de la imagen se utilizan técnicas de interpolación. La estructura del esqueleto propuesto implica ciertas acciones en el proceso de «Render», que serán descritas en el Apartado 6. Una vez establecidas las orientaciones de todos los huesos, se debe calcular la deformación de la malla correspondiente a la postura adoptada por el esqueleto. A continuación se realizan los cálculos de iluminación para la geometría y se crea la imagen final. Cada una de estas imágenes 2D (fotogramas de un vídeo) se genera en intervalos de tiempo. La duración de estos intervalos depende de la aproximación que se esté utilizando:

- La primera aproximación es el render en tiempo real. Cuando se utiliza este método, una imagen se visualiza en cuanto se termina de procesar por el «Render». Por lo tanto el intervalo entre imágenes consecutivas se establece por el tiempo requerido para procesar cada una de ellas. Esto puede dar lugar a intervalos irregulares entre imágenes o excesivamente largos, dando lugar a una falta de fluidez en la animación. Esta aproximación se utiliza en los juegos tridimensionales, por lo que, un dispositivo con suficientes recursos gráficos y de procesamiento podría utilizarla.
- La segunda aproximación es el render en diferido. En este caso, la generación de la animación en su totalidad se realiza previamente a su visualización. El intervalo entre imágenes puede establecerse a voluntad para ajustarse al estándar de visualización requerido. Además, el tiempo necesario para generar cada imagen no está restringido, con lo que puede optarse por utilizar técnicas con mayor coste computacional, pero que generan imágenes de una mayor calidad y realismo. El resultado se almacena en un fichero de vídeo que será visualizado con posterioridad. Esta aproximación requiere menor capacidad de proceso del

sistema ya que, aunque el tiempo necesario para generar la animación será superior, si el dispositivo utilizado permite visualizar vídeo, el resultado obtenido será aceptable.

5.1.6. Visualización

La visualización de la animación resultante es el último paso del proceso. Como hemos visto, el proceso de render presenta dos alternativas distintas que implican dos métodos de visualización distintos. El primero es una visualización ligada al proceso de «Render», en el que las imágenes son generadas en tiempo real y el segundo es la visualización de una secuencia de vídeo. El resultado observado por el usuario será el mismo en ambos casos, salvo incidencias en la reproducción de vídeo o en el módulo de «Render».

5.2. Síntesis distribuida

Debido a la diversidad de dispositivos existentes en el mercado y a sus distintas capacidades gráficas y computacionales, se plantea una estrategia en la arquitectura que facilite su adaptación a la mayoría de estos sistemas. Así mismo, esta estrategia también permitirá adaptar el sistema en función del tipo de conexión del que disponga el cliente en el momento de solicitar el servicio de síntesis. La Figura 5.3 muestra un esquema de las posibles alternativas que se pueden adoptar.

A continuación se describen dichas alternativas:

- La primera opción (escenario α) libera completamente al servidor, ya que el mensaje HLSML se envía al cliente y es éste el que lo descodifica, sintetiza y renderiza en tiempo real. Esta solución implica gran carga computacional para el cliente y una carga alta para la red. Como se ha explicado, la síntesis del mensaje en lengua de signos requiere una conexión a la base de datos relacional en la que se almacena la descripción de los signos, debido a que el proceso de recuperación de la información realiza una serie de consultas a la misma. Además del envío de esta información, el fichero con el modelo del avatar debe descargarse una vez al comienzo de cada ejecución de la aplicación. Esta solución es adecuada para un ordenador personal de gama media-alta con una conexión de banda ancha y tarifa plana, como sería el caso de una conexión ADSL.

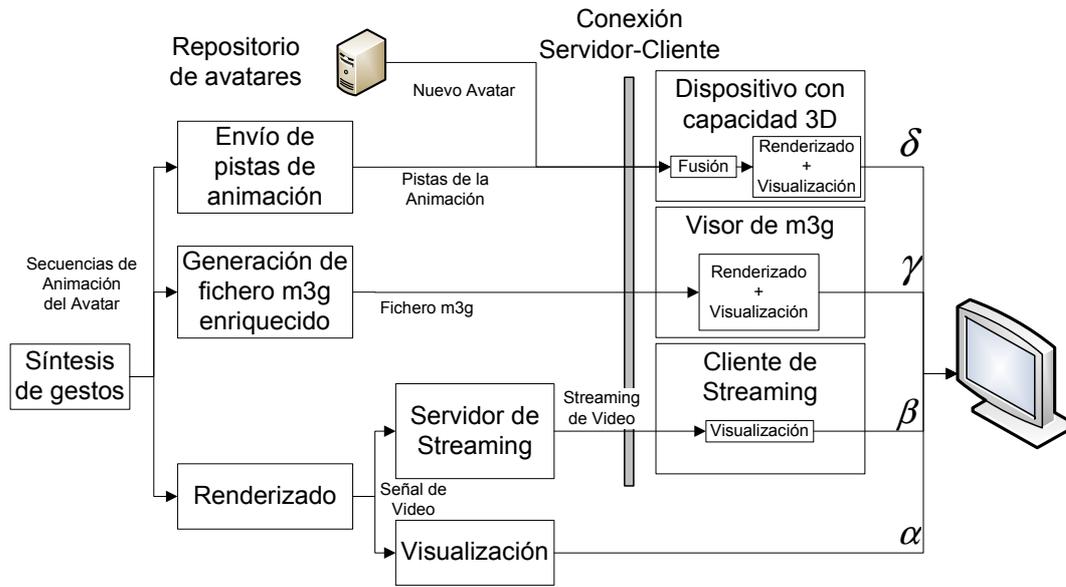


Figura 5.3: Reparto de los módulos de «Síntesis de Gestos», «Render» y «Visualización» entre cliente y servidor.

- La segunda alternativa planteada (escenario β) contempla que el trabajo de síntesis y renderizado del mensaje se realice en su totalidad en el servidor. El resultado se envía hasta el cliente en forma de secuencia de vídeo, envío que puede realizarse mediante dos técnicas distintas. La primera define un servicio de *streaming* de vídeo en el que la señal se envía según se va generando. En la segunda, el fichero de vídeo se enviará una vez se haya generado en su totalidad. Lógicamente en esta alternativa todo el peso de la síntesis y renderizado recae en el servidor. Además se genera una carga alta para la red ya que el envío de vídeo implica la transmisión de una gran cantidad de información. Por el contrario, la ventaja de esta aproximación reside en que el dispositivo cliente sólo necesita software de reproducción de vídeo. La interfaz del cliente puede basarse en un navegador web o en otra aplicación sencilla. Así pues, el cliente no requiere capacidad 3D, tener una librería 3D o una aplicación específica.
- La tercera de las opciones propuestas (escenario γ) reparte la carga de trabajo entre el servidor y el cliente. La labor de síntesis recae por completo en el servidor, mientras que el cliente debe llevar a cabo la tarea de renderizado de la escena tridimensional. El envío desde el servidor al cliente consiste un fichero «m3g» enriquecido con respecto al que se obtiene del «Servidor Web». En este fichero se combina la descripción de la escena con el avatar y las pistas de animación para cada uno de los huesos. Por lo tanto, una vez que el servidor ha generado las pistas de animación para cada uno de los huesos involucrados, se

genera el fichero «m3g» enriquecido. Este fichero es reproducido por el cliente, dado que contiene la animación 3D que debe ser renderizada.

Con esta estrategia no se requiere el uso de un dispositivo de altas prestaciones para realizar todo el proceso de «Síntesis de Gestos», siendo suficiente una aplicación de visualizado de ficheros «m3g». Dado que la comunicación entre el servidor y el cliente se limita al envío de un único fichero, la carga de la red se ve reducida en gran medida.

- La cuarta y última opción propuesta (escenario δ) es una variante del escenario γ . Al igual que en dicho escenario, el servidor realiza el proceso de síntesis y el cliente el de renderizado. La diferencia estriba en la información que se envía al cliente y en el procesado que éste debe hacer previamente al renderizado.

El servidor únicamente envía al cliente el vector de pistas de animación para cada uno de los huesos. Estas pistas de animación indican la orientación que debe establecerse para cada hueso en instantes de tiempo determinados. El cliente debe integrar esta información con la descripción de la escena y del avatar obtenidos de manera independiente. Esto permite que la librería 3D que utiliza el cliente no tenga que ser necesariamente la JSR-184 [59], lo que define este escenario como la solución ideal para plantear la interacción con clientes en VRML [56, 57] o XNA [78]. Las pistas de animación para el avatar, se definirán en función de la librería 3D utilizada por el cliente, que deberá obtener la descripción de la escena del «Servidor web».

El estudio de las cargas de proceso para el servidor, la red y el cliente en las distintas alternativas planteadas se muestra resumido en la Tabla 5.1 y de manera gráfica en la Figura 5.4.

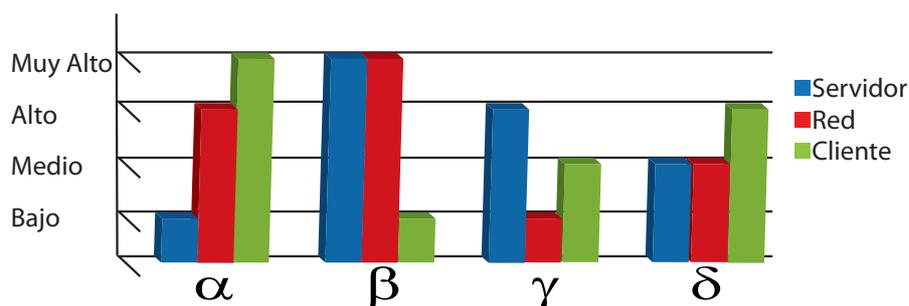


Figura 5.4: Distribución de cargas de proceso en las cuatro alternativas.

	α	β	γ	δ
Servidor	BAJA El mensaje en HLSML se envía al cliente.	MUY ALTA Síntesis, Renderizado y envío del vídeo.	ALTA Síntesis y generación del m3g enriquecido.	MEDIA Síntesis y envío de información de animación.
Red	ALTA Descarga del avatar y consultas a BD.	MUY ALTA Transmisión de vídeo.	BAJA Envío de fichero enriquecido.	MEDIA Pistas de animación y descarga del avatar.
Cliente	MUY ALTA Síntesis, Renderizado y Visualización.	BAJA Visualización.	MEDIA Renderizado y Visualización.	ALTA Procesado, Renderizado y Visualización.

Tabla 5.1: Esta tabla muestra la distribución de carga de proceso en cada una de las alternativa definidas para la distribución del proceso de síntesis.

5.3. Resumen

En este apartado hemos presentado los distintos elementos que componen la arquitectura modular del sintetizador, presentando la relación existente entre ellos: «Parser de HLSML», «Base de Datos Relacional», «Servidor Web», «Síntesis de Gestos», «Render» y «Visualización». Así mismo, hemos definido cómo distribuyendo la ejecución de estos módulos entre un servidor y el dispositivo cliente, podemos adaptar el proceso de síntesis a los recursos de dicho dispositivo.

Apartado 6

Avatar

Este apartado se centra en las características del avatar desarrollado en esta tesis. Como puede verse en el Apartado 4, la formalización realizada implica la modificación de la estructura esquelética estándar utilizada para la animación de un humanoide. En este apartado presentamos el modo en que se ha plasmado esta formalización en el diseño del esqueleto. A continuación se presenta la malla diseñada para el avatar.

Las dos últimas secciones de este apartado, repasarán las partes más importantes del avatar. En este repaso se describirá la estructura de dichas partes, presentando cómo los distintos elementos del esqueleto y de la malla, descritos en las secciones iniciales de este apartado, se combinan. En la última sección de este apartado, se presentarán brevemente los dos avatares que se han desarrollado en el periodo de la tesis.

6.1. Esqueleto

Hemos definido el esqueleto de un avatar como una estructura jerárquica de control que permite modificar la forma de la malla y adecuarla a las distintas posturas requeridas. Esto implica que la función de un hueso en el esqueleto es principalmente la de establecer deformaciones en la malla. Veremos cómo la definición de esta estructura jerárquica permite realizar más funciones que la antes mencionada.

El estudio anatómico realizado por Levangie y Norkin [67] ha sido la base para la definición del esqueleto de estos avatares.

6.1.1. Descripción

El esqueleto que se ha desarrollado tiene como función principal permitir la animación del avatar. La estructura que se ha seguido al definir la componente de animación del esqueleto es comparable a la definición H-ANIM, la cual puede dividirse en cuatro secciones: manos, brazos, cabeza y tórax. Esta división se realiza en función de la relevancia que tienen para la síntesis de lengua de signos e influye directamente en la complejidad del diseño. La Tabla 6.1 presenta la distribución de los huesos del esqueleto desarrollado en este trabajo en las cuatro secciones antes mencionadas.

	Cabeza	Manos	Brazos	Cuerpo	Total
Número de Huesos	79	64	28	39	210
Porcentaje	37,62 %	30,48 %	13,33 %	18,57 %	100 %

Tabla 6.1: Desglose de los huesos utilizados en el avatar según su posición. Los huesos de la cabeza son todos los huesos que provienen del hueso "NECK", incluido, los huesos de cada mano heredan del hueso "X_P_WRIST", éste incluido y los huesos de cada brazo desde el hueso del húmero, incluido, hasta el hueso de posición de la muñeca, exclusive. Los restantes huesos han sido agrupados en el cómputo del cuerpo.

Se ha optado por simplificar la definición de la columna vertebral y utilizar sólo dos articulaciones, una a la altura de la cadera, y otra en el comienzo de las costillas. Además, no se ha considerado la definición del esqueleto del avatar por debajo de las rodillas dado que la animación de las piernas no es relevante para la lengua de signos. La variación en la orientación de todos estos «huesos de animación» tiene como consecuencia un conjunto de deformaciones en la malla del avatar.

En adición a los «huesos de animación» se han definido otra serie de «huesos estructurales», que no existiendo en la estructura esquelética humana, o no siendo relevantes para la animación, facilitan en gran medida la definición de la estructura del esqueleto. Este es el caso de los huesos que representan los metacarpos de los dedos índice, medio, anular y meñique, sobre los no se define animación, pero su inclusión facilita la definición de la estructura esquelética de la mano del avatar. Por otro lado, estos huesos facilitan la definición de las deformaciones de la malla en la sección de la mano. También se han definido estos «huesos estructurales» en el cuerpo del avatar, para la definición de los fémures del avatar.

Una diferencia sustancial con otros avatares se encuentra en el modo en que se generan las expresiones faciales. Como hemos visto, la técnica más extendida es la de *morphing de malla* o deformación de la malla para adaptarse a determinadas expresiones objetivo. Cada uno de estos objetivos puede definir una expresión completa de la cara o simplemente la modificación de una parte. Cada uno de ellos es una copia de la sección de la malla que describe. Por lo tanto, el tamaño del fichero que almacena

el avatar, aumenta en función del número de objetivos incluidos. De igual modo, la inclusión de una nueva expresión requiere que, tras su modelado, se incluya en el fichero de descripción del avatar y posteriormente la nueva versión deberá ser distribuida para que esta nueva expresión esté disponible. Además, la librería de animación debe gestionar dos sistemas de animación distintos: la animación del *morphing* y la animación esquelética.

Con el objeto de evitar el incremento del tamaño del fichero que describe el avatar, así como las complicaciones relativas a una actualización y lograr una simplificación en el manejo de la librería de animación, se ha optado por desarrollar otra aproximación para la animación facial. El método de animación facial propuesto se basa en la animación esquelética. De esta manera, una expresión queda definida por un conjunto de orientaciones de huesos, lo que reduce el tamaño del fichero de descripción del avatar en comparación con el método de *morphing*. De igual modo, cada nueva expresión requiere una combinación diferente de valores para las orientaciones de los huesos implicados. Teniendo en cuenta la arquitectura del sistema, estas nuevas orientaciones quedan almacenadas en la base de datos relacional y están disponibles instantáneamente. Por último, al utilizar la animación esquelética se unifica con la animación utilizada para el resto del avatar. Como consecuencia de esta aproximación, aumenta el número de huesos que se utilizan en la cabeza (Tabla 6.2).

6.1.2. Nuevas clases de huesos

En el punto anterior se ha tratado la estructura del esqueleto desde el punto de vista de la animación. A continuación se describen las nuevas clases de huesos. La función de estos huesos no es generar la animación del avatar, sino que se centra en cubrir una serie de requisitos en la síntesis de signos. Esas funciones son:

1. Eliminar la dependencia existente entre el PFQ *orientación* y los PFQ *plano*, *localización* y *movimiento*.
2. Facilitar la definición de la orientación de los ojos y de la cabeza del avatar.
3. Establecer los puntos anatómicos necesarios para la definición del PFQ *localización*.
4. Crear un conjunto de poliedros que permitan simplificar los cálculos orientados a evitar la colisión de las manos del avatar con el cuerpo y la cabeza.

6.1.2.1. Hueso de la muñeca

En el apartado de formalización, establecimos la necesidad de mantener la independencia de todos los PFQ. En todas las estructuras esqueléticas que se han revisado en el estado del arte se mantiene la herencia de transformaciones en todos los huesos. Esto supone que un cambio en la posición de la mano definido por los PFQ *localización*, *plano* o *movimiento*, implica una alteración en el valor del PFQ *orientación*.

La solución que se propone en este trabajo es introducir un hueso auxiliar en la definición de la muñeca, denominado “X_P_WRIST”. La función que realiza este hueso es romper la herencia de transformaciones de forma parcial, en particular, provoca que la orientación de la mano sea independiente de los valores heredados del antebrazo. Las dos transformaciones que si se heredan son las de escalado y la de traslación, siendo ésta última la más importante.

La acción de este hueso queda patente cada vez que se actualiza el conjunto de transformaciones que heredan cada uno de los huesos del esqueleto. En este momento, cada uno de los huesos “X_P_WRIST” desecha la orientación heredada del antebrazo y la sustituye por un valor constante. Esto implica que el valor de orientación que hereda el hueso siguiente en la jerarquía y responsable de la *orientación*, “X_R_WRIST”, será siempre el mismo, con independencia de la posición que haya tomado el brazo. Con esto se logra que la orientación de la mano sea la que se ha definido en el PFQ *orientación* y la independencia de este PFQ con los mencionados antes. La posición que ocupa este nuevo hueso dentro de la estructura de la muñeca del avatar, puede verse en la comparación de la Figura 6.1.

6.1.2.2. Huesos direccionales

Cuando se han tratado los *clasificadores* en la Sección *Clasificación de las unidades de las LS* se ha expuesto la necesidad de que el avatar pueda dirigir la cara o la vista hacia un punto en particular. Este hecho está directamente relacionado con el punto anterior. La herencia de orientaciones a lo largo de los huesos del esqueleto complica el establecimiento de una orientación en particular. Además, los elementos que deben ser orientados hacia un punto en concreto también se incluyen dentro del PFQ *no manual*, lo que implica que también deben poder establecer una orientación específica si fuera necesario.

El método empleado para gestionar este problema tiene la misma base que la solución tomada para independizar el PFQ *orientación*: eliminar la herencia de orientaciones. La diferencia respecto al punto anterior consiste en que el hueso que rompe esa

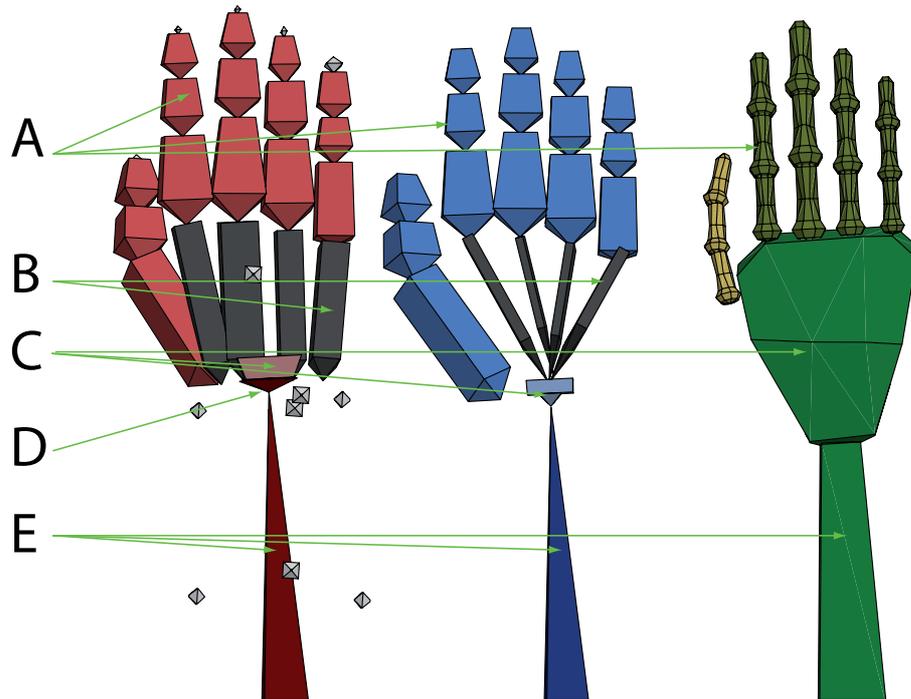


Figura 6.1: Comparación del diseño esquelético aplicado a las manos del avatar. A la izquierda el modelo propuesto en este trabajo, en el centro una estructura compatible con H-ANIM y a la derecha la estructura utilizada en eSIGN.

herencia no define una orientación constante, sino que se aprovecha dicha orientación para mover los ojos y la cabeza haciendo que miren a un punto concreto del espacio. El punto al que se orientan estos huesos está definido por dos objetos no visibles en la escena. Tanto en el caso de la cabeza como en el de los ojos, el siguiente hueso es el encargado de establecer la orientación necesaria para la generación de la ‘expresión simple’ a la que pertenecen.

6.1.2.3. Huesos para el PFQ *localización*

En la Sección 2.1 se establece que el PFQ *localización* se define mediante el uso de referencias anatómicas. Este método de descripción hace que sea necesaria la obtención de la posición de los puntos relevantes de la anatomía del avatar. El estándar H-ANIM permite obtener la posición de determinados puntos anatómicos, ya que define estas posiciones relacionándolas con el vértice de la malla más cercano. Como se mencionó anteriormente, el número de estos puntos no cubre todas las posibles ubicaciones necesarias en la descripción de la lengua de signos. El avatar utilizado en

el proyecto eSIGN incrementa el número de estas posiciones cubriendo las descripciones de HamNoSys; sin embargo continúa utilizando los vértices de la malla para esta tarea.

El uso de los vértices para la definición de las ubicaciones anatómicas presenta una serie de inconvenientes. El proceso de modelado de la malla queda condicionado debido a que es necesario establecer una serie de vértices en posiciones definidas. Esto dificulta el uso de avatares que no hayan sido definidos específicamente. Además dicha condición impide ciertas simplificaciones que puedan ser útiles en el desarrollo de un avatar específico para un dispositivo con recursos limitados. Todo ello implica que durante la síntesis de signos se debe gestionar la deformación de la malla para obtener las *localizaciones*, lo que implica un porcentaje significativo del tiempo de proceso.

La aproximación que se sigue en esta tesis tiene por objeto evitar los inconvenientes antes citados y eliminar la necesidad de utilizar la malla durante la síntesis de signos. Al omitir toda referencia a la malla, únicamente podemos basarnos en la definición del esqueleto para obtener la posición de los puntos anatómicos definidos en la LS. Por lo tanto, se ha optado por introducir una serie de huesos en la definición del esqueleto a partir de los cuales se obtiene la posición de dichos puntos anatómicos. Estos huesos reciben el nombre de «huesos de localización».

Los «huesos de localización» se definen dentro de la jerarquía del esqueleto. Ninguno de ellos tiene huesos herederos y no están asociados a ningún vértice de la malla para el proceso de deformación. Para definir el hueso antecesor se selecciona el hueso más relevante para la deformación de los vértices colindantes; de esta manera se logra que la posición del hueso sea lo más próxima a la superficie de la malla. Durante el proceso de síntesis, cuando se desea obtener las coordenadas de un punto en la anatomía del avatar, se calcula la posición del hueso correspondiente.

Por último, cabe señalar que, cuando existe un hueso de animación que coincida con una de estas ubicaciones anatómicas, también puede utilizarse para la definición del PFQ *localización*, con el consecuente ahorro de huesos en el esqueleto.

6.1.2.4. Huesos para la detección de colisiones

La última de las clases especiales de huesos definidos en el esqueleto es la que se utiliza para los cálculos de detección de colisiones. La primera aproximación para realizar los cálculos necesarios para la detección y solución de colisiones sería utilizar la propia malla del avatar. Debido al alto número de polígonos que componen la malla, la duración del proceso de detección de colisiones aumentaría considerablemente.

Además, como ya se ha manifestado, se desea que el proceso de síntesis no requiera de la definición de la malla. Por último, los cálculos para la detección de colisiones se simplifican cuando se utilizan volúmenes convexos, algo que no se puede asegurar para todos los diseños de la malla.

La aproximación utilizada ha consistido en definir una serie de poliedros convexos que cubran las principales partes del avatar. Estos poliedros se definen utilizando una fracción del número de caras con las que se define la malla; los vértices de dichos poliedros se definen utilizando huesos del esqueleto del avatar. Estos huesos reciben el nombre de «huesos de colisión». Al igual que los «huesos de localización», no tienen huesos herederos y no alteran en ningún modo la malla. La Figura 9.4 muestra los poliedros definidos gracias a estos huesos.

En ciertos casos determinados huesos pueden utilizarse bien como «huesos de localización» o como «huesos de colisión». Tal es el caso del hueso que define el punto anatómico en la parte superior de la cabeza que, además, es uno de los vértices del poliedro que cubre la misma.

	Cabeza	Manos	Brazos	Cuerpo	Total
Animación	24	32	4	4	65
Estructura	3	8	0	5	16
Muñeca	0	2	0	0	2
Direccional	3	0	0	0	3
Localización	41	52	24	17	134
Colisión	14	0	0	18	32

Tabla 6.2: Los huesos han sido clasificados por su posición y su función dentro del esqueleto. La agrupación en distintas zonas sigue el mismo criterio que en la Tabla 6.1. Algunos huesos cumplen varias funciones y por lo tanto han sido incluidos en todas las secciones relevantes.

6.2. Malla

La malla es el componente del avatar que define su apariencia. La impresión que obtenga el usuario que utilice el sintetizador dependerá en gran medida de dicha apariencia. Es importante conseguir la generación de una malla tan detallada y realista como sea posible.

El aumento en el detalle de la malla para mejorar la apariencia del avatar se compensa con las restricciones técnicas que define el motor gráfico. Todos los polígonos que contiene una escena tridimensional deben procesarse cada vez que se genera una imagen. Si la tasa de imágenes por segundo que se alcanza no llega a ciertos valores mínimos no se observará una animación fluida. Esta tasa de refresco mínima está

entre quince y veinte imágenes por segundo, dependiendo de las personas. Si una animación no alcanza estos valores se ve penalizada en su calidad y aceptación por el usuario.

Por tanto, se debe establecer un número máximo de polígonos para que la definición del avatar consiga una animación fluida. Para ello se planteó el análisis de la frecuencia de refresco de diversas versiones del avatar *Yuli*¹ con distinto número de polígonos. Los resultados de este experimento se pueden ver en la Sección 10.1.

Las partes del cuerpo que mayor grado de detalle requieren son la cara y las manos, como quedó confirmado con los estudios de Lu *et al.* [73] y de Cherniavsky *et al.* [17], ya que sobre estas secciones recae la mayor atención. Como puede observarse en la Tabla 6.3, es a estas partes del avatar a las que se ha dedicado mayor número de vértices y polígonos.

	Cabeza	Manos	Brazos	Cuerpo	Total
Polígonos	3816	2920	740	291	7767
Porcentaje	49,13 %	37,59 %	9,53 %	3,75 %	100 %
Vértices	1994	1466	382	137	3979
Porcentaje	50,11 %	36,84 %	9,60 %	3,44 %	100 %

Tabla 6.3: Descripción de la malla del avatar *Daniel*².

La Figura 6.2 muestra la estructura esquelética del avatar junto con la malla de uno de los avatares definidos. Mientras que los huesos relacionados con la deformación de la malla se sitúan en el interior de la misma, la mayoría de los nuevos huesos definidos («huesos de localización» y «huesos de colisión») se encuentran en la superficie.

¹ *Yuli* es uno de los avatares creados en este trabajo, ver la Sección 6.4

² *Daniel* es uno de los avatares creados en este trabajo, ver la Sección 6.4.

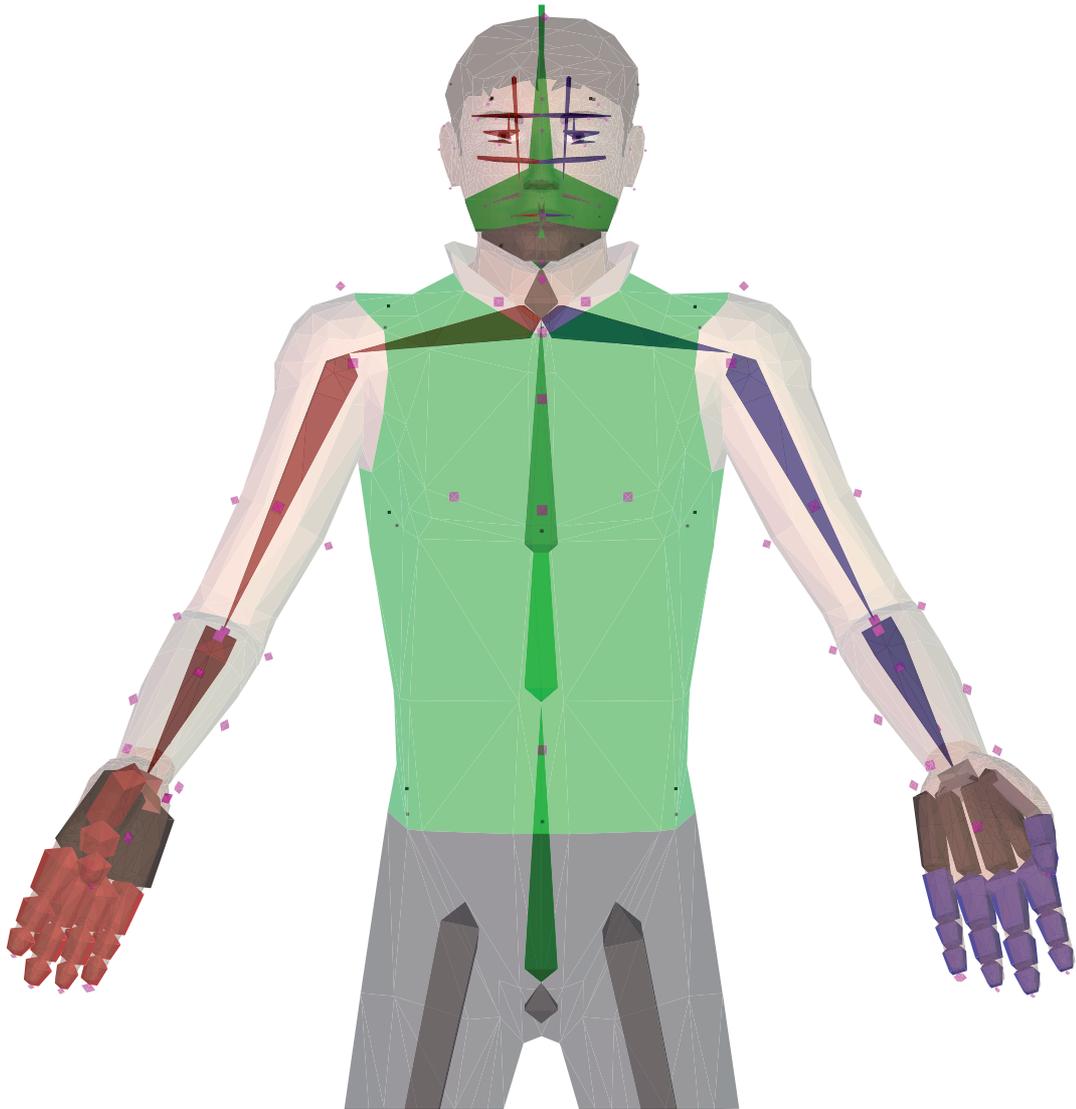


Figura 6.2: Representación de la alineación entre los distintos elementos que componen el avatar: los «huesos de localización» y los «huesos de colisión» (capa superior), la malla del avatar (capa intermedia) y el esqueleto definido (capa inferior).

6.3. Principales secciones del avatar

Las dos secciones anteriores han descrito las categorías de huesos que se han definido y la descripción de la malla. Esta sección se centra en el estudio de las distintas partes que componen el avatar, con lo que se verá con mayor claridad la aplicación que han tenido las distintas definiciones planteadas en las secciones anteriores de este apartado.

6.3.1. Mano

Observando con detenimiento el diseño de la estructura esquelética de la mano (Figura 6.3), se distingue el uso de cuatro categorías de huesos: «huesos de animación», «de localización», «estructurales» y el hueso especial de la muñeca. Aunque hemos definido que la mano comienza en el hueso “X_P_WRIST”, por claridad empezaremos el análisis de la estructura de la mano describiendo brevemente el antebrazo del avatar.

- El antebrazo del avatar se modela mediante el hueso “FOREARM”, (A) en la Figura 6.3. Dicho hueso es directamente la representación del Cúbito y del Radio, por lo tanto no aporta ningún elemento novedoso sobre la definición anatómica.
- Los siguientes dos huesos representan la muñeca, (C) y (D) en la Figura 6.3. Como puede verse en la Figura 6.1, normalmente se utiliza un único hueso para emular la rotación de la muñeca pudiendo además, servir de base para los huesos que llevan a cabo la función de las falanges. El hueso directamente dependiente en la jerarquía al antebrazo es el hueso “X_P_WRIST”. Este hueso cumple una doble función: por una parte, es el hueso que se utiliza como extremo en los cálculos de cinemática inversa y, por la otra, es el hueso que se utiliza para eliminar la herencia de las transformaciones de rotación entre el antebrazo y la muñeca. El siguiente elemento en la jerarquía es el hueso “X_R_WRIST”. Las rotaciones que se efectúan sobre este hueso se traducen en las rotaciones que puede efectuar la muñeca. Dado que su hueso anterior “X_P_WRIST” anula la rotación heredada del brazo, se puede asegurar que la mano siempre estará orientada en la misma dirección, con independencia de la orientación del brazo y del antebrazo. Este hueso cumple únicamente funciones de animación, ya que para obtener la posición de la muñeca se utiliza el hueso “X_P_WRIST”.

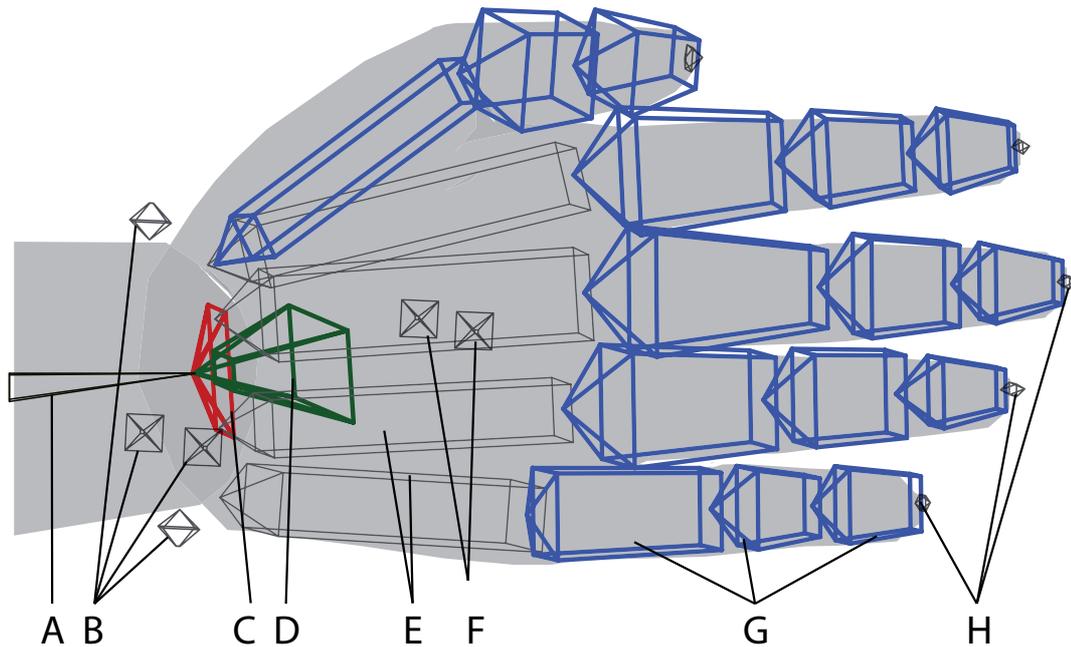


Figura 6.3: Diagrama de los huesos que componen la mano: (A) hueso del antebrazo, (B) huesos de localización de la muñeca, (C) hueso "X_P_WRIST", (D) hueso "X_R_WRIST", (E) huesos estructurales equivalentes a los metacarpos, (F) huesos de localización del dorso y de la palma, (G) huesos de animación de los dedos, (H) huesos de localización en la punta de los dedos.

- A continuación encontramos huesos de localización para obtener distintas posiciones relativas a la muñeca; son los huesos marcados con (B). Similares a estos huesos, son los marcados con (F) que permiten obtener la posición tanto de la palma de la mano como del dorso.

Dependiendo igualmente del hueso "X_R_WRIST", encontramos un conjunto de huesos estructurales, que equivalen a los huesos metacarpianos: huesos (E). En el diseño actual, su función es estructural y no se define animación para los mismos, pero en un futuro desarrollo se puede contemplar la posibilidad de generar animación para estos huesos; en especial para los metacarpos correspondientes a los dedos anular y meñique, con lo que se puede llegar a mejorar la expresividad de la mano.

- La animación de la flexión de los dedos se realiza mediante el hueso metacarpo del pulgar, las dos falanges del mismo dedo y las doce falanges del resto de los dedos. Estos huesos se han dibujado en azul en la Figura 6.3, etiquetados como (G), y cumplen la doble función de animación y localización. Su rotación genera la flexión de los dedos de la correspondiente configuración de la mano y su posición nos permite obtener la localización de los puntos anatómicos definidos

dentro de la misma. Así por ejemplo, la posición de los huesos correspondientes a las falanges proximales nos dan el valor de la posición de los nudillos.

- Por último, se observa en los extremos de los cinco dedos un hueso extra, etiquetado como (H). Este hueso, ajeno a la anatomía humana, nos permite obtener fácilmente las coordenadas del extremo de los dedos.

Para la definición de la malla correspondiente a una mano se han usado un total de 1460 polígonos (Tabla 6.3). A esta sección de la malla se le han aplicado tres materiales distintos: uno para el dorso de la mano, otro para la palma y el último para las uñas. El hecho de definir dos materiales distintos para la piel de la mano tiene ventajas inmediatas: por una parte mejora la percepción de la mano y aumenta la inteligibilidad del modelo. Por otra, sin aplicación en la presente tesis y trabajo futuro, permite variar el tono de la piel para mejorar la percepción de la tensión de la mano, esto simularía la constricción de los vasos sanguíneos que, al disminuir el riego, hace palidecer la piel de la mano.

6.3.2. Cabeza y cara

La estructura esquelética de la cabeza incluye todos los tipos de huesos definidos en la Sección 6.1, con excepción del definido para la muñeca.

Debido a la aproximación utilizada para definir la animación facial, la estructura esquelética de la cabeza requiere un gran número de huesos de animación. Dentro de la anatomía de la cabeza las únicas articulaciones móviles entre dos huesos son las del cuello y la mandíbula; la alteración en la forma del resto de elementos de la cara responde a contracciones musculares que arrastran secciones de piel. La animación esquelética de estos movimientos no articulares se define de tal manera que el movimiento del hueso recorra el mismo arco que recorre la piel. La Figura 6.4 presenta los huesos que definen la animación facial y las secciones de la malla a las que afectan.

Los «huesos direccionales» se concentran todos en la cabeza del avatar. El primero de ellos es el hueso «NECK», antecesor del hueso «HEAD». Este hueso permite definir la dirección hacia la que se orienta la cara del avatar. Los otros dos «huesos direccionales» forman parte de la definición de los ojos y permiten definir la dirección hacia la que mira el avatar. Tanto en el caso de la cabeza como en el de los ojos, el hueso siguiente al «direccional» es un «hueso de animación». Este hueso animable recibe su orientación de la información definida en la base de datos relacional para la definición de la ‘expresión simple’ correspondiente.

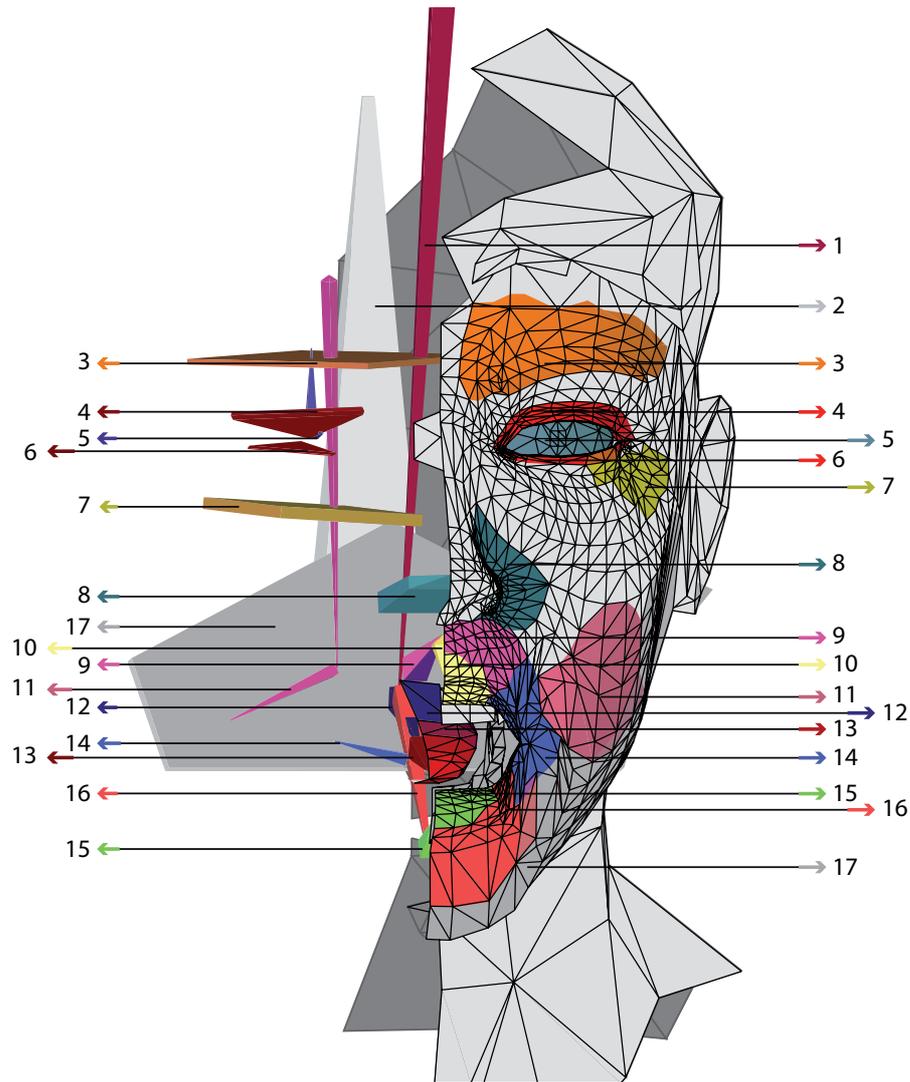


Figura 6.4: Diagrama de las partes de la malla que son deformadas por la rotación de los huesos de la cara. En este diagrama, las secciones de la cara y el hueso responsable de la deformación han sido coloreados del mismo color: (1) hueso "TONGUE_INI_ANI", (2) hueso "HEAD", (3) hueso "X_EYEBROW_ANI", (4) hueso "X_EYELID_UP_ANI", (5) hueso "X_EYE_ANI", (6) hueso "_EYELID_DOWN_ANI", (7) hueso "X_CHEEK_UP_ANI", (8) hueso "DOWN_NOSE_ANI", (9) hueso "NOSE_LIP_ANI", (10) hueso "UP_LIP_ANI", (11) hueso "X_CHEEK_ANI", (12) hueso "TONGUE_MID_ANI", (13) hueso "TONGUE_END_ANI", (14) hueso "X_LIP_ANI", (15) hueso "DOWN_LIP_ANI", (16) hueso "CHIN_LIP_ANI", (17) hueso "CHIN_ANI".

El elemento de mayor complejidad dentro de la cara es la boca. Para definir la animación de los labios y lengua se han requerido un total de diez huesos de animación. Como puede observarse en la Figura 6.5, el movimiento de los labios necesita de un total de seis huesos de animación, al que hay que añadir el hueso que define la articulación de la mandíbula. Los tres huesos restantes se utilizan para la animación de la lengua: el primero de ellos controla el porcentaje de lengua que sale de la boca

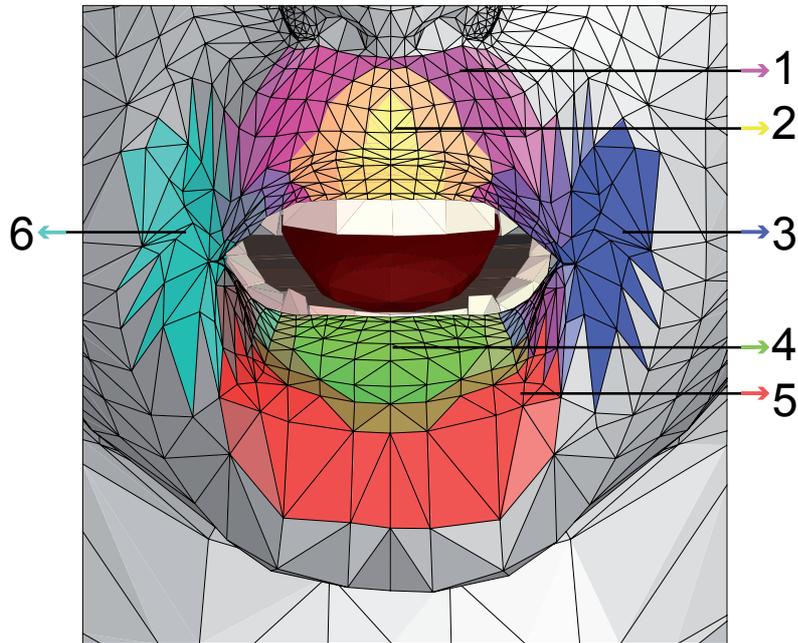


Figura 6.5: Secciones dependientes de cada hueso del contorno de la boca: (1) hueso “NOSE_LIP_ANI”, (2) hueso “UP_LIP_ANI”, (3) hueso “L_LIP_ANI”, (4) hueso “LIP_DOWN_ANI”, (5) hueso “CHIN_LIP_ANI”, (6) hueso “R_LIP_ANI”.

y los otros dos definen la forma de la misma. Esta compleja estructura permite la definición de todas las formas de la boca necesarias para la lengua de signos, entre las que incluimos los visemas³.

Al igual que en las manos, el número de polígonos utilizados para la definición de la cabeza del avatar es elevado (Tabla 6.3). Así mismo, el número de materiales que se han utilizado en esta parte del avatar también es grande en comparación con otras partes del mismo. Como es el caso de las manos, se ha optado por utilizar un material diferente para la cara del avatar; de esta manera se podrá modificar el color de la misma para reflejar distintos estados emocionales.

6.3.3. Brazos y cuerpo

Después de la revisión de la definición de la cabeza y las manos, las dos últimas secciones del avatar que restan por definir son los brazos y el tronco. El esqueleto en estas secciones no aporta novedades desde el punto de vista de la animación. Las diferencias entre el diseño aquí planteado y el resto de aproximaciones son relativas a la inclusión de huesos para la definición del PFQ *localización* y a la definición

³La palabra 'visema' se deriva de las palabras 'visual' y 'fonema': un visema es la referencia visual de un fonema.

del poliedro requerido para evitar las colisiones. Dicho poliedro puede verse en la Figura 9.4. En lo referente a la malla del avatar y dado que son las partes del avatar menos relevantes, son las secciones que han requerido menor porcentaje de polígonos. Por esta misma razón, son las secciones a las que se asignó una prioridad más baja en el momento de reducir el número de polígonos para optimizar la animación.

6.4. Avatares definidos

En esta sección se presentan los dos avatares que se han desarrollado en el transcurso del trabajo de la tesis. El primer avatar que se desarrolló fue *Yuli* y el segundo *Daniel*. Estos dos avatares comparten la mayoría de las propiedades básicas descritas a lo largo de este apartado en cuanto a la definición del esqueleto y estructura de la malla.

	<i>Yuli</i>	<i>Daniel</i>
Animación	63	65
Estructura	17	16
Muñeca	2	2
Direccional	0	3
Localización	50	134
Colisión	0	32
Total	123	210

Tabla 6.4: Comparación entre los esqueletos de los avatares *Yuli* y *Daniel*. Al igual que en la Tabla 6.2, los huesos se incluyen en todas las categorías a las que pertenecen.

	<i>Yuli</i>	<i>Daniel</i>
Vértices	4180	3979
Polígonos	7866	7767
Materiales	10	16

Tabla 6.5: Comparación entre las mallas de los avatares *Yuli* y *Daniel*.

La Tabla 6.4 muestra que el número de huesos con función de animación no ha variado en gran medida a lo largo del trabajo realizado. La única diferencia entre los avatares son los huesos que controlan la parte superior de la mejilla. Los «huesos estructurales» se han mantenido casi constantes entre ambas versiones. En el momento de crear al avatar *Daniel*, se incrementó el número de «huesos de localización», no sólo añadiendo más huesos al esqueleto, sino utilizando para esta función huesos originalmente incorporados con otro objetivo. La gran diferencia entre ambos esqueletos es el uso de «huesos direccionales» y «huesos de colisión», ya que su definición es posterior a la creación de *Yuli*. Las figuras 6.6(a) y 6.6(b) permiten ver la evolución del diseño del avatar.

Como puede verse en la Tabla 6.5, el número de vértices y polígonos no difiere mucho entre los dos avatares. El avatar *Yuli* se utilizó en el experimento de la Sección 10.1 para obtener el número óptimo de polígonos para una visualización fluida.

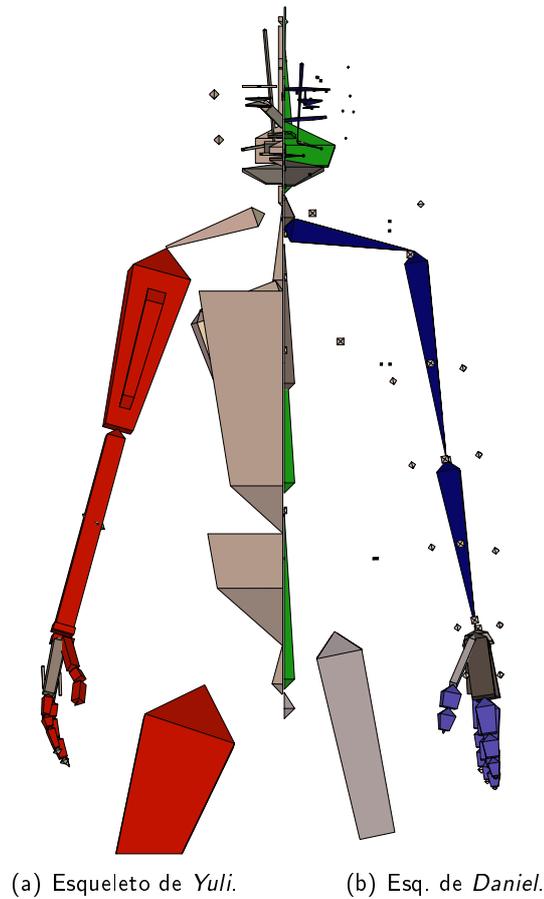


Figura 6.6: Comparación de los esqueletos utilizados para los avatares *Yuli* y *Daniel*.

Cuando se diseñó el avatar *Daniel* se utilizó el número de polígonos de *Yuli* como cota superior del número de vértices. La mayor diferencia entre ambos avatares es su apariencia (*Yuli* presenta a una mujer y *Daniel* a un hombre) y la percepción de los usuarios. La apariencia de *Daniel* fue mejorada en gran medida con respecto a la de *Yuli*, esto hizo que la percepción de las manos y de las expresiones mejorara. Por último, se ha utilizado un mayor número de materiales en el avatar *Daniel*. La mayor diferencia entre ambos avatares en este aspecto fue el uso de distintos materiales para la piel de la cara y las dos manos, diferenciando el material usado en las palmas de las manos, el avatar *Yuli* utiliza el mismo material para estos tres elementos. Las figuras 6.7(a) y 6.7(b) permiten ver la apariencia de los avatares *Yuli* y *Daniel*.

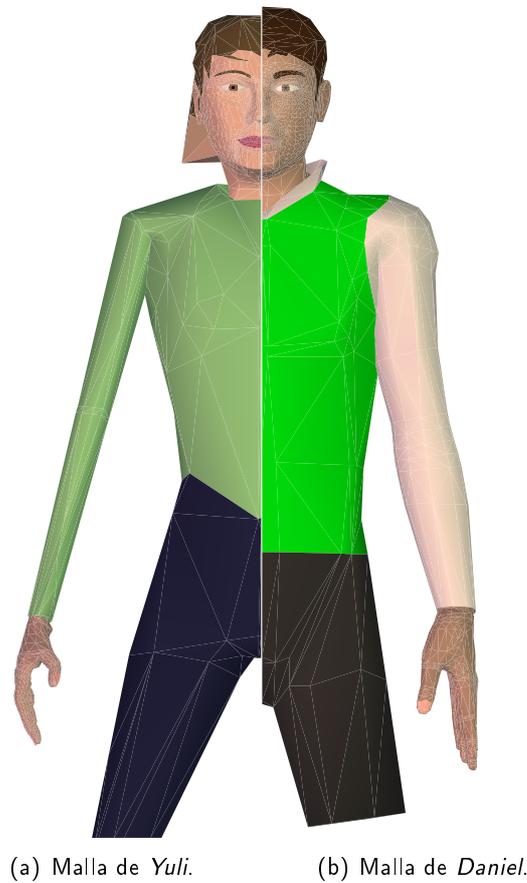


Figura 6.7: Comparación de las mallas utilizados para los avatares *Yuli* y *Daniel*.

6.5. Resumen

Este apartado describe los nuevos tipos huesos que hemos introducido en la definición del esqueleto del avatar para realizar distintas funciones en la síntesis de LSE. Este apartado también recoge una descripción detallada de las distintas secciones que componen el avatar, describiendo tanto la estructura esquelética, como la definición de la malla de las mismas. Por último hemos presentado los dos avatares que se han desarrollado en este trabajo.

Apartado 7

Base de datos relacional

Uno de los criterios seguidos en este trabajo para la síntesis de la LSE ha sido mantener la independencia de los siete PFQ tanto en su definición como en su gestión. De igual modo, se ha seguido el criterio de separar la definición del mensaje a sintetizar de la descripción de los PFQ que componen cada signo del mensaje.

La solución que se ha considerado más adecuada para cumplir con estos criterios ha sido el uso de una base de datos relacional para almacenar la descripción de los signos con lo que se logra que el mensaje a sintetizar no requiera de la definición de los PFQ de cada signo.

Este capítulo se centra en la descripción de la base de datos relacional implementada. La primera sección recoge de manera detallada la estructura de la base de datos; la segunda presenta el modo en que este diseño permite almacenar definiciones alternativas de los signos, ya sea para representar variaciones emocionales o el mismo concepto en distintos idiomas o dialectos.

7.1. Estructura

El diseño simplificado de la base de datos puede observarse en sus líneas generales en la Figura 7.1 y de manera más detallada en la Figura 7.2.

La base de datos relacional está organizada en cuatro niveles conceptuales, como puede observarse en la Figura 7.1, detallándose a continuación:

- El primer nivel cumple la función de una entrada de diccionario, utilizando glosas para definir cada una de éstas. Cada una de las entradas de la base de datos

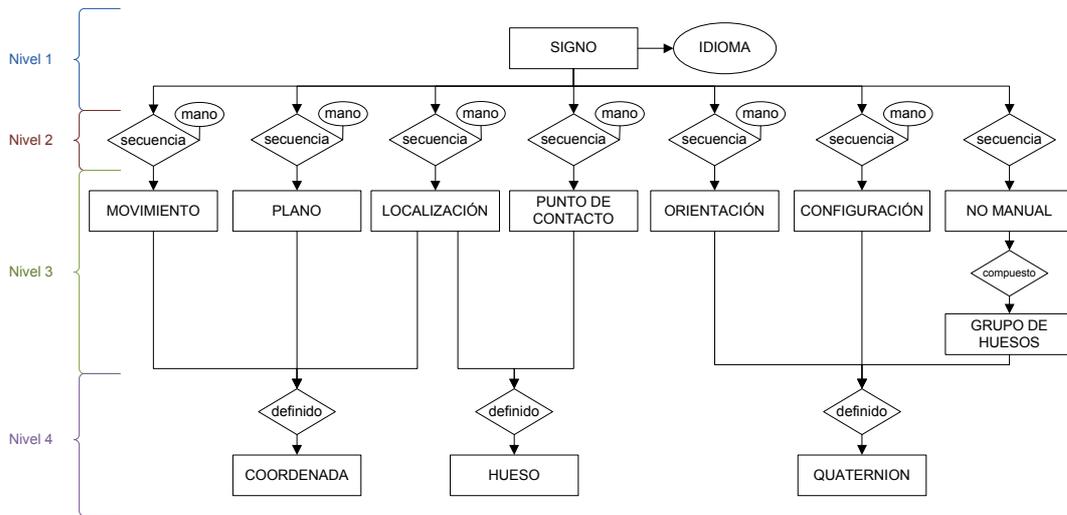


Figura 7.1: Diagrama simplificado de la base de datos.

relacional establece una relación con el idioma en que está definido. Aunque el objetivo de esta tesis es la de la síntesis de la LSE, la variación dialectal entre distintas regiones de España hace necesario contemplar la posibilidad de definir un mismo concepto en las distintas lenguas y dialectos, lo que permitirá en un futuro ampliar el contenido de esta base de datos relacional para su uso con otros idiomas.

Este nivel también almacena la descripción de definiciones alternativas para un mismo signo, cuyo principal objetivo es contemplar distintas ejecuciones del mismo en función de la componente emocional que se desea simular. Dado que el disponer de distintas variaciones de un signo pueden implicar una variación en la duración del signo, este nivel contiene además la definición de la duración total del signo.

- El segundo nivel de la base de datos contiene la descripción de los signos mediante secuencias de cada uno de los siete PFQ que lo componen. La información de estas secuencias se recoge en siete tablas, una para cada PFQ. El instante en que cada PFQ debe ejecutarse dentro de la representación del signo se establece mediante valores porcentuales sobre la duración total del signo representada en el nivel superior.

A diferencia de la secuencia del PFQ *no manual*, para la descripción de las secuencias de los seis PFQ que hacen referencia a las manos es necesario definir cuál de las dos manos se está describiendo. Por último hay que destacar que la descripción de cada entrada de la secuencia del PFQ *movimiento* requiere

que se especifique el instante inicial y final del *movimiento* a ejecutar, definidos ambos mediante el uso de valores porcentuales.

- El tercer nivel de la base de datos contiene las unidades de cada uno de los PFQ. Todos los PFQ manuales se especifican en la base de datos mediante el uso de una única tabla para cada uno. Por el contrario, la definición del PFQ *no manual* requiere el uso de dos tablas. En la descripción de la formalización aplicada en este trabajo expuesta en el Apartado 4, se establece que el PFQ *no manual* se define mediante la agrupación de unidades menores, designadas como ‘expresiones simples’, en ‘expresiones complejas’ (las unidades de este PFQ). Es por esta razón que la definición de este PFQ requiere el uso de dos entidades relacionadas: la primera de ellas recoge las unidades de este PFQ (‘expresiones complejas’) y la segunda, la descripción para todas las ‘expresiones simples’.
- El cuarto nivel de la base de datos contiene la información necesaria para la síntesis de cada una de las unidades descritas en el nivel tres. En el caso de los PFQ *configuración* y *movimiento*, en este nivel se incluyen las estructuras necesarias para la descripción de distintas variaciones emocionales en la definición de una misma unidad.

La descripción de las unidades correspondientes a cada PFQ se realiza mediante uno de los tres tipos de definición siguientes: *Quaternions*, *Nombres de Huesos* o *Coordenadas*. La definición de los PFQ se realiza mediante una de éstas, salvo el PFQ *localización* que permite la combinación de dos. A continuación se detallan los tres tipos:

- El *Quaternion* es un vector tetradimensional que se utiliza para describir la orientación de un objeto en un entorno tridimensional. La definición detallada y el criterio de elección de este tipo de representación se han descrito en el Apéndice B.

Los PFQ *configuración*, *orientación* y *no manual* establecen relaciones entre una determinada articulación y un *quaternion* para definir la unidad correspondiente. El número de *quaternions* necesarios para dicha definición está en relación uno a uno con los huesos implicados en dicha unidad. Los PFQ *configuración* y *orientación* definen una relación constante de quince y uno respectivamente. En el caso del PFQ *no manual* este número dependerá de las ‘expresiones simples’ que se requieran para definir la ‘expresión compleja’.

- Un *Nombre de Hueso* del esqueleto se utiliza en la definición de dos PFQ: *localización* y *punto de contacto* para definir un punto en la anatomía del avatar. Se ha comentado cómo la definición del PFQ *localización* se realiza

mediante referencias a una parte de la anatomía del signante. La aproximación seguida para este tipo de definición ha sido la utilización de los huesos del esqueleto en lugar de los vértices de la malla. Los movimientos del avatar impiden que las posiciones anatómicas requeridas para este PFQ sean definidas de manera estática. Por lo tanto, estas posiciones o coordenadas se obtienen de manera dinámica a partir de la posición de los huesos requeridos (ver Punto 6.1.2.3). El PFQ *punto de contacto* establece la parte activa de la mano durante la ejecución de un signo, definiendo dicho PFQ mediante la utilización del nombre del hueso que modela la articulación definida como parte activa.

- Una *Coordenada* es un vector 3D utilizado para los PFQ *plano* y *movimiento*. En el caso del plano, las coordenadas definen la distancia horizontal al cuerpo del signante, de manera que sólo la componente *Z* de la coordenada utilizada es relevante.

Para definir un *movimiento*, es necesario establecer la trayectoria del mismo junto con determinada información dinámica relativa a su aceleración. Para cumplir ambos requisitos se utiliza una secuencia de cambios de posición en distintos instantes de la definición del movimiento, lo que se representa mediante vectores 3D. Estos vectores toman como origen del desplazamiento la posición previa de la mano sobre la que están definiendo el movimiento.

Para que los valores asignados a estas coordenadas sean independientes del avatar utilizado inicialmente, se ha optado por el hecho de que los valores almacenados en la base de datos no representen distancias absolutas: las coordenadas almacenadas están expresadas como fracción de la longitud del húmero del avatar. De igual modo, la definición de las coordenadas del PFQ *plano* es independiente de la posición del avatar con respecto al origen de coordenadas de la escena. La distancia horizontal definida para este PFQ toma como referencia la posición de uno de los huesos del esqueleto; se ha optado por utilizar el hueso “CHEST_LOCATION_BONE” como origen de coordenadas para el PFQ *plano*.

El hecho establecido de que el PFQ *localización* utilice para su definición tanto el *Nombre de un Hueso* como una *Coordenada* permite establecer localizaciones relativas a una referencia anatómica. De este modo, podremos definir *localizaciones* como, por ejemplo, sobre el hombro o a la derecha de la oreja.

Para completar la descripción de la estructura de la base de datos y facilitar su interpretación, se propone el siguiente símil que pretende describir los niveles de la

base de datos relacional implementada en esta tesis en los términos equivalentes a la síntesis de voz:

El nivel 1 almacenaría cada palabra o concepto, de manera similar a una entrada de diccionario. El nivel 2 establecería la secuencia de unidades básicas (fonemas) que forma dicha palabra. El nivel 3 contendría la lista de todos los fonemas que se requieren para dicho idioma. Por último, el nivel 4 especificaría cómo deben ser generadas cada una de las unidades del idioma. Por lo tanto, esta aproximación tiene un gran parecido al planteamiento que rige el sintetizador de voz por formantes de Klatt [64]. La gran diferencia es que, mientras que para la síntesis de voz sólo se define una secuencia de parámetros¹, la síntesis de signos requiere un número mayor; siendo cinco el número mínimo de secuencias para un signo monomanual y trece el máximo en el caso de signos bimanuales con definición del PFQ *no manual*.

La analogía anterior muestra que la relación entre los niveles tres y cuatro de la base de datos es de «1-a-N»: cada fonema puede generarse de distintas maneras según el usuario o, incluso, el estado emocional, lo que implica que la representación de la LS puede verse alterada de igual manera. La definición básica de cada uno de los signos equivale, en cierta medida, a la definición que pueda obtenerse de las notaciones SiGML o SWML. La principal diferencia estriba en que la definición almacenada en la base de datos ha sido modificada, a partir de la versión básica, para generar un resultado más natural. Como veremos en el Apartado 12, las herramientas desarrolladas permiten aplicar matices a dichas definiciones. De igual manera, el diseño de la base de datos permite almacenar definiciones alternativas de los signos, que se aplican en función de los parámetros definidos en la notación HLSML expuestas en el Apartado 8.

La Figura 7.2 muestra el diagrama de tablas de la base de datos relacional que proponemos. Para facilitar su estudio, el Apéndice C muestra este mismo diagrama con un tamaño mayor.

7.2. Descripciones paralelas

La base de datos contiene diferentes representaciones para cada uno de los signos, lo que responde a tres objetivos:

¹Suponiendo la síntesis de un idioma no tonal y sin modificaciones prosódicas.

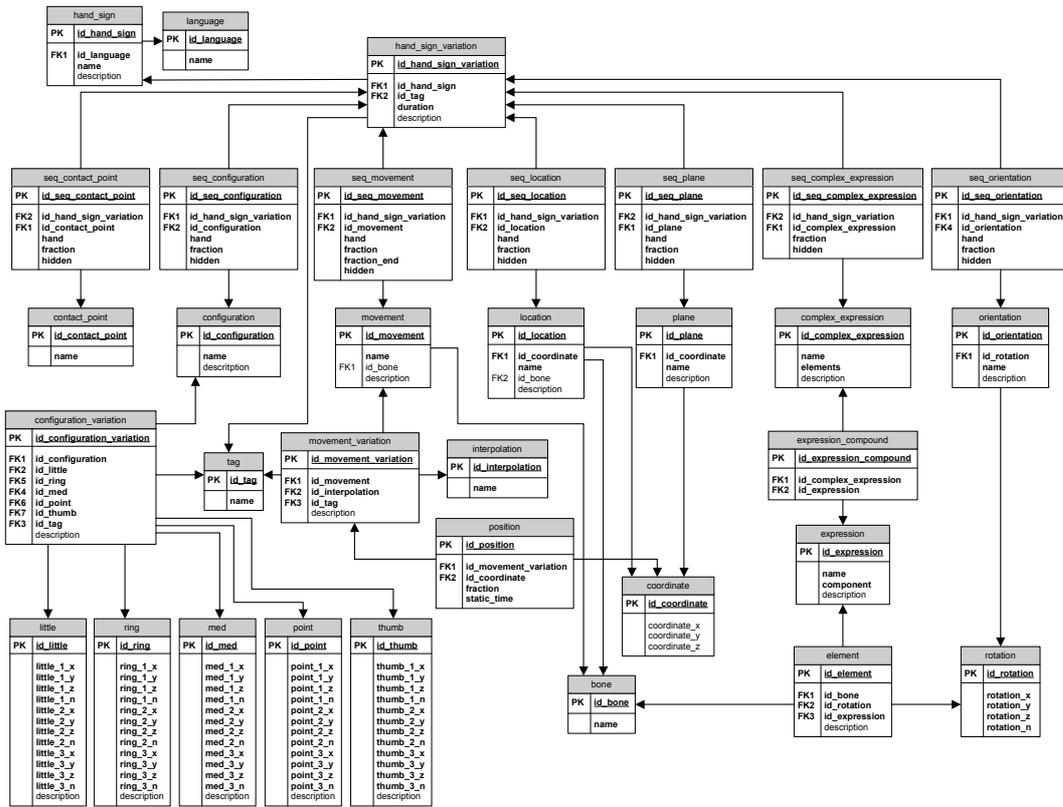


Figura 7.2: Diagrama detallado de la base de datos.

1. Posibilitar el almacenamiento de las distintas variaciones dialécticas que existen en la LSE, lo que se podría aplicar para almacenar en la base de datos relacional otras lenguas de signos diferentes de la LSE.
2. Permitir la representación de distintos estados emocionales en la síntesis de LSE.
3. Generar variabilidad en la síntesis de signos. Esta variabilidad hace referencia a la imperfección inherente en las personas cuando signan, ya que no lo hacen siempre de idéntica manera.

El método de elección entre las distintas alternativas se comentará en el Apartado 8.

7.2.1. Variaciones de idioma y dialectales

Al igual que pasa con el español, la LSE presenta variaciones dialectales entre distintas regiones de España. Esto implica que si se desea alcanzar una solución lo más amplia posible, se debe contemplar dicha variabilidad. Por otra parte, se debe

simplificar el proceso de descripción del mensaje a sintetizar, evitando duplicidades innecesarias en el número de entradas del diccionario.

Es en el primer nivel de la base de datos relacional donde se realizan las modificaciones relativas al idioma. La Figura 7.3 muestra su implementación en la base de datos.

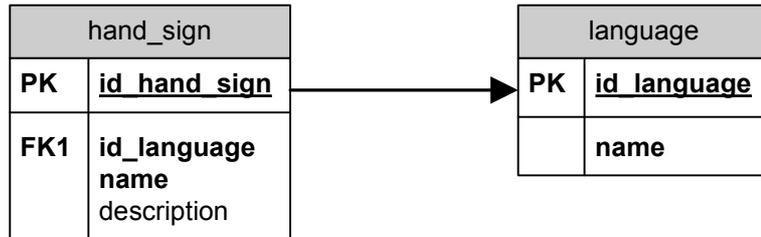


Figura 7.3: Estructura de la base de datos que permite la definición de variaciones dialectales y de idioma.

7.2.2. Variaciones emocionales

El segundo objetivo para proponer la existencia de representaciones alternativas en la base de datos relacional es el de definir las variaciones pertinentes para la síntesis emocional. La primera modificación está en el primer nivel. Como puede observarse en la Figura 7.4, la definición de variaciones emocionales («hand_sign_variation»), junto con la tabla que describe estas variaciones («tag»), establece la duración del signo. Así, por ejemplo, una variación en la que se muestre una mayor tensión del signante, se caracteriza por reducir la duración total del signo, mientras que una ejecución relajada del mismo signo tendrá una duración mayor.

Como hemos visto antes, cada variación emocional en la síntesis de un signo genera una nueva entrada en la tabla «hand_sign_variation». Si se observa la Figura C.1, todas las secuencias para cada uno de los siete PFQ establecen una relación con esta tabla. Es decir, cada variación emocional permite definiciones diferentes de estas secuencias, lo que mejora el resultado. Esto se produce sin que haya modificación en las unidades referenciadas sino mediante alteraciones de la fracción de tiempo en que se ejecutan, lo que implica alterar la aceleración de los distintos PFQ durante la ejecución del signo.

No todos los PFQ se ven afectados por las variaciones emocionales, la *orientación*, *punto de contacto*, *localización* y *plano* son PFQ que no lo están. Por el contrario, la definición de los PFQ *configuración*, *movimiento* y *no manual* si se ven modificados por la componente emocional. Dichas modificaciones se definen en el nivel cuatro de la base de datos.

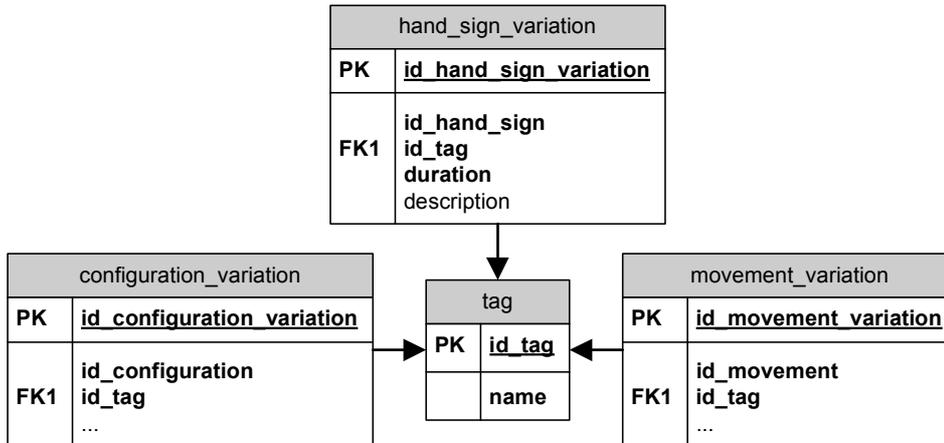


Figura 7.4: Estructura de la base de datos que permite la definición de variaciones emocionales.

Como puede verse en la Figura 7.4, la *configuración* y el *movimiento* están relacionados con la tabla de descripción de estados emocionales. Lógicamente, las modificaciones que apliquen a estos PFQ no deben disminuir la inteligibilidad del signo. Por ejemplo, una variación en la emoción con que se representa un signo hará que la tensión muscular de la mano del signante varíe, aunque el significado se mantiene. La modificación produce pequeñas variaciones en la flexión de los dedos a consecuencia de un grado distinto de tensión muscular.

El siguiente PFQ que depende de la tabla «tag» es el *movimiento*. En función del estado emocional del signante se producen alteraciones en la trayectoria, aceleración y tensión con la que se ejecuta un movimiento. Las modificaciones sobre este parámetro, deben realizarse con mayor cuidado que en el caso de la *configuración* ya que, en la LS, una variación en la tensión o aceleración de un movimiento puede modificar el significado del signo. Por lo tanto, la definición de una variación en el *movimiento*, sólo se implementará en aquellos signos en los que no pueda dar lugar a confusión.

Por último queda la definición del PFQ *no manual*. No consideramos necesario definir variaciones de una expresión en función del estado emocional ya que, conceptualmente hablando, se trata de expresiones diferentes. Esto se aplica tanto a la expresión facial como a la corporal. Al igual que con el *movimiento*, se debe prestar atención en no modificar elementos relevantes para el significado del signo.

7.2.3. Variaciones genéricas

La organización de la información de una base de datos relacional permite múltiples definiciones de un mismo elemento, lo que permite dotar al avatar de cierto grado de

variabilidad en la síntesis de signos. De esta manera, la ejecución de un signo no será exactamente igual todas las veces que se sintetice. Esto se logra mediante distintas definiciones para cada una de las unidades de un PFQ, siendo una modificación que afecta únicamente al nivel cuatro de la base de datos.

7.3. Resumen

Este apartado presenta la estructura de la base de datos relacional utilizada para almacenar la definición de los PFQ, los *signos establecidos* y el diccionario *dactilológico*. Esta base de datos, organizada en cuatro niveles lógicos, permite recoger definiciones alternativas de los signos y los PFQ para representar las variantes de idioma de las LS, variantes emocionales y variantes genéricas que aporten naturalidad a la representación del mensaje.

Apartado 8

HLSML. Notación de entrada

Este apartado presenta la notación HLSML (High Level Signing Markup Language) que ha sido desarrollada para este trabajo y que cumple los requisitos planteados en la Sección 4.4. En primer lugar se muestra la forma en que se definen los distintos elementos semánticos de la LSE, así como los elementos y atributos en la definición de HLSML que permiten matizar la síntesis de la LSE aplicando variaciones emocionales y prosódicas. A continuación se presenta una comparación entre HLSML y las notaciones utilizadas para la síntesis de LS en formato XML.

8.1. Descripción del HLSML

Los objetivos planteados para la creación de HLSML se resumen de la siguiente manera:

- Crear una notación basada en XML que pueda ser utilizada por personas con escaso conocimiento de la LSE y que al mismo tiempo sirva de comunicación entre un módulo de traducción y este sintetizador.
- Definir las tres unidades semánticas (los signos *establecidos*, el diccionario *dactilológico* y los *clasificadores*) en la misma notación.
- Permitir el uso de modificadores para variar la prosodia y la componente emocional en la representación del mensaje.
- Contemplar el comportamiento existente en la LSE que permite definir de manera paralela a la ejecución de una secuencia de signos *establecidos*, *clasificadores* y acciones no manuales.

La definición de las tres unidades semánticas de la LSE en la notación HLSML requiere el uso de dos aproximaciones distintas, una para los signos *establecidos* y el diccionario *dactilológico* y otra para los *clasificadores*. Tanto los signos *establecidos* como las unidades del diccionario *dactilológico* están descritos en la base de datos relacional (ver Apartado 7), por lo que, HLSML sólo necesita definir la glosa del signo o la letra a representar. Esto permite que, para definir un mensaje utilizando esta notación, no sea necesario conocer cómo se realizan los signos. La descripción de los clasificadores varía en complejidad según la naturaleza del clasificador que se esté definiendo.

8.1.1. Descripción de una secuencia de signos *establecidos*

La definición de un signo en HLSML se realiza mediante la etiqueta `<sign>` en el que se define la glosa del signo a realizar como un valor de su atributo `name`. Un conjunto de estas etiquetas pueden agruparse bajo `<sentence>`, que representa un bloque de síntesis que será procesado y generado en su conjunto.

La Figura 8.1 muestra un ejemplo de la notación HLSML que, en este caso, describe la secuencia de tres signos YO CASA IR. La descripción y la duración de cada uno de estos signos se obtiene de la base de datos relacional. El tiempo de transición entre dos signos consecutivos es de 300ms, valor establecido por defecto.

8.1.2. Descripción de un signo mediante sus PFQ

Aunque el objetivo principal del HLSML es definir un mensaje de la manera más sencilla posible y sin necesidad de conocer la descripción de los signos, no resulta conveniente obviar la posibilidad de definir un signo a bajo nivel. Podríamos plantearnos situaciones en las que esta posibilidad es necesaria, por ejemplo, para definir los signos utilizados para hacer referencia a personas o conceptos relacionados con la conversación. Estos signos no pertenecen a un diccionario oficial, ya que son creados por el signante y por tanto, no están contemplados en la base de datos relacional. Otra situación se produce con los nuevos conceptos ya que el signo que los representa no se establece de manera inmediata, por lo que carecen inicialmente de definición en la base de datos relacional, y por esto debe permitirse su definición en la notación de entrada.

La definición de un signo en HLSML se realiza mediante la etiqueta `<signDefinition>` que agrupa el conjunto de secuencias de cada uno de los PFQ que lo componen. La escala de tiempo utilizada para describir estas secuencias es la misma

```

<!DOCTYPE hlsml SYSTEM "hlsml.dtd">
<hlsml>
<sentence language="es"
           tag="standard">
  <!-- Figura (b) -->
  <sign name="yo" />

  <!-- Figura (c) -->
  <sign name="casa" />

  <!-- Figura (d) -->
  <sign name="ir" />
</sentence>
</hlsml>

```



(a) Definición HLSML.

(b) Signo YO.



(c) Signo CASA.



(d) Signo IR.

Figura 8.1: Ejemplo de la definición de una secuencia de signos para ser sintetizados. Se ha resaltado la sección del mensaje que define los *signos establecidos* que deben sintetizarse.

que se aplica en la base de datos relacional: un porcentaje sobre la duración total del signo. Todos los PFQ se definen mediante una lista de etiquetas `<fraction>`, que define este porcentaje sobre la duración del signo. Cada `<fraction>` contiene una etiqueta `<element>` que representa la unidad del PFQ que se debe ejecutar. En la Figura 8.2 se presenta un fragmento de mensaje en HLSML que define el signo LIBRO.

La descripción del PFQ *no manual* en HLSML, mediante la etiqueta `<nonHand-Sequence>`, permite utilizar junto a `<fraction>` la etiqueta `<talk>`. Éste permite definir de manera sencilla una secuencia de visemas, utilizando `<viseme>`, con el objetivo de generar la animación labial en el avatar. Cada `<viseme>` describe el visema que se debe presentar y el instante de tiempo en el que se alcanza dicha

```

<!DOCTYPE hlsml SYSTEM "hlsml.dtd">
<hlsml>
<sentence language="es"
      tag="standard">
  <signDefinition time="500">
    <configurationSequence side="left">
      <fraction percentage="0">
        <element name="config_043"/>
      </fraction>
    </configurationSequence>
    <orientationSequence side="left">
      <fraction percentage="0">
        <element name="h_p_000_b"/>
      </fraction>
      <fraction percentage="36">
        <element name="h_p_000_a"/>
      </fraction>
      <fraction percentage="73">
        <element name="h_p_000_b"/>
      </fraction>
      <fraction percentage="100">
        <element name="h_p_000_a"/>
      </fraction>
    </orientationSequence>
    <locationSequence side="left">
      <fraction percentage="0">
        <element name="chest"/>
      </fraction>
    </locationSequence>
    <planeSequence side="left">
      <fraction percentage="0">
        <element name="pla_2"/>
      </fraction>
    </planeSequence>
    <contactSequence side="left">
      <fraction percentage="0">
        <element name="little_4"/>
      </fraction>
    </contactSequence>
  </signDefinition>
</sentence>
</hlsml>

```

(a) Definición HLSML.



(b) Signo LIBRO.

Figura 8.2: Ejemplo de la definición de un signo en el mensaje HLSML. El código aquí mostrado recoge únicamente la definición de la mano izquierda; el mensaje real duplica todas las etiquetas <(PFQ)Sequence> para la mano derecha. Se ha destacado la etiqueta que define una descripción basada de PFQ.

unidad. Se ha optado por seguir este formato para facilitar la adaptación de la salida obtenida producida por un reconocedor de voz.

8.1.3. Descripción del diccionario dactilológico

Cada elemento del diccionario *dactilológico* se almacena en la base de datos relacional bajo una glosa de la forma ‘letter_?’, donde ‘?’ corresponde a cada una de las letras del alfabeto español. Por lo tanto, la definición una secuencia deletreada equivaldría a una secuencia de signos *establecidos*. Esta aproximación penaliza en gran medida la usabilidad de la notación, por lo que se ha optado por definir una secuencia a deletrear utilizando directamente la propia secuencia de letras.

En la notación HLSML se establece la etiqueta `<spellSign>` para definir una palabra a deletrear mediante el diccionario dactilológico, la cual se recoge en el valor del atributo `name`. La Figura 8.3 muestra un ejemplo de la notación HLSML para generar una animación en la que se deletrea la palabra «libr» o.

```
<!DOCTYPE hlsml SYSTEM "hlsml.dtd">
<hlsml>
<sentence language="es" tag="standard">
  <spellSign name="libro" />
</sentence>
</hlsml>
```

(a) Definición HLSML.

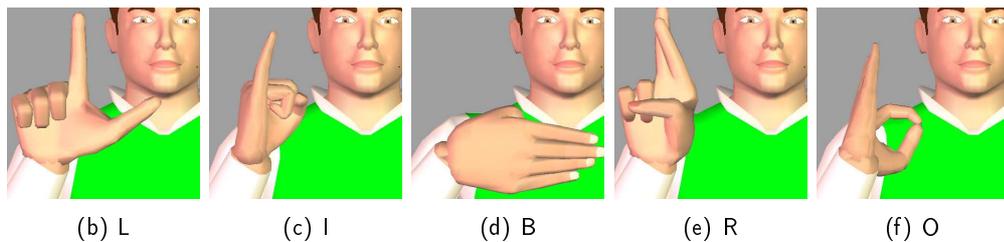


Figura 8.3: Ejemplo de deletreo de una palabra con el diccionario dactilológico. Se destaca la etiqueta que define la secuencia de deletreo.

8.1.4. Descripción de los *clasificadores* en HLSML

Los *clasificadores* son la unidad de la LS que mayor componente semántica posee, lo que impide que puedan ser definidos con anterioridad al proceso de síntesis. Por esta razón, la descripción de los *clasificadores* debe realizarse en la notación HLSML.

En este punto se describirá a los distintos *clasificadores* siguiendo el orden en que fueron expuestos en la Subsección 2.2.1.

8.1.4.1. Clasificadores nominales

Esta categoría de *clasificadores* se define como una secuencia de dos signos *establecidos* que unen sus significados para generar un nuevo concepto. Al ser morfológicamente equivalentes a una secuencia de dos signos cualesquiera, la definición en HLSML de esta construcción es similar a la descrita en la Subsección 8.1.1. No obstante, como definimos en la Sección 4.2, se puede disminuir el tiempo de transición entre estos dos signos para enfatizar la construcción clasificatoria. La Figura 8.4 muestra la definición en HLSML de una construcción de clasificador nominal: el signo AGUA actúa como nombre clasificatorio sobre el signo CAUCE. Además, se ha establecido que la transición entre ambos signos sea de 100ms en lugar de los 300ms que se aplican por defecto.

8.1.4.2. Clasificadores introflexivos

La clasificación *introflexiva* de un signo *establecido* supone una alteración en la definición de uno o varios de sus PFQ, normalmente del PFQ *configuración*. El modo de definir esta clasificación es indicar el PFQ del signo que se altera (signo modificado) y cuál es el nuevo valor que toma. El nuevo valor se extrae en la mayoría de los casos de la definición de un segundo signo (signo modificador), por lo tanto, el método más sencillo de definir este tipo de *clasificadores* es establecer el PFQ modificado y la glosa del signo modificador.

En HLSML se describe la clasificación *introflexiva* mediante la etiqueta `<sign-Classifier>`, que aparece supeditado a la etiqueta `<sign>` que modifica. El valor del atributo `name` define el PFQ afectado, mientras que el nuevo valor para dicho PFQ se define mediante la etiqueta hija de `<signClassifier>`, pudiendo ser una de los dos siguientes: `<sign>` establece la glosa del signo modificador o bien `<element>` especifica directamente el nombre de una unidad para el PFQ a modificar.

En la Figura 8.5 se muestra un ejemplo de esta construcción clasificatoria. Al sintetizar el signo TELEVISION se modifica su PFQ *configuración* tomando el valor utilizado en la *configuración* del signo VOLCARSE, para representar que la televisión se ha volcado.

```

<!DOCTYPE hlsml SYSTEM "hlsml.dtd">
<hlsml>
<sentence language='es'
           tag='standard'>
  <sign name="agua">
    <signMod timeNext="100" />
  </sign>
  <sign name="cauce"/>
</sentence>
</hlsml>

```

(a) Definición HLSML.



(b) Signo AGUA.



(c) Signo CAUCE.

Figura 8.4: Ejemplo de un *clasificador nominal*. El signo AGUA (marcado en azul) actúa como un nombre clasificatorio sobre el signo CAUCE. La combinación de ambos genera el concepto «río».

8.1.4.3. Clasificadores icónicos

Los *clasificadores icónicos* son representaciones mímicas realizadas por el signante de una determinada escena. Su definición puede requerir tanto de elementos estáticos: una mano se sitúa en una determinada posición y adquiere una forma y orientación definidas o que el avatar mire hacia un punto determinado del espacio o adquiera una determinada expresión, como dinámicos: el avatar realiza un determinado movimiento con las manos, varía su expresión, o sigue un objeto con la mirada. Por lo tanto, la descripción del *clasificador icónico* debe contemplar tanto una definición estática como una dinámica.

La definición estática de un *clasificador icónico* se realiza a través de la etiqueta `<classifier>`, el cual permite definir la forma, orientación y posición de ambas

```

<!DOCTYPE hlsml SYSTEM "hlsml.dtd">
<hlsml>
<sentence language='es'
           tag ='standard'>
  <sign name="television">
    <signClassifier
      name="configuration">
      <sign name="volcarse"/>
    </signClassifier >
  </sign>
</sentence>
</hlsml>

```

(a) HLSML definition.



(b) Signo original.



(c) Signo modificado.

Figura 8.5: Ejemplo de un *clasificador intreflexivo*, la ejecución del signo TELEVISIÓN modifica su *configuración* por la del signo VOLCARSE (marcado en azul).

manos (mediante las etiquetas `<hand>`), el componente no manual (mediante la etiqueta `<nonHand>`) y la dirección a la que debe orientar la cara y los ojos del avatar (`<headLookAt>` y `<eyesLookAt>` respectivamente). La etiqueta `<classifier>` define, entre otros, el atributo `armDivision` con un valor entero. Este valor establece el número de partes en las que se debe dividir la longitud del húmero para definir las coordenadas en el espacio de signado.

Cada mano (`<hand>`) requiere definir tres PFQ (*configuración*, *orientación* y *punto de contacto*) y una posición en el espacio. Esta posición sigue el planteamiento definido en la Sección 4.2 (ver Figura 4.6, y queda representada en la notación HLSML por la etiqueta `<coordinate>`). Las coordenadas (X', Y', Z') en el sistema de referencia de la escena se calculan mediante la Ecuación 8.1, siendo (X_p, Y_p, Z_p) la posición del «hueso de localización» que define el centro del pecho

("CHEST_BONE_LOCATION"), AD establece el número de divisiones de la longitud del húmero (*longHúmero*) lo que permite variar la precisión con la que se define la posición. Por último los valores X, Y, Z siempre son enteros. El punto al que el avatar dirige su mirada o la cara se define de igual manera que las posiciones de las manos, utilizando la etiqueta `<coordinate>`. Por último, la definición del componente no manual en el *clasificador* se realiza de manera similar a la definición de signos mediante sus PFQ, indicando el nombre de la unidad de las definidas para el PFQ *no manual*.

```

<classifier
  armDivision="AD">
  ...
  <coordinate
    width ="X"
    height="Y"
    depth ="Z"/>
  ...
</classifier>

```

$$(X', Y', Z') = \frac{\text{longHúmero}}{AD} \cdot (X, Y, Z) + (X_p, Y_p, Z_p) \quad (8.1)$$

Para representar un *clasificador icónico* dinámico, se genera una secuencia de posiciones estáticas separadas por intervalos de tiempo. Esta secuencia se representa en el mensaje HLSML mediante la etiqueta `<classifierSequence>` y el intervalo de tiempo que debe transcurrir entre distintas posiciones estáticas mediante el uso de `<classifierTransition>`. La Figura 8.6 muestra el código HLSML necesario para definir un sencillo *clasificador icónico* dinámico.

8.1.4.4. Clasificadores afijales

El último tipo de *clasificador* es el *afijal*, que previamente hemos descrito como una unión entre los *clasificadores introflexivos* y los *clasificadores icónicos*. La descripción de este tipo de clasificadores implica incorporar a la notación HLSML un elemento que defina la ejecución concurrente de las etiquetas hijas, dicha etiqueta es `<compound>`. Para establecer la modificación de este clasificador sobre un signo, se reutiliza la etiqueta `<signClassifier>` definida para los *clasificadores introflexivos*. En este caso, al atributo `name`, que define el PFQ a modificar, se le asigna como valor la cadena *affixal*.

La Figura 8.7 muestra el fichero HLSML necesario para representar el siguiente mensaje: el signante realiza el signo correspondiente a una naranja; a continuación, utiliza su mano izquierda para situar en el espacio la posición de esta naranja; inmediatamente después y sin mover la mano izquierda, comienza a ejecutar el signo

```

<!DOCTYPE hlsml SYSTEM "hlsml.dtd">
<hlsml>
<sentence language="es" tag="standard" >
  <classifierSequence>
    <!-- Figura (b) -->
    <classifier armDivision="7" time="700">
      <hand side="right">
        <configuration>
          <element name="config_006" />
        </configuration>
        <orientation>
          <element name="f_p_000_d" />
        </orientation>
        <coordinate width="-5" height="2"
                    depth="4" />
        <contactPoint>
          <element name="point_4" />
        </contactPoint>
      </hand>
    <headLookAt>
      <coordinate width="-5" height="0"
                  depth="4" />
    </headLookAt>
    <eyesLookAt>
      <coordinate width="-5" height="2"
                  depth="4" />
    </eyesLookAt>
  </classifier>
  <!-- Figura (c) -->
  <classifierTransition time="400" />
  <!-- Figura (d) -->
  <classifier armDivision="7" time="700">
    <hand side="right">
      <configuration>
        <element name="config_006" />
      </configuration>
      <orientation>
        <element name="f_p_000_d" />
      </orientation>
      <coordinate width="1" height="2"
                  depth="4" />
      <contactPoint>
        <element name="point_4" />
      </contactPoint>
    </hand>
    <headLookAt>
      <coordinate width="1" height="0"
                  depth="4" />
    </headLookAt>
    <eyesLookAt>
      <coordinate width="1" height="2"
                  depth="4" />
    </eyesLookAt>
  </classifier>
</classifierSequence>
</sentence>
</hlsml>

```

(a) Definición HLSML.



(b) Posición inicial.



(c) Transición.



(d) Posición final.

Figura 8.6: Ejemplo de un *clasificador icónico*. El avatar describe una escena en la que hay una persona caminando a lo largo de un camino recto.

MIRAR pero modificando la parte final del signo para que la mano termine el movimiento orientada y situada junto a la mano que representa la naranja.

8.1.5. Definición de variaciones y elementos prosódicos

Aunque hemos presentado la etiqueta `<compound>` en el apartado de los *clasificadores afijales*, su uso no se restringe a la definición de éstos. Este elemento incluye entre los posibles descendientes de la etiqueta `<nonHandSequence>`, con lo que se puede especificar las expresiones que se desea que el avatar adquiera mientras representa un mensaje signado.

En el Apartado 7 se definió la forma en que pueden almacenarse representaciones alternativas de un mismo signo, dejando pendiente la descripción del método para escoger cuál de estas alternativas debe ser recuperada de la base de datos relacional. A continuación veremos los atributos utilizados en el mensaje HLSML para concretar la recuperación de información de la base de datos y además, otros que permiten modificar parámetros de la síntesis como son la velocidad de reproducción de un signo o el tipo de interpolación que debe utilizarse en el proceso de «Render» (ver Apartado 5), y son los siguientes:

- **anim.** Define el modo en que se realiza la interpolación en la animación, modificando los parámetros de la interpolación, se consigue variar la fluidez con que se produce el movimiento del avatar.
- **compensate.** Establece la opción de utilizar o no el PFQ *punto de contacto* para definir la posición de las manos, compensando así la posición de la muñeca.
- **language.** Define el idioma o dialecto en que se representan los signos, modificando el proceso de recuperación de información desde la base de datos relacional.
- **speed.** Modifica la velocidad de reproducción de un signo.
- **tag.** Define la variación emocional que debe aplicarse al recuperar la descripción de un signo de la base de datos relacional.
- **tagCriteria.** Establece el criterio de selección cuando se plantean definiciones alternativas de un signo o de un PFQ para una misma variante emocional y de idioma.

```

<!DOCTYPE hlsml SYSTEM "hlsml.dtd">
<hlsml>
<sentence language="es" tag="standard">
  <!-- Figura (b) -->
  <sign name="naranja_fruto"/>
  <!-- Figura (c) -->
  <classifierSequence>
    <classifier armDivision="3"
      time="700">
      <hand side="left">
        <configuration>
          <element name="config_078"/>
        </configuration>
        <orientation>
          <element id="91"/>
        </orientation>
        <coordinate width="2" height="1"
          depth="3"/>
        <contactPoint>
          <element name="med_4"/>
        </contactPoint>
      </hand>
    </classifier>
  </classifierSequence>
  <!-- Figura (d) -->
  <compound>
    <classifierSequence>
      <classifier armDivision="3"
        time="700">
        <hand side="left">
          <configuration>
            <element name="config_078"/>
          </configuration>
          <orientation>
            <element id="91"/>
          </orientation>
          <coordinate width="2" height="1"
            depth="3"/>
          <contactPoint>
            <element name="med_4"/>
          </contactPoint>
        </hand>
      </classifier>
    </classifierSequence>
    <sentence>
      <sign name="mirar">
        <signClassifier
          name="affixal">
          <element name="l_palm"/>
        </signClassifier>
      </sign>
    </sentence>
  </compound>
</sentence>
</hlsml>

```

(a) Definición HLSML.



(b) Signo NARANJA.



(c) Posición del objeto.



(d) Signo MIRAR modificado.



(e) MIRAR original.

Figura 8.7: Ejemplo de un *clasificador afijal*. El signo MIRAR se modifica para que la mano se oriente hacia la posición de la naranja y se sitúe junto a la misma.

- `timeInter`, `timeNext` y `timePrev`. Estos tres atributos se utilizan para definir el tiempo de transición entre dos signos consecutivos. El atributo `timeInter` se utiliza a nivel global, mientras que `timeNext` y `timePrev` definen en un signo el tiempo que se debe dejar con el siguiente y el anterior, respectivamente.

Los atributos que hemos descrito antes pueden aplicarse a una frase en su totalidad, a un conjunto de signos o a un único signo en particular:

- Si pertenecen a la etiqueta `<sentence>`, todos los signos se modifican.
- Si están en `<globalMod>`, sólo se modifican los signos que están después de este elemento.
- Si pertenecen a la etiqueta `<signMod>` las modificaciones afectan únicamente al signo que contiene dicho elemento.

8.2. Comparación de HLSML con SiGML y SWML

El objetivo de esta sección es comparar HLSML con las otras dos notaciones basadas en XML: SiGML y SWML. Para ello se presenta el mismo mensaje utilizando las tres notaciones. El mensaje descrito consta de tres signos YO AQUÍ CRECER. Las notaciones SiGML y SWML no requieren definir la glosa del signo, sin embargo, para facilitar la comparación, se ha incorporado la especificación de dicho valor en el mensaje. Con el mismo propósito se ha resaltado en azul la sección relativa al signo YO en cada uno de los tres mensajes.

El nivel de abstracción de la definición de HLSML es muy superior a SiGML y SWML, lo que permite definir un mensaje compuesto por una secuencia de signos *establecidos* de forma más sencilla en HLSML. Además, dado que la definición de dichos signos está almacenada en la base de datos relacional, la longitud del mensaje en HLSML es apreciablemente inferior.

Aunque en este ejemplo sólo se ha propuesto una secuencia de signos, SiGML y SWML también permiten definir elementos del diccionario dactilológico y *clasificadores introflexivos*. Sin embargo, la complejidad de estas definiciones es elevada debido a que deben ser descritos en función de los PFQ que los forman. La descripción de los *clasificadores icónicos* y *afijales* no es posible utilizando estas notaciones, dado que no permiten especificar la posición de las manos de manera continua o el punto al que deben mirar los ojos o dirigir la cara. Además las notaciones SiGML y SWML no permiten definir una composición paralela de distintos elementos.

HLSML

```
<?xml version="1.0"
      encoding="iso-8859-1"?>
<!DOCTYPE hlsml SYSTEM "hlsml.dtd">
<hlsml>
  <sentence language="es"
            tag="standard">
    <sign value="yo"/>
    <sign value="aqui"/>
    <sign value="crecer"/>
  </sentence>
```

SiGML

```
<?xml version="1.0"
      encoding="iso-8859-1"?>
<!DOCTYPE sigml SYSTEM
  "http://www.visicast.cmp.uea.ac.
  uk/sigml/sigml.dtd">
<sigml>
  <hamgestural_sign gloss="Y0">
    <sign_manual both_hands="false">
      <handconfig handshape="finger2"/>
      <handconfig extfidir="i"/>
      <handconfig palmor="r"/>
      <location_bodyarm location="chest"
                        contact="touch"/>
    </hamgestural_sign>
  <hamgestural_sign gloss="AQUI">
    <sign_manual both_hands="false"/>
    <handconfig handshape="finger2"/>
    <handconfig extfidir="d"/>
    <handconfig palmor="r"/>
    <location_bodyarm
                      location="stomach"/>
    <rpt_motion
              repetition="fromstart">
      <directedmotion direction="d"/>
    </rpt_motion>
  </hamgestural_sign>
  <hamgestural_sign gloss="CRECER">
    <sign_manual both_hands="false"/>
    <handconfig handshape="flat"/>
    <handconfig extfidir="ol"/>
    <handconfig palmor="d"/>
    <location_bodyarm
                      location="stomach"/>
    <tgt_motion >
      <location_bodyarm
                    location="headtop"/>
    </tgt_motion>
  </hamgestural_sign>
</sigml>
```

SWML

```

<?xml version="1.0"?>
<swml version="1.0"
      file-type="text">
<!-- Y0 -->
<signbox>
  <symb x="50" y="53" x-flop="1"
        y-flop="1">
    <category>01</category>
    <group>01</group>
    <symbnum>001</symbnum>
    <variation>01</variation>
    <fill>05</fill>
    <rotation>01</rotation>
  </symb>
  <symb x="29" y="74" x-flop="0"
        y-flop="0">
    <category>02</category>
    <group>01</group>
    <symbnum>001</symbnum>
    <variation>01</variation>
    <fill>01</fill>
    <rotation>01</rotation>
  </symb>
</signbox>
<!-- AQUÍ -->
<signbox>
  <symb x="31" y="51" x-flop="1"
        y-flop="1">
    <category>01</category>
    <group>01</group>
    <symbnum>001</symbnum>
    <variation>01</variation>
    <fill>02</fill>
    <rotation>01</rotation>
  </symb>
  <symb x="20" y="88" x-flop="1"
        y-flop="0">
    <category>02</category>
    <group>03</group>
    <symbnum>002</symbnum>
    <variation>01</variation>
    <fill>01</fill>
    <rotation>01</rotation>
  </symb>
</signbox>
<!-- CRECER -->
<signbox>
  <symb x="38" y="68" x-flop="0"
        y-flop="0">
    <category>01</category>
    <group>05</group>
    <symbnum>007</symbnum>
    <variation>01</variation>
    <fill>06</fill>
    <rotation>02</rotation>
  </symb>
  <symb x="43" y="40" x-flop="0"
        y-flop="0">
    <category>02</category>
    <group>03</group>
    <symbnum>001</symbnum>
    <variation>02</variation>
    <fill>01</fill>
    <rotation>01</rotation>
  </symb>
  <symb x="16" y="19" x-flop="0"
        y-flop="0">
    <category>10</category>
    <group>01</group>
    <symbnum>001</symbnum>
    <variation>01</variation>
    <fill>01</fill>
    <rotation>03</rotation>
  </symb>
</signbox>
</swml>

```

8.3. Resumen

Este apartado describe la notación HLSML que hemos definido para describir los mensajes de entrada al sintetizador. Esta notación permite definir de manera sencilla todos los elementos semánticos descritos en la LSE, ya sea haciendo referencia a la base de datos relacional o definiéndolos en el propio mensaje. Esta notación permite definir variaciones emocionales y prosódicas.

Apartado 9

Módulo de síntesis de gestos

En este apartado se expone el módulo de síntesis de gestos, prestando especial atención a la parte de «Generación de Pistas de Animación» comentado en la Subsección 5.1.4 (Figura 5.2). En la primera sección se define el conjunto de dependencias entre los distintos PFQ involucrados. En la segunda, se explica el algoritmo de Cinemática Inversa (IK) que se ha desarrollado para este trabajo. En la última sección se muestra el método implementado para eliminar las colisiones entre distintas partes del cuerpo del avatar.

9.1. Proceso de síntesis y dependencias entre PFQ

Los signos *establecidos* y las unidades del diccionario *dactilológico* se definen utilizando secuencias de PFQ. Como se presentó anteriormente, esta definición puede estar contenida en la base de datos relacional o en el mensaje HLSML. Aunque sería preferible su inclusión en la base de datos relacional donde puede ser validado por un experto, ambas aproximaciones son equivalentes y deben dar lugar al mismo resultado.

Aunque los PFQ se definen y gestionan de manera independiente, existen algunas relaciones de precedencia entre ellos que deben ser tenidas en cuenta durante la síntesis. Tras procesar las secuencias de PFQ, se obtiene una lista con las unidades de cada uno de los PFQ que intervienen en el signo en cuestión. Para obtener la descripción de estas unidades, se recupera la información desde el cuarto nivel de la base de datos relacional. Este proceso de consulta se realiza en cualquier caso, independientemente de si el signo ha sido recogido en la base de datos relacional o

en el mensaje HLSML. Una vez recuperadas las descripciones, el proceso de síntesis se desarrolla en las siguientes etapas:

- Primera etapa. En esta etapa se definen las pistas de animación relacionadas con los PFQ *no manual*, *configuración* y *orientación*, definidos mediante quaternions y completamente independientes de cualquier otro PFQ. Los valores obtenidos de la base de datos relacional se aplican directamente a los huesos correspondientes.
- Segunda etapa. Se establece el conjunto de posiciones estáticas de la mano no dominante, lo que requiere el proceso de los PFQ *localización* y *plano*. El PFQ *no manual* establece animaciones de la cabeza, cara o cuerpo del avatar que alteran la posición de las referencias anatómicas definidas por el PFQ *localización*, lo que explica la precedencia de dicho PFQ. Como se comentó en la Sección 2.1 la *localización* se define mediante referencias anatómicas, que nosotros describimos mediante huesos del esqueleto (Sección 4.1), para obtener las coordenadas del espacio de referencia de la escena en este plano frontal, se obtiene la posición del hueso correspondiente a dicha referencia anatómica; en caso necesario, se le suma el desplazamiento definido en la base de datos relacional. En este punto del proceso hemos definido los valores (X, Y) de la coordenada tridimensional; el valor Z se obtiene del *plano*, ya sea utilizando el valor definido en la base de datos relacional, una vez procesado según la posición y dimensión del avatar, o bien mediante el valor de la coordenada Z del «hueso de localización», en caso de que el *plano* así lo defina. Tras fijar la posición objetivo, el sintetizador debe calcular la posición que debe tomar la muñeca para que la parte activa de la mano, definida por el PFQ *punto de contacto*, alcance dicho objetivo. Para este cálculo se requiere que los PFQ *configuración* y *orientación* se hayan procesado previamente.
- Tercera etapa. Esta etapa comienza aplicando la secuencia de *movimientos* a la mano no dominante. Este PFQ se define utilizando vectores 3D que establecen el desplazamiento de la mano con respecto a la posición anterior. Esta aproximación no incluye la definición de la primera posición, ya que debe ser definida por los PFQ *localización* y *plano*. Por lo tanto, la gestión de éstos debe ser anterior a la gestión del *movimiento*.

En este punto del proceso de generación de las pistas de animación, la secuencia de posiciones que debe alcanzar la muñeca de la mano no dominante ha quedado definida. Dicha secuencia se procesa para detectar y evitar posibles colisiones entre la mano no dominante y el cuerpo o la cabeza del avatar, en

la transición entre dos posiciones consecutivas. Este proceso se describirá en la Sección 9.3. Para obtener las rotaciones de las articulaciones del hombro y del codo necesarias para que la muñeca alcance dicha secuencia de posiciones, se ha utilizado un proceso de IK simplificado. Esta simplificación ha sido posible debido a la estructura esquelética del avatar definida en el Punto 6.1.2.1. Este proceso de IK se detalla en la Sección 9.2.

- Etapas cuarta y quinta. Estas etapas son similares a la segunda y tercera, pero aplicadas a la mano dominante. Este orden surge de la posibilidad de que la *localización* de la mano dominante se describa haciendo referencia a la no dominante.

La definición de los signos *establecidos* y las unidades del diccionario *dactilológico* se diferencian con respecto a los clasificadores en la definición de la posición de las manos. En el caso de los primeros se realiza mediante PFQ y en el caso de los segundos mediante un sistema de coordenadas propio. La consecuencia directa es que las etapas dos y tres se unifican, del mismo modo que las etapas cuatro y cinco, manteniéndose la precedencia de la mano no dominante sobre la dominante.

Una vez realizado el proceso de síntesis de una unidad del mensaje mediante estas cinco etapas, la transición entre unidades se genera a través una animación aplicada a todos los «huesos de animación» del avatar entre la última orientación definida para la unidad i y la primera de la $i + 1$.

9.2. Cinemática inversa

La Cinemática Inversa (IK) es el proceso de cálculo de los valores que deben adoptar las articulaciones de una estructura esquelética para que su extremo, efector final, se sitúe en una determinada posición espacial (P_{destino}). Partiendo de esta condición se pueden añadir un conjunto de restricciones adicionales tales como incluir la definición de la orientación o posición de elementos intermedios de la cadena esquelética.

La estructura cinemática del brazo humano consta de tres articulaciones: hombro, codo y muñeca. Los movimientos del hombro, flexión-extensión, abducción-aducción y rotación longitudinal, definen tres grados de libertad para dicha articulación. En el caso del codo es una articulación más sencilla con un sólo grado de libertad, dado que la pronación-supinación del antebrazo se implementa en la muñeca. Por último, la muñeca posee dos grados de libertad, inherentes a su anatomía, en los movimientos de flexión-extensión y de aducción-abducción, a los que se añade un

tercero, el movimiento de pronación-supinación del antebrazo. Por lo tanto, en el brazo humano se definen siete grados de libertad, cuyos valores establecen la posición y orientación de la mano.

Con la estructura esquelética planteada en este trabajo, en la que la orientación de la muñeca se ha desvinculado de la cadena cinemática del brazo, el número de grados de libertad se reduce de siete a cuatro, lo que conlleva una simplificación del proceso de cinemática inversa. Sin embargo, una vez establecida la posición del hombro y definida la posición objetivo de la muñeca, existe indeterminación en la solución final: las posibles posiciones del codo definen una circunferencia sobre el eje hombro-muñeca como puede observarse en la Figura 9.1.

Para obtener la posición final del codo es preciso eliminar todas las posibles soluciones que impliquen una rotación anormal de alguna de las articulaciones del brazo o que den lugar a la colisión del brazo con el cuerpo (ver Sección 9.3) es decir, se eliminan aquellas que incumplen las restricciones establecidas por las limitaciones anatómicas de las articulaciones del hombro y de la muñeca. Aun así, la solución obtenida no es única, es necesario aplicar otro proceso adicional para establecer la posición final del codo, que además permite representar distintos grados de tensión muscular en las articulaciones del brazo lo que permite aumentar la expresividad del avatar.

9.2.1. Definiciones

En esta sección se presenta la notación y definiciones utilizadas en el proceso de IK y en el cálculo posterior para la resolución de la indeterminación que implica establecer la posición del codo.

Articulaciones. Se consideran puntuales y su posición se establece por el punto de pivote para el hueso relacionado. Se define el punto S como la articulación del hombro, el punto E como la articulación del codo y, por último, el punto W como la articulación de la muñeca.

Huesos. Se definen mediante los siguientes vectores: $\vec{l}_1 = \vec{SE}$ como el vector que representa al brazo (hueso húmero) y $\vec{l}_2 = \vec{EW}$ al antebrazo (huesos radio y cúbito). Los vectores \vec{l}_1 y \vec{l}_2 son, por tanto, abstracciones de dos elementos anatómicos reales como son el brazo y en antebrazo. Además se define el vector $\vec{l}_3 = \vec{SW} = \vec{l}_1 + \vec{l}_2$ como el vector hombro-muñeca, que define el eje de rotación sobre el que el hombro puede

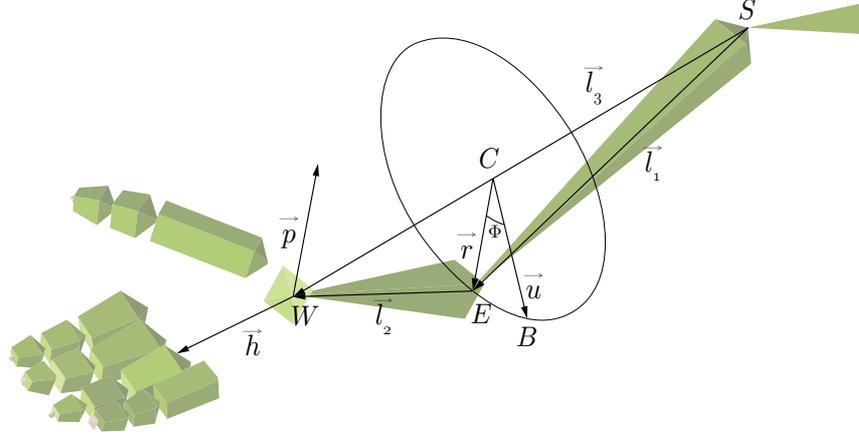


Figura 9.1: Diagrama de la estructura esquelética del brazo.

rotar sin variar la posición de la muñeca. El último vector relativo a la configuración del brazo y el antebrazo es \vec{e} , que define el semiespacio exterior resultado de la división del espacio por el plano definido por el brazo y el antebrazo. Este vector es calculado según la Ecuación 9.1.

$$\vec{e} = \begin{cases} \vec{l}_1 \times \vec{l}_2 & \text{brazo derecho} \\ \vec{l}_2 \times \vec{l}_1 & \text{brazo izquierdo} \end{cases} \quad (9.1)$$

Para definir la orientación de la mano son necesarios dos vectores: \vec{h} , como el vector longitudinal a la dirección de la mano y \vec{p} , como el vector normal a la palma de la mano.

Acto seguido definiremos un sistema de coordenadas ortonormal cuyo origen de coordenadas C (Ecuación 9.2) es el centro de la circunferencia formada por el conjunto de las posibles soluciones para la posición del codo y cuyos vectores base sean \vec{n} , \vec{u} y \vec{v} dados, respectivamente por las Ecuaciones 9.3 a 9.5, donde \vec{Y} es el vector del sistema de referencia estándar de la escena en dirección vertical (suelo-techo).

$$C = S + |\vec{l}_1| \cos(\vec{l}_1, \vec{l}_3) \frac{\vec{l}_3}{|\vec{l}_3|} \quad (9.2)$$

$$\vec{n} = \frac{\vec{l}_3}{|\vec{l}_3|} \quad (9.3)$$

$$\vec{u} = \frac{-\vec{Y} - (-\vec{Y} \cdot \vec{n}) \vec{n}}{|-\vec{Y} - (-\vec{Y} \cdot \vec{n}) \vec{n}|} \quad (9.4)$$

$$\vec{v} = \vec{n} \times \vec{u} \quad (9.5)$$

$$(9.6)$$

Definimos el vector $\vec{r} = \overrightarrow{CE}$ como el vector que une el centro de la circunferencia con la posición del codo. El ángulo formado por los vectores \vec{r} y \vec{u} , denominado Φ , establece la rotación del codo con respecto a la vertical. El ángulo Φ será positivo hacia el exterior del cuerpo y negativo hacia el interior. Se define el punto B tal que $\overrightarrow{CB} = (|\vec{r}| \vec{u})$, como la posición en la que el codo tienen la menor altura posible.

9.2.2. Cálculo cinemático

El objetivo principal del cálculo cinemático es determinar los ángulos de rotación que deben tomar las articulaciones de la cadena cinemática para que su extremo, o efector final, alcance una posición determinada (P_{destino}). Es evidente que en función de las coordenadas objetivo y la complejidad del brazo articulado, el número de soluciones posibles varía.

La estructura anatómica del brazo humano que se ha definido para el avatar establece la muñeca como efector final de la cadena cinemática. Las restricciones anatómicas establecen que no existe solución si la distancia entre el objetivo y el hombro es mayor que la suma de las longitudes del brazo y del antebrazo o, por el contrario, si dicha distancia es muy pequeña, menor que aproximadamente $0,38 |\vec{l}_1|^1$. En caso de que el objetivo solicitado se encuentre en una región en la que no existe solución posible, se realiza un desplazamiento de dicho objetivo a la posición más cercana en la que existe solución.

Una vez establecida una posición objetivo en la que existe al menos una solución, se deben establecer los ángulos del codo y del hombro.

¹Calculado por trigonometría a partir del valor máximo del ángulo flexión del codo.

9.2.2.1. Cálculo del ángulo del codo

El ángulo que se debe flexionar el codo para que la muñeca (W) alcance la posición deseada ($P_{destino}$) se calcula mediante trigonometría. Definiendo el vector $\vec{v}_{destino} = \vec{SP}_{destino}$, se establece que el ángulo que debe adoptar el codo se obtiene de la Ecuación 9.7. Esta articulación puede rotar desde un ángulo de 0° grados, que corresponde a una extensión completa, hasta 145° que es el valor máximo de flexión activa de esta articulación [67].

$$\text{ángulo}_{\text{codo}} = \arccos \left(\frac{|\vec{l}_1|^2 + |\vec{l}_2|^2 - |\vec{v}_{destino}|^2}{2 |\vec{l}_1| |\vec{l}_2|} \right) \quad (9.7)$$

9.2.2.2. Cálculo de la rotación del hombro

Una vez definido el ángulo de rotación que debe tomar el codo, la rotación que debe realizar el hombro para que la muñeca alcance la posición de destino es equivalente a la rotación que debe realizar \vec{l}_3 para convertirse en $\vec{v}_{destino}$. Por lo tanto el eje de giro queda definido por la Ecuación 9.8 y el valor de dicho ángulo de giro se obtiene mediante la Ecuación 9.9.

$$\vec{e}j_{\text{hombro}} = \vec{l}_3 \times \vec{v}_{destino} \quad (9.8)$$

$$\text{ángulo}_{\text{hombro}} = \arccos \left(\frac{\vec{l}_3 \cdot \vec{v}_{destino}}{|\vec{l}_3| |\vec{v}_{destino}|} \right) \quad (9.9)$$

9.2.3. Relaciones

Como se ha explicado antes, la posición final del codo deberá determinarse en base a un conjunto de restricciones anatómicas como son los límites de rotación del hombro y de la muñeca, a los que se añade una limitación adicional: evitar la colisión del codo con el torso y la cabeza. Las restricciones asociadas a la resolución de colisiones se estudiarán con más detalle en la Sección 9.3

Una vez aplicadas esas restricciones, el rango de posiciones posibles permite establecer distintas soluciones anatómicamente válidas pero con diferentes grados de tensión para las articulaciones del hombro y de la muñeca. En este trabajo se ha considerado únicamente la tensión ejercida sobre el hombro ya que es la que produce un efecto visual más llamativo.

9.2.3.1. Hombro: rango válido y rango óptimo

Las limitaciones que se aplican a la rotación del hombro se establecen de manera que el avatar incorpore en esta articulación los desplazamientos que implican rotaciones de clavícula o desplazamientos del omóplato. Esto define un rango válido para el ángulo Φ entre -20° y 160° que equivale a los límites absolutos de rotación que puede adquirir el hombro. Dado que estos límites incluyen los valores de tensión máxima de la articulación, se han definido unos rangos óptimos aplicando el grado de tensión que representa el hombro. Hemos establecido cuatro posibles grados de tensión, cada uno definido mediante un rango de valores óptimos del ángulo Φ . El rango de cada estado (estándar, relajado, tenso y extra tenso) es descrito mediante gaussianas descritas en el conjunto de Ecuaciones 9.10 y que pueden verse en la Figura 9.2. Los valores que definen cada una de las gaussianas han sido establecidos manualmente.

$$g(x, \text{gradoTensión}) = \begin{cases} \textit{estándar} & \begin{cases} 1,5e^{-\frac{(x-20)^2}{2(10)^2}} & \text{si } x < 20^\circ \\ 1,5e^{-\frac{(x-20)^2}{2(34)^2}} & \text{si } x \geq 20^\circ \end{cases} \\ \textit{relajado} & \begin{cases} 1,5e^{-\frac{(x)^2}{2(5)^2}} & \text{si } x < 0^\circ \\ 1,5e^{-\frac{(x)^2}{2(15)^2}} & \text{si } x \geq 0^\circ \end{cases} \\ \textit{tenso} & \begin{cases} 1,5e^{-\frac{(x-90)^2}{2(10)^2}} & \text{si } x < 90^\circ \\ 1,5e^{-\frac{(x-90)^2}{2(10)^2}} & \text{si } x \geq 90^\circ \end{cases} \\ \textit{extra tenso} & \begin{cases} 1,5e^{-\frac{(x-120)^2}{2(10)^2}} & \text{si } x < 120^\circ \\ 1,5e^{-\frac{(x-120)^2}{2(10)^2}} & \text{si } x \geq 120^\circ \end{cases} \end{cases} \quad (9.10)$$

La aplicación del valor óptimo del ángulo Φ sobre la definición de la solución final depende además de otros dos valores denominados S_1 y S_2 . El primero de ellos, S_1 , está determinado por el valor del módulo del vector \vec{r} ; cuanto mayor sea dicho valor, es mayor el par que ejerce el peso del brazo sobre el hombro. Es decir, si el brazo está completamente extendido, rotar el hombro sobre el vector \vec{l}_3 no supondrá gran esfuerzo y ninguna variación visual. Por el contrario, cuando la flexión del codo es máxima, rotar sobre dicho vector implicará un esfuerzo mayor y una mayor diferencia visual (Ecuación 9.11). El segundo elemento, S_2 , que influye en la definición final de la posición del codo y de la tensión del hombro depende de la orientación del vector \vec{l}_3 ya que, cuanto más cerca se halle este vector de la vertical, menos esfuerzo conlleva para el hombro una variación en el ángulo Φ (Ecuación 9.12, Figura 9.3).

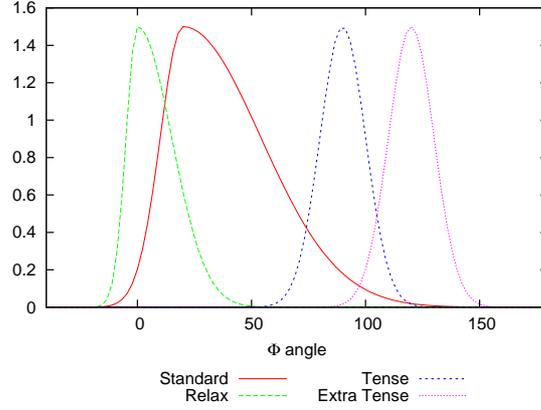


Figura 9.2: Gráfica que representa las cuatro gaussianas de definición del ángulo Φ óptimo en función del estado de tensión que se desea simular.

$$S_1 = \frac{|\vec{r}|}{|\vec{l}_1|} \quad (9.11)$$

$$S_2 = \sin(\vec{l}_3, \vec{Y}) \quad (9.12)$$

Por tanto, el valor que aporta el hombro a los solución de la posición del codo está definido por la Ecuación 9.13, en la que queda reflejado el valor óptimo definido por la función $g(\Phi)$, según el grado de tensión que se desee elegir, y los valores S_1 y S_2 .

$$S(\Phi, gradoTensión) = g(\Phi, gradoTensión) S_1 S_2 \quad (9.13)$$

9.2.3.2. Muñeca

Al igual que para el hombro, el primer paso necesario para establecer la posición que debe tomar el codo nos lleva a eliminar todas las posibles soluciones que implican una rotación imposible de la muñeca. Para ello se ha definido un conjunto de relaciones entre los vectores que definen la orientación de la misma y los que modelan el brazo. El valor W_1 (Ecuación 9.14) puede representar tanto la flexión-extensión de la muñeca como el grado de aducción-abducción. Para definir cuál de los dos movimientos se está representando, recurrimos al valor definido por W_2 (Ecuación 9.15). Si el valor de $|W_2|$ es próximo a 0, los valores de W_1 serán de aducción-abducción, mientras que si se aproximan a 1 serán de flexión-extensión. Dado que los límites máximos en las rotaciones de flexión-extensión y de aducción-abducción son diferentes, el valor W_4

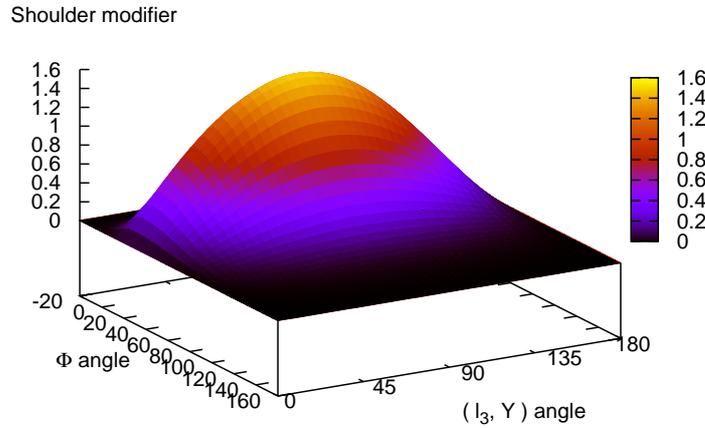


Figura 9.3: Valor óptimo del hombro en función del ángulo Φ , para una tensión estándar, y el ángulo entre los vectores \vec{l}_3 y \vec{Y} .

(Ecuación 9.17) relaciona dichos valores para establecer la tensión de la muñeca en función del tipo de rotación. El valor de W_3 (Ecuación 9.16) modela el movimiento de pronación-supinación del antebrazo. Para ello se establece la relación entre la orientación de la palma de la mano (\vec{p}) con el vector \vec{e} .

$$W_1 = \vec{l}_2 \cdot \vec{h} \quad (9.14)$$

$$W_2 = \vec{l}_2 \cdot \vec{p} \quad (9.15)$$

$$W_3 = \vec{p} \cdot \vec{e} \quad (9.16)$$

$$W_4 = W_1 - [0,6(1 - W_2)] \quad (9.17)$$

Una vez obtenidos los valores antes mencionados relativos a la muñeca, se establecen los límites para representar las restricciones anatómicas. En primer lugar, el ángulo máximo de la mano con respecto al antebrazo está limitado por el ángulo máximo de flexión-extensión de la muñeca [67], como queda expresado en la Ecuación 9.18. En segundo lugar, dado que la palma de la mano siempre debe estar orientada hacia el interior del brazo, el valor de W_3 deberá ser siempre menor o igual que 0, tal y como se indica en la Ecuación 9.19.

$$W_1 \geq 0 \quad (9.18)$$

$$W_3 \leq 0 \quad (9.19)$$

La situación de menor tensión para la articulación de la muñeca es aquella que maximiza el valor de W_1 .

9.2.3.3. Fórmula final

Una vez establecidas las restricciones que se aplican a las articulaciones del hombro y de la muñeca en función del valor de Φ y las definiciones que formalizan los valores óptimos de dichas articulaciones, el valor final del ángulo Φ será aquel que maximice la Ecuación 9.20 siempre y cuando cumpla las restricciones establecidas en esta sección.

$$K(\Phi, \text{gradoTensión}) = 0,35W_1 + 0,55S(\Phi, \text{gradoTensión}) + 0,1W_4 \quad (9.20)$$

9.3. Resolución de colisiones

Debido a la naturaleza intrínseca de los objetos representados en entornos virtuales, una definición incorrecta de sus posiciones puede dar lugar a una intersección entre las mallas de los mismos. Esto también es aplicable cuando la malla de un objeto se deforma, pudiendo atravesarse a sí misma, lo que provoca una imagen anormal y por tanto debe eliminarse la colisión para mejorar el resultado final.

9.3.1. Trabajo relacionado

El único sistema de síntesis de LS que incluye algún tipo de detección/eliminación de colisiones ha sido presentado por van Zijl y Raitt [109]. Esta aplicación plantea un algoritmo para evitar las colisiones en una mano que realiza deletreo dactilológico. La mano no está implementada como una malla deformable, sino que es un conjunto de elipsoides de dimensiones fijas. El algoritmo se basa en planificar las transiciones entre dos *configuraciones* distintas para evitar la colisión entre los distintos dedos.

El sintetizador definido por Kennaway *et al.* [63] asume que las definiciones hechas en SiGML no implican ningún tipo de colisión ni situación imposible. Esto implica

que su avatar generará colisiones debidas una definición incorrecta. Además, también puede darse el caso de que la colisión se produzca durante la transición entre dos posiciones correctas. Su proyecto sólo tiene en consideración evitar la colisión del codo con el tórax limitando los valores de rotación del hombro.

Para evitar estas situaciones, se ha estudiado el método planteado por Badler *et al.* [5]. Esta técnica se basa en el uso de potenciales para la detección de colisiones y está orientada al manejo de la transición entre posturas estáticas. Para evitar obstáculos consideran dos soluciones: la de planificar una trayectoria libre de colisiones y simular un comportamiento repulsivo en el avatar. De manera resumida, se calcula el potencial para todos los puntos susceptibles de entrar en colisión, que será mayor que uno si dicho punto se encuentra dentro de un objeto. El cálculo de potenciales queda simplificado ya que usa cilindros para la definición del avatar. Sin embargo, este algoritmo presenta un problema al poder caer en un mínimo local, solucionándolo mediante el uso de movimientos aleatorios para intentar salir del mismo. El sistema implementado en nuestro sintetizador sigue una técnica similar.

9.3.2. Volumen de colisión

Se define que un punto se encuentra dentro de un objeto si se sitúa en la parte interior de todas las caras del mismo. En los gráficos por ordenador se define la parte exterior e interior de un polígono mediante la definición del vector normal a dicho polígono. El sentido de dicho vector se establece por el orden en que se utilizan los vértices del polígono para calcular el vector normal. Se establece que un punto se encuentra en el exterior de un polígono si el producto escalar del vector que une dicha cara con el punto, por la normal al polígono es mayor o igual a cero.

La aproximación ideal para detectar la colisión es realizar la comprobación utilizando la propia malla que compone el avatar. Sin embargo, esto supone varios inconvenientes: en primer lugar, el número de polígonos que se utilizan para definir la malla es demasiado grande como para que las comprobaciones se realicen en un tiempo razonable; el segundo, es que al construir las mallas no siempre se describen volúmenes cerrados y convexos, concretamente cuando el modelado del avatar se realiza dividiéndolo en bloques principales. Por último, utilizar la malla para la detección de colisiones hace que el proceso dependa del trabajo del diseñador gráfico, no siendo dicho proceso una de las consideraciones en el momento de modelar el avatar.

La solución propuesta para tratar de paliar las limitaciones antes planteadas ha sido la definición de poliedros de colisión. Estos poliedros (Figura 9.4) no dependen de la definición de la malla, sino de la posición de los «huesos de colisión» (ver

Punto 6.1.2.4). Estos huesos independizan el volumen de colisión del detalle con el que se modele la malla del avatar. Además, se reduce el número de polígonos con los que se debe operar para buscar las colisiones.

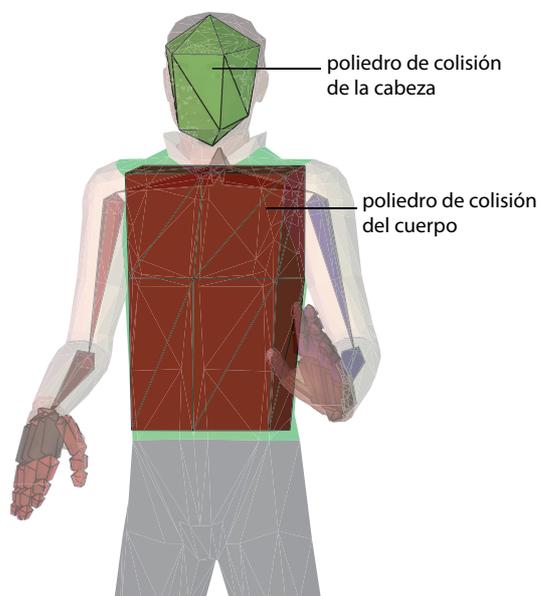


Figura 9.4: Esta figura muestra los poliedros que definen los volúmenes de colisión para la cabeza y el torso del avatar.

9.3.3. Algoritmo aplicado

El algoritmo que se ha desarrollado en este trabajo aplica la resolución de colisiones en dos momentos concretos de la síntesis:

- En el primero se plantea la generación de una trayectoria que evite colisiones entre las manos y el resto del cuerpo del avatar. Este procedimiento se aplica tras la generación del vector de posiciones de la muñeca. Su cometido es buscar y resolver, si fuera necesario, las posibles colisiones que pudieran existir en la transición entre dos posiciones consecutivas.
- En el segundo, se aplica durante la ejecución de la Cinemática Inversa con el objetivo de evitar colisiones entre el codo y el cuerpo del avatar.

Trayectoria libre de colisiones

El primer paso en todo algoritmo diseñado para evitar colisiones es la detección de éstas. En nuestro caso, se define que hay colisión si alguno de los huesos de la mano

se encuentra dentro de algún volumen de colisión. Para reducir el número de cálculos, se define un subconjunto de los huesos de la mano para realizar la comprobación.

El algoritmo planteado para modificar las trayectorias que presentan colisión tiene dos etapas distintas, diferenciándose ambas en el criterio de elección del vector que se utiliza para desplazar la mano y evitar la colisión. En la primera etapa (que coincide con la primera iteración del algoritmo) se define el vector de desplazamiento como el vector de tendencia \vec{t} tal, que se haga sortear el obstáculo por el mismo lado de la mano o bien por la parte delantera del avatar. En la segunda etapa (que corresponde a las siguientes iteraciones) el vector de desplazamiento se define como el vector de salida \vec{s} . Este vector establece su origen en el centro del volumen en colisión y final en la posición de la muñeca. Una vez definido el vector de desplazamiento, en ambos casos la mano se mueve sobre la dirección definida hasta que se elimina la colisión, incorporando esta nueva posición de la muñeca al vector de posiciones original (definido en las etapas 3 y 5 del proceso de «Generación de Pistas de Animación»).

Cálculo del vector de tendencia

En la primera iteración, el vector de tendencia \vec{t} (Figura 9.5) se calcula a partir de los vectores \vec{f} (dirección y sentido frontal del avatar) y \vec{l} (dirección y sentido del lado de la mano procesada), siendo además ortogonales. El vector \vec{t} es la suma de las proyecciones de \vec{f} y \vec{l} sobre el plano ortogonal a la trayectoria (P_i, P_{i+1}) en el punto en que se ha detectado la colisión P_c .

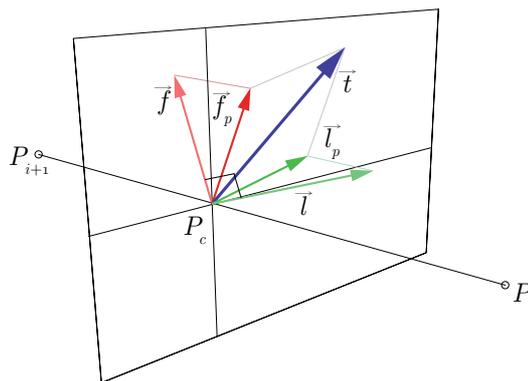


Figura 9.5: Esta figura muestra el vector de desplazamiento calculado en la primera etapa del algoritmo.

Cálculo del vector de salida

En las siguientes iteraciones, el vector de salida \vec{s} (Figura 9.6) es el vector que tiene como origen el centro del volumen C y como destino la posición de la muñeca en la que se detectó la colisión P_c .

9.3.4. Resolución de colisiones durante la cinemática inversa

Como se ha planteado en la Sección 9.2, el algoritmo de cinemática inversa aplicado al avatar debe resolver la indeterminación que implica el posicionamiento del codo: de entre todas las posiciones posibles que puede adoptar con las limitaciones anatómicas impuestas por la flexibilidad de las articulaciones, se deben evitar todas aquellas que impliquen una colisión entre el codo y el cuerpo del avatar.

El algoritmo desarrollado para esta resolución es simple: para cada posible posición del codo se descartan todas aquellas en las que se produzca una colisión con el poliedro del tórax. La versión actual del algoritmo utiliza únicamente el punto C descrito en la sección anterior (ver Figura 9.1) para la comprobación, lo que puede dar lugar a una ligera colisión entre el codo y el cuerpo. Si dicha colisión se percibiera por los usuarios como perjudicial para la apariencia del avatar, deberían utilizarse los «huesos de localización» relativos al codo y al antebrazo para el proceso.

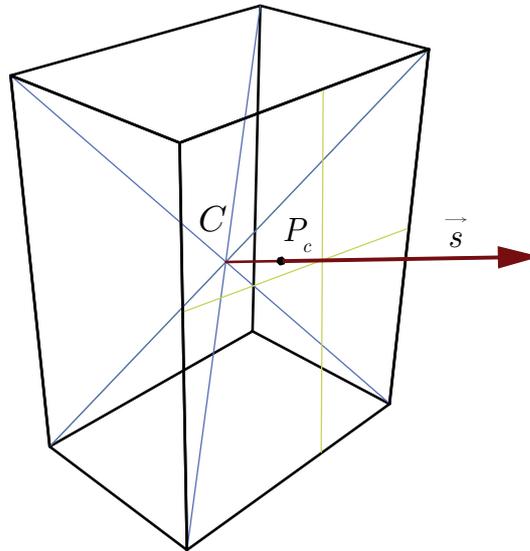


Figura 9.6: Esta figura muestra el vector de salida aplicado en caso de detectar una colisión en alguna de las iteraciones $2, n$ del algoritmo.

9.4. Resumen

En este apartado hemos presentado la dependencia entre los distintos PFQ y la secuencia de acciones definida para realizar la síntesis de LSE. Así mismo hemos presentado el algoritmo de Cinemática Inversa que hemos utilizado en este trabajo y cómo se aprovechan las distintas soluciones posibles para simular distintos grados de tensión en el hombro. Por último hemos descrito la aproximación seguida para evitar la colisión entre las manos del avatar y el resto del cuerpo.

Capítulo III

Experimentación y Evaluación

Apartado 10

Evaluaciones objetivas

El sintetizador de LSE que aquí se propone se basa en una aproximación novedosa en comparación a otros sintetizadores existentes. El desarrollo del mismo se ha realizado en su totalidad, lo que incluye el modelado del avatar, el diseño de la base de datos relacional, la implementación del módulo de síntesis y las directivas aplicadas al módulo de «Render». Esto implica que, antes de realizar cualquier prueba con usuarios, se debe validar el rendimiento del sistema de manera objetiva.

En este apartado se presentan dos evaluaciones: en la primera se mide la capacidad del sistema implementado para animar el avatar en tiempo real, mientras que la segunda determina las secciones del módulo de síntesis que requieren mayor tiempo de proceso. Estos resultados permitirán, además, definir los objetivos de futuras mejoras.

10.1. Evaluación de la animación en tiempo real

El primer experimento realizado tiene como objetivo validar la capacidad del sistema para generar animaciones fluidas en tiempo real. En esta sección presentamos dos experimentos que calculan la tasa de refresco en función de dos condicionantes: Por una parte, como se precisó en la Sección 6.2 el tiempo requerido por el proceso de «Render» depende en gran medida del número de polígonos que contiene la escena; en nuestro caso, todos los polígonos pertenecen a la malla del avatar. En segundo lugar, el tiempo requerido en este proceso depende también de la resolución de la ventana de visualización.

10.1.1. Tasa de refresco según el número del polígonos de la malla

Este experimento se realizó en paralelo con el proceso de diseño del avatar *Yuli* (ver Sección 6.4). El objetivo planteado ha sido encontrar un equilibrio entre el número de polígonos de la malla que permite maximizar el detalle de la misma y la minimización del tiempo consumido al procesar cada imagen. Para ello se obtuvieron medidas de la tasa de refresco con distintas versiones del avatar, manteniendo las dimensiones de visualización en una resolución de 400×400 píxels. Las características técnicas del ordenador que se utilizó para este experimento se describen en la Tabla 10.1.

Ordenador	Procesador	Frecuencia	RAM	Mem. de Vídeo	Java
A	Pentium IV	2 GHz	512 MB	32 MB	1.5.4

Tabla 10.1: Descripción del ordenador utilizado para la evaluación 10.1.1.

En la primera versión, el avatar se creó con una malla de aproximadamente 10000 polígonos, obteniéndose una tasa de refresco media de 10 imágenes por segundo (fps). Sucesivas reducciones en el número de polígonos del avatar permitieron alcanzar una media de 14 fps al aproximarse a los 7800 polígonos. Para que la reducción en el número de polígonos no mermara drásticamente el nivel de detalle del avatar, se optó por definir grupos de suavizado¹ a lo largo de las distintas partes del mismo. El resultado permitió mejorar tanto la apariencia de la malla como la tasa de refresco, pasando de 14 a 20 fps, valor que puede considerarse dentro de los límites de una animación fluida. Los resultados de este experimento se muestra en la Figura 10.1.

10.1.2. Tasa de refresco según el tamaño de la ventana

El segundo experimento para la evaluación del rendimiento del sistema se centró en medir la tasa de refresco en función del tamaño de la ventana utilizada. Para ello se utilizó el avatar *Daniel*, realizando la misma animación en todos los casos. Las mediciones fueron realizadas en dos ordenadores: uno de gama media (ordenador A) y otro de gama alta (ordenador B), cuyas características técnicas se presentan en la Tabla 10.2.

Ordenador	Procesador	Frecuencia	RAM	Mem. de Vídeo	Java
A	Pentium IV	2 GHz	512 MB	32 MB	1.5.4
B	Core 2 Duo	2,4 GHz	3,12 GB	512 MB	1.6.0

Tabla 10.2: Descripción de los ordenadores utilizados para la evaluación 10.1.2.

¹Un grupo de suavizado es un conjunto de polígonos que componen una misma superficie y se gestionan conjuntamente al realizar los cálculos de iluminación, con el objetivo de dar una sensación de continuidad en la representación.

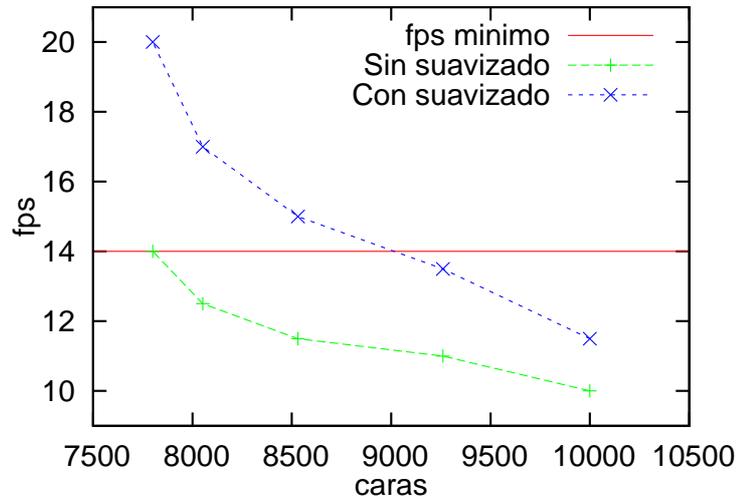


Figura 10.1: Evaluación de la tasa de refresco (fps) generada por el ordenador en función del número de caras del avatar y el uso de grupos de suavizado.

El avatar *Daniel* fue diseñado utilizando grupos de suavizado, ya que, según quedó demostrado en el experimento 10.1.1 su uso mejora el rendimiento del sistema. La tasa de refresco se ha medido utilizando una ventana cuadrada, cuyo lado se varió desde 100 a 650 puntos, en incrementos de 50 puntos.

Los resultados obtenidos permiten afirmar que el avatar puede animarse con una tasa de refresco válida cuando la resolución es inferior a 400×400 píxeles, lo que permite obtener un grado de detalle aceptable. El hecho de que la resolución de la pantalla se defina en tiempo de ejecución, permite su adaptación a la capacidad gráfica del dispositivo del cliente. La gráfica que se presenta en la Figura 10.2 muestra los resultados de este experimento.

Por último, se ha realizado también este experimento en la implementación desarrollada para el dispositivo PocketPC. El dispositivo utilizado es un HTC P3600, con procesador Samsung SC32442A 400 MHz, 64 MB de memoria RAM, sin memoria ni procesador gráfico dedicados. En este dispositivo, la animación se ejecuta a 2 fps en una ventana de 240×240 píxeles, lo que constituye un valor demasiado bajo como para obtener una animación fluida. Se plantea, para un desarrollo futuro, el uso de un avatar con menor detalle y/o dispositivos con mayor capacidad gráfica.

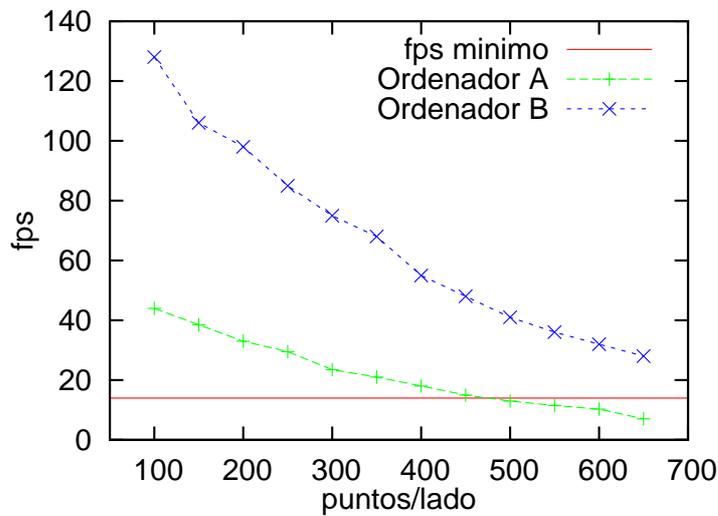


Figura 10.2: Evaluación de la tasa de refresco en función del tamaño de la ventana.

10.2. Tiempos de ejecución de cada sección del proceso de síntesis

La segunda evaluación de rendimiento realizada sobre el sistema consiste en medir el tiempo de ejecución de las distintas partes del proceso de síntesis:

1. Parser de HLSML.
2. Recuperación desde la base de datos relacional.
3. Procesado de signos y generación de pistas de animación
4. Cinemática Inversa (IK).
5. Detección de Colisiones.

Para realizar las mediciones se ha utilizado el sistema instalado en un ordenador de las mismas características que el descrito en la Tabla 10.2 como de gama alta (ordenador B). Se ha registrado el rendimiento de distintos tipos de signos: monomanuales y bimanuales, con variantes estáticas y dinámicas. Además se han evaluado dos tipos de clasificadores icónicos bimanuales: uno estático (sencillo) y otro dinámico (complejo). Para los signos se compuso una sentencia en HLSML con diez repeticiones del mismo signo; cada frase se ejecutó cuatro veces, lo que da un total de cuarenta mediciones por cada clase de signo. En el caso de los clasificadores se compuso una sentencia HLSML para cada clasificador en estudio, ejecutándose diez veces. El

clasificador «Sencillo» consta de una descripción estática de una postura del avatar, mientras que el «Complejo» contiene una descripción de una animación del avatar incluyendo la animación de los ojos y la cabeza. La media de los valores obtenidos se reflejan en la Tabla 10.3.

	Signo				Clasificador		
	Monomanual		Bimanual		Sencillo	Complejo	
	Estático	Dinámico	Estático	Dinámico			
Total	240,6	1400	904,3	5026	297,2	4613	
Parte	1	14				15,5	101,5
	2	65,1	74,2	97,8	101,3	132,7	56
	3	11,7	12,4	126,9	114,7	11,8	11,9
	4	161,3	718	673,2	2674	238,5	2312
	5	0,1	589,4	0,8	2130	0	2058
Ejemplo	YO / TÚ	PASADO	TELEVISIÓN	FELIZ			

Tabla 10.3: Tiempo medio de ejecución en ms para distintos tipos de signos y clasificadores en las distintas partes del proceso de síntesis.

A la vista de los resultados obtenidos en esta evaluación, podemos concluir que las partes del proceso que más penalizan el tiempo final son la IK (4) y la detección de colisiones (5). Estos dos procesos deberán mejorarse en futuros desarrollos para reducir el tiempo de cálculo. De igual modo, podría valorarse la efectividad de una simplificación de ambos en una implementación orientada a dispositivos móviles.

Apartado 11

Evaluaciones subjetivas

La evaluación de la calidad de la síntesis, la naturalidad de los movimientos del avatar o su expresividad es un proceso que sólo puede ser llevado a cabo por usuarios, lo que lo convierte en un proceso subjetivo. En este apartado se presentan una serie de evaluaciones destinadas a medir diferentes facetas del avatar.

Las evaluaciones realizadas se han agrupado en cuatro secciones atendiendo a aspectos concretos de la síntesis y del avatar. La primera de ellas mide la capacidad expresiva y comunicativa de la cara del avatar. En la segunda se evalúa la viabilidad de la aproximación utilizada para que el avatar muestre distintos grados de tensión, base para la síntesis emocional. Las secciones tercera y cuarta se centran en las evaluaciones realizadas sobre la síntesis de LSE, presentando los resultados obtenidos en distintas pruebas de comprensión de los mensajes generados por el avatar.

11.1. Capacidad expresiva y comunicativa de la cara del avatar

En la LS, la cara del signante tiene dos labores fundamentales. La principal es utilizar las expresiones faciales como parte del PFQ *no manual*, necesario para la definición de los signos *establecidos*. De manera secundaria, el movimiento de los labios complementa la comunicación entre aquellas personas que dominan la lectura de los mismos.

Se han planteado tres evaluaciones distintas en relación a la cara del avatar. Las dos primeras se centran en el movimiento de los labios y la tercera en la capacidad expresiva del avatar.

11.1.1. Comprensión de la lectura de labios

La lectura de labios se utiliza en ocasiones como complemento de la LS, por lo que un avatar signante con movimiento labial aumenta su capacidad de comprensión. En nuestro sistema, dentro de la notación HLSML, se describe el movimiento labial como una secuencia de visemas¹ con una marca temporal asociada. Para la generación de este mensaje se han utilizado dos reconocedores de voz, uno entrenado con modelos de alófonos y otro con modelos basados en grafemas. El uso de estas dos alternativas en el español se justifica por el trabajo de Tejedor *et al.* [105]. En ambos casos, el resultado se convierte en secuencias de visemas mediante una tabla de equivalencias (ver Tabla 3.2). A continuación, el avatar genera los movimientos correspondientes a dicha secuencia de visemas.

La evaluación de comprensión en la lectura de labios se ha realizado mediante un experimento en el que el avatar reproduce diez frases con movimiento labial. Estas frases fueron extraídas de la base de datos Albayzin [79] y procesadas con ambos reconocedores generando un total de veinte casos de prueba. Para cada una de ellos se muestra un vídeo en primer plano de la cara del avatar y un conjunto de cinco respuestas posibles de las que sólo una es verdadera. Dado que los usuarios potenciales son sordos, los vídeos utilizados no presentan ningún sonido. En este experimento se contó con la colaboración de once personas, no expertos en la lectura de labios, obteniéndose los resultados de la Tabla 11.1; para cada uno de los dos reconocedores se muestra la tasa aciertos en el reconocimiento y en la conversión a visemas, así como el porcentaje de aciertos de los usuarios al escoger la frase correcta.

Aciertos			
	Reconocimiento	Visemas	Usuarios
Alófonos	72,5 %	83,9 %	57 %
Grafema	82,5 %	82,3 %	64 %

Tabla 11.1: Tasa de acierto de los reconocedores de voz, conversión a visemas y reconocimiento por los usuarios.

Los resultados obtenidos conceden un margen a la mejora en la animación generada para el avatar, tanto en el modelado de los distintos visemas como en la realización de la articulación entre unidades consecutivas. Una vez conseguida la mejora en los resultados de este tipo de experimento, se deberá incrementar la dificultad proponiendo preguntas de respuesta libre.

¹Un visema es el patrón de referencia visual de un fonema.

11.1.2. Naturalidad en el movimiento de labios

La segunda evaluación, relativa al movimiento de los labios del avatar se centró en medir la naturalidad de estos movimientos. Para ello se realizaron dos experimentos, uno para medir la calidad de los movimientos utilizando una aproximación similar a la de otros proyectos de *talking head*, es decir como complemento de una señal acústica y otro en el que se informa al usuario de la frase que el avatar esta reproduciendo, pero sin incorporar audio en la frase.

El primer experimento consiste en la visualización de vídeos en los que el avatar mueve los labios acorde con la frase que se está escuchando. Estos vídeos alternan dos encuadres distintos del avatar, usando un primer plano de la cara del avatar, similar al de la evaluación 11.1.1², o bien un encuadre general como en el caso de la visualización de mensajes signados. Se le solicita a cada usuario una valoración entre uno y cinco de la calidad con la que percibe el movimiento labial. Los resultados obtenidos para un conjunto de treinta personas se recogen en la Tabla 11.2 donde se compara, para cada frase, la evaluación en los distintos encuadres. La apreciación de la calidad es mejor en un encuadre cercano, significante al 2% con un test de significancia de «T de Student», como cabría esperar, ya que la boca se observa con mayor detalle.

	Frase							
	1	2	3	4	5	6	7	8
cerca	3,5	4,0	3,8	3,7	3,8	3,8	3,7	3,3
lejos	3,4	3,5	3,7	3,4	3,7	3,5	3,5	3,2
	Frase							Med
	9	10	11	12	13	14	15	
cerca	3,4	3,6	3,4	3,8	3,6	3,6	3,9	3,7
lejos	3,3	3,4	3,7	3,8	3,5	3,6	3,6	3,5

Tabla 11.2: Resultados de la calidad, en una escala de 1 a 5, del movimiento labial *con* sonido.

El segundo experimento fue realizado por los mismos usuarios que participaron en la evaluación 11.1.1, no expertos en la lectura de labios. Este experimento utiliza también los mismos encuadres que en dicha evaluación, pero presentado los vídeos de la cara del avatar sin sonido, informando de la frase de manera escrita. Los resultados obtenidos en valoraciones de uno a cinco se presentan en la Tabla 11.3, mostrando una disminución en la percepción media de la calidad de un 20% relativo con respecto al experimento anterior.

²Este encuadre es el utilizado por las aplicaciones de *talking head*.

	Calidad
alófono	2,91
grafema	2,84

Tabla 11.3: Resultados de calidad, en una escala de 1 a 5, para el movimiento labial *sin* sonido.

Podemos concluir que la incorporación del movimiento labial al avatar otorga una calidad aceptable cuando el avatar se utiliza como complemento a una señal acústica. Sin embargo, cuando se presenta únicamente el movimiento labial, no se percibe un nivel de calidad suficiente para permitir una lectura de labios precisa.

11.1.3. Expresiones faciales

La última de las evaluaciones relacionadas con la cara del avatar se centra en el estudio de la capacidad expresiva. Este factor tiene aplicación tanto para la síntesis de signos (definición del PFQ *no manual*) como en la mejora de la naturalidad del avatar.

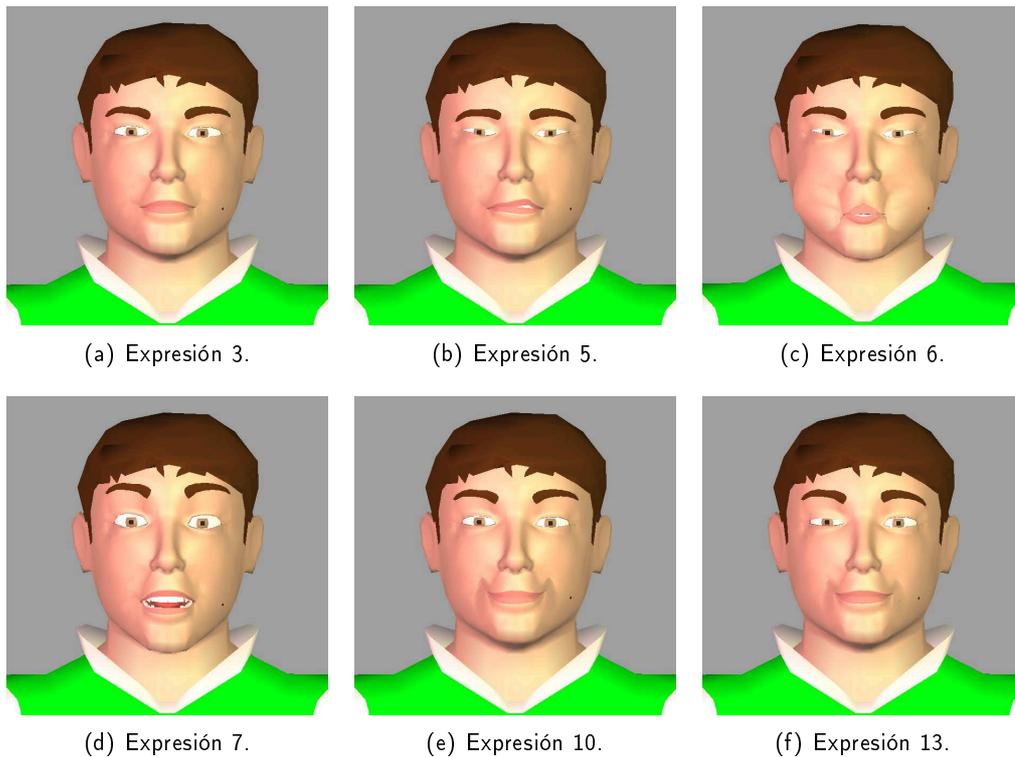


Figura 11.1: Algunas de las expresiones utilizadas en la evaluación 11.1.3.

En este test se presenta al usuario catorce animaciones de la cara del avatar (ver Figura 11.1). Se comienza con una expresión neutra, pasando a representar la expresión deseada, que mantiene durante 600ms, y retorna a la expresión neutra. A cada expresión se acompaña con un conjunto de diez posibles soluciones de las que el usuario debe escoger sólo una. Los resultados obtenidos con cuarenta y cinco usuarios encuestados, se recogen en la Tabla 11.4, en la que se ha resaltado en azul la expresión que se trata de simular. Se han agrupado las expresiones correspondientes a estados emocionales cercanos, dado que son las que han generado mayor confusión entre los usuarios.

Como conclusión podemos afirmar que la capacidad que tiene el avatar para generar expresiones es notable. No obstante, en determinadas expresiones, se debe mejorar el modelado, ya que en algunos casos se produce ambigüedad en la expresión evaluada (2), en otros se ha confundido con una expresión cercana ($1 \leftrightarrow 8$) o se ha interpretado erróneamente (4).

11.2. Representación de la tensión

En la Sección 4.1, se han referido distintos PFQ susceptibles de modificación para representar distintos estados emocionales. Además, en la Sección 9.2 se han planteado soluciones diferentes para el proceso de Cinemática Inversa (IK) para representar distintos grados de tensión en el hombro. En la sección anterior se verificó la capacidad del avatar para mostrar diferentes expresiones faciales, componente principal del PFQ *no manual*. En esta sección se presentan los resultados obtenidos en la evaluación de las variaciones incorporadas al PFQ *configuración* y la aproximación propuesta para la IK.

11.2.1. Tensión en la *configuración*

En la Sección 4.1 se ha descrito cómo la alteración en el estado emocional y el grado de tensión en la síntesis de signos se reflejan en la tensión muscular con la que se realizan las *configuraciones*. Para validar la representación de distintas variantes para una misma *configuración* se ha planteado el siguiente experimento:

El usuario observa tres imágenes correspondientes a distintas representaciones de la misma *configuración*. La diferencia entre ellas se corresponde al grado de tensión diferente (tensa, normal y relajada) reflejado en la flexión de las articulaciones de la

	tristeza	miedo	sorpresa	ironía	placer	felicidad	asco	enfado	neutro	soplo
1	6,7%		2,2%				26,7%	60%	4,4%	
2	4,4%			37,8%	4,4%		13,3%	0%	4,4%	35,6%
3				2,2%					97,8%	
4		2,2%	4,4%	66,7%	22,2%	4,4%				
5		4,4%					17,8%	77,8%		
6			2,2%	2,2%			2,2%		2,2%	91,1%
7		15,6%	80%	2,2%					2,2%	
8	44,4%	55,6%								
9	20%	66,7%	11,1%						2,2%	
10			6,7%	17,8%	17,8%	51,1%	2,2%	2,2%	2,2%	
11		2,2%		13,3%			55,6%	26,7%		2,2%
12	2,2%			20,0%	24,4	46,7%	4,4%		2,2%	
13	2,2%	2,2%	8,9%			4,4%	2,2%		6,7%	
14	66,7%	24,4%			2,2%		2,2%		4,4%	

Tabla 11.4: Porcentaje de respuestas para la evaluación de expresiones faciales. Indicamos en azul la expresión que se está representando.

mano. En este experimento no se ha realizado modificación alguna sobre la apariencia de la textura de las manos. La Figura 11.2 muestra una de las ternas utilizadas.

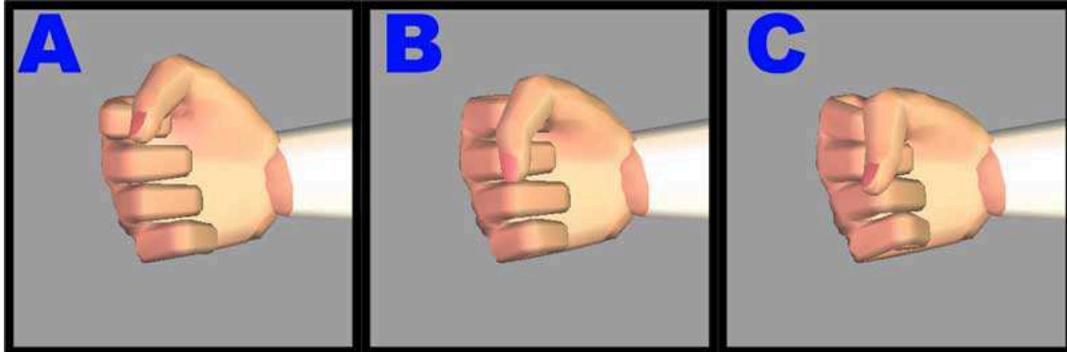


Figura 11.2: Imagen utilizada en la pregunta 6 de la evaluación 11.2.1.

Los resultados de este experimento (ver Tabla 11.5) confirman que el avatar es capaz de representar variaciones perceptibles en la tensión de las manos. En todas las preguntas de los veinte usuarios encuestados, más de un noventa por ciento siempre señalaron correctamente la mano con mayor tensión. En dos ocasiones, las variaciones normal y relajada se permutaron.

	A>B>C	A>C>B	B>A>C	B>C>A	C>A>B	C>B>A
1	90%	10%				
2			30%	60%	5%	5%
3			5%		80%	15%
4	65%	30%		5%		
5		10%	15%	85%		
6			5%			95%
7		5%			95%	

Tabla 11.5: Resultados de la evaluación de tensión en la mano.

11.2.2. Tensión en el hombro

La segunda evaluación relacionada con la representación de la tensión se centra en el proceso de IK. La posición del codo se define para mostrar mayor o menor tensión en el hombro.

El experimento propuesto presenta al avatar realizando el mismo movimiento, con los dos brazos, definido con la misma amplitud, velocidad y aceleración. En cada cuestión, los hombros pueden representar el mismo o distinto grado de tensión en la ejecución de los movimientos, los cuales se realizan de manera sincronizada sobre dos objetos, uno blanco y otro negro. Se insta al usuario a identificar si existe alguna

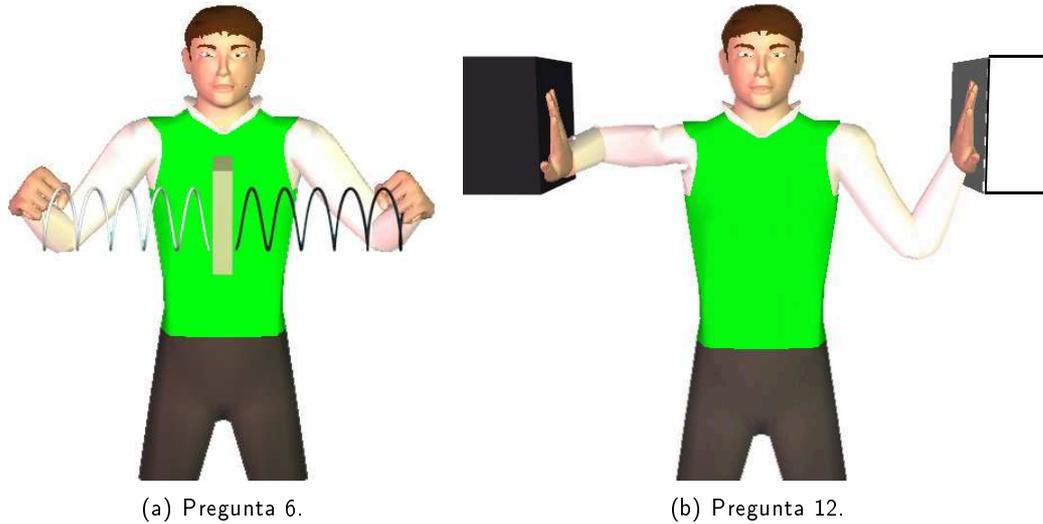


Figura 11.3: Fotogramas de los vídeos utilizados en la evaluación 11.2.2.

diferencia en la percepción del esfuerzo realizado por ambos brazos y en caso afirmativo, cuál realiza mayor esfuerzo. La Figura 11.3 muestra dos ejemplos de los vídeos presentados. Para limitar la evaluación a la percepción de la postura de los brazos, no se presenta ningún tipo de expresión facial que pueda influir en la decisión final.

Los resultados obtenidos al encuestar a veintiún usuarios (ver Tabla 11.6) muestran que, en general, se señala la opción correcta, aunque sin una tendencia definida. Debido a que las personas interpretan principalmente el esfuerzo por la expresión facial y la velocidad de movimientos. Por tanto, se puede concluir que la posición del codo refleja la tensión del hombro, pero no puede ser utilizada de manera aislada para representar una variación en la tensión del avatar. Nuestra hipótesis es que si este experimento se repitiese solicitando la comparación entre dos vídeos del avatar con las distintas alternativas, los resultados mejorarían significativamente.

11.3. Evaluación signos aislados

La primera evaluación realizada para medir la inteligibilidad del sintetizador utiliza exclusivamente signos aislados. En este tipo de evaluaciones, se presenta un vídeo del avatar en el que parte de una posición de reposo, ejecuta el signo determinado y retorna a la posición de reposo. Al presentar sólo un signo existe la certeza de que se evitan confusiones relativas al comienzo y final del signo, se evita que un

	blanco	negro	igual
1	23,8 %	23,8 %	52,4 %
2	14,3 %	14,3 %	71,4 %
3	38,1 %	47,6 %	14,3 %
4	47,6 %	19,1 %	33,3 %
5	57,1 %	23,8 %	19,1 %
6		19,1 %	80,9 %
7	28,6 %		71,4 %
8	33,3 %	61,9 %	4,8 %
9	57,1 %	42,9 %	
10	52,4 %	42,9 %	4,7 %
11	33,3 %	61,9 %	4,8 %
12	23,8 %	76,2 %	

Tabla 11.6: Resultado de la evaluación de tensión en el brazo, en porcentaje de respuestas por opción. La opción marcada como azul es la considerada correcta.

uso incorrecto de la gramática penalice el resultado así como que signos incorrectos alteren la apreciación de los correctos. Como se elimina la información de contexto hay que tener en cuenta que en los resultados obtenidos en la evaluación se debe compensar la posible ambigüedad de ciertos signos.

11.3.1. Respuesta abierta

El primer experimento de la evaluación de signos aislados se propuso dando libertad a los usuarios para exponer sus respuestas. Por lo tanto, es una evaluación relativamente cercana a una situación real, aunque no presenta la complejidad de una frase completa.

11.3.1.1. Resultados del avatar *Yuli*

El primer experimento realizado planteó un conjunto de veinte signos distintos para los que cada usuario debía identificar el signo. Para este experimento se contó con la colaboración de seis profesores de la academia SIGNAR, tres de ellos intérpretes de signos y el resto nativos de la LSE. Este experimento presentó al avatar *Yuli* ejecutando una secuencia de signos entre los que se realizaban pausas para diferenciar correctamente el inicio y el final del signo. En este experimento se observó que los usuarios sólo respondían cuando tenían la certeza del signo que estaban visualizando. Los resultados obtenidos en este experimento se muestran en la Tabla 11.7.

Signo		Usuario						Aciertos		
		1	2	3	4	5	6	1-3	4-6	total
1	HOLA	✓	✓	✓	✓	✓	✓	3	3	6
2	YO	✓	✓	✓	✓	✓	✓	3	3	6
3	SORDO	✓			✓			1	1	2
4	CRECER		✓		✓	✓	✓	1	3	4
5	AQUÍ	✓	✓	✓	✓	✓	✓	3	3	6
6	MADRID	✓	✓	✓	✓	✓	✓	3	3	6
7	EDAD							0	0	0
8	25	✓	✓	✓	✓	✓	✓	3	3	6
9	FELIZ	✓	✓	✓	✓	✓	✓	3	3	6
10	COMPAÑERO			✓	✓			1	1	2
11	COLEGIO	✓				✓	✓	1	2	3
12	TOLEDO	✓						1	0	1
13	HORA		✓	✓	✓	✓	✓	2	3	5
14	CERCA			✓	✓			1	1	2
15	MÍO		✓		✓	✓		1	2	3
16	VERDE	✓	✓	✓	✓	✓	✓	3	3	6
17	ROJO	✓		✓	✓	✓	✓	2	3	5
18	MAÑANA		✓		✓	✓	✓	1	3	4
19	TODO EL DÍA				✓	✓	✓	0	3	3
20	HOY	✓		✓	✓	✓	✓	2	3	5
Aciertos		60%	55%	60%	85%	75%	70%	67,5%		
Aciertos por grupo		58,3%			75%					

Tabla 11.7: Esta tabla muestra los resultados obtenidos en el test del avatar *Yuli*. Los usuarios 1 a 3 son profesores nativos, mientras que los usuarios 4 a 6 son profesores de LSE oyentes.

Los resultados de este experimento permiten confirmar que las evaluaciones realizadas por personas sordas nativas obtienen una tasa de aciertos inferior a las realizadas por personas oyentes. Esto es debido a que las personas que utilizan la LSE como segundo idioma admiten mejor los posibles errores en la síntesis. Por esta razón, todos los experimentos restantes se realizaron con personas sordas, aunque se contó con la colaboración de expertos en LSE oyentes, en el proceso de desarrollo del sintetizador.

11.3.1.2. Resultados del avatar *Daniel*

Tras comprobar que la calidad demostrada por el avatar *Yuli* no era lo suficientemente alta, se optó por diseñar un nuevo avatar que mejorara los resultados obtenidos por *Yuli*. Una vez terminado el nuevo avatar *Daniel*, se repitió el experimento 11.3.1.1. En esta ocasión los signos se presentaron individualmente y se permitió a los usuarios que escribieran libremente el signo que se estaba visualizando. Esta libertad generó una serie de respuestas que no corresponden exactamente con los signos propuestos, pero que sí tienen una relación semántica con el originalmente propuesto, o se trata de un signo similar. Por esta razón, en los resultados de este experimento se contemplan las respuestas alternativas que se han considerado correctas, midiendo la tasa de aciertos exactos y de aciertos relajada, incluyendo éstas alternativas.

Tanto para este experimento, como para todas las evaluaciones siguientes, se ha contado con un grupo de expertos en LSE de la FCNSE y algunas personas sordas que han participado de manera voluntaria en estos experimentos. Este grupo de usuarios utilizan la LSE como medio de comunicación en su vida diaria. Este conjunto, de entre seis y once usuarios (según el experimento) pertenece al colectivo que, llegado el momento, pueden beneficiarse del uso de un sintetizador de LSE.

	Signo	Alternativas	Acierto exacto	Acierto relajado
1	FELIZ		100 %	100 %
2	BEBER/AGUA		100 %	100 %
3	HERMANO/IGUAL		100 %	100 %
4	PERO		71,4 %	71,4 %
5	HORA	RELOJ	57,1 %	100 %
6	VENIR		100 %	100 %
7	COCHE		100 %	100 %
8	LIBRO		100 %	100 %
9	PASEAR		100 %	100 %
10	LEER		100 %	100 %
11	ABRIGO		85,7 %	85,7 %
12	MÉDICO/ENFERMO		100 %	100 %
13	YO		100 %	100 %
14	MAÑANA	DÍA	57,1 %	100 %
15	CASA		100 %	100 %
16	SORDO		100 %	100 %
17	ABANICO		71,4 %	71,4 %
18	CUADRO	TELEVISIÓN	71,4 %	100 %
19	SEMANA		100 %	100 %
20	PASADO		100 %	100 %
Promedio			90,7 %	96,4 %

Tabla 11.8: Esta tabla muestra los resultados obtenidos en el test de respuesta abierta del avatar *Daniel*. Todos los valores se expresan en porcentajes.

Los resultados obtenidos en este experimento con siete usuarios expertos en LSE (Tabla 11.8), mejoran sustancialmente los obtenidos con el avatar *Yuli*. En el caso de las preguntas 2, 3 y 12, inicialmente se consideraba uno de los dos significados; pero se ha verificado que ambos se representan con el mismo signo en la LSE y por lo tanto no se considera como alternativa.

11.3.2. Respuesta cerrada

Para limitar la variabilidad de las respuestas realizadas por los usuarios, hemos realizado un experimento equivalente a los dos anteriores, pero utilizando respuestas cerradas. Cada cuestión incluía cinco posibles respuestas, de las que sólo una se ha

dado por válida³. Este experimento consta de veinte signos, presentados en un vídeo equivalente a los utilizados en el experimento 11.3.1.2 y ha sido realizado por el mismo perfil de personas que en dicho experimento. En este caso, el número de personas que realizaron este experimento es de diez.

El aumento en el porcentaje de acierto (ver Tabla 11.9), con respecto a la evaluación de respuesta abierta, se debe a la pequeña ayuda que ofrece el tener todas las opciones posibles. En próximas evaluaciones se debe incluir una opción extra que permita presentar el desacuerdo con todas las opciones planteadas⁴.

	Signo	Acierto total
1	NUEVE	100 %
2	METRO	100 %
3	HOY	100 %
4	CRECER	100 %
5	ÉL	100 %
6	LÍNEA	100 %
7	GRANDE	100 %
8	INTERESANTE	100 %
9	ARREGLAR	100 %
10	GORDO	100 %
11	AGUA	100 %
12	TÚ	100 %
13	CUATRO	100 %
14	AQUÍ	71,4 %
15	HERMANO	100 %
16	ATARDECER	100 %
17	TODO EL DÍA	100 %
18	CORRER	100 %
19	VER	100 %
20	MÍO	100 %
	Promedio	98,6 %

Tabla 11.9: Esta tabla muestra los porcentajes de aciertos obtenidos en el test de respuesta cerrada del avatar *Daniel*.

11.4. Evaluacion frases

Tras medir la tasa de reconocimiento de los signos generados por el sintetizador de manera individual, esta evaluación se centra en medir la tasa de reconocimiento de frases sintéticas en LSE. Las siguientes secciones recogen tres experimentos diferentes relacionados con esta evaluación: la primera de ellas plantea una serie de frases compuestas exclusivamente por signos *establecidos*; la segunda contiene frases

³En el caso de la pregunta 11, correspondiente al signo AGUA, se utilizó como distractor el valor BEBER. Dado que en LSE ambos signos son iguales, se ha contabilizado la única respuesta a este segundo signo como correcta.

⁴Esta modificación ha sido sugerida también en el informe de la FCNSE, ver Sección 11.5.

compuestas principalmente por construcciones clasificatorias y la tercera presenta dos frases en distintas variaciones emocionales.

El conjunto de usuarios que han realizado estos experimentos es el mismo que en el caso de los experimentos de signos aislados del avatar *Daniel*: expertos en LSE y personas que utilizan esta lengua en su vida diaria.

11.4.1. Frases compuestas por secuencias de signos *establecidos*

El primero de los experimentos tiene por objeto valorar la capacidad que tiene el avatar para generar frases mediante la concatenación de signos. En este experimento se plantea un conjunto de diez secuencias de signos *establecidos*, palabras deletreadas y números cardinales. Para cada una de las preguntas se presentan cinco opciones, consistentes en frases en español⁵. Se ha procurado que la dificultad de las oraciones planteadas sea incremental para permitir al usuario cierta adaptación al avatar. Estas oraciones han sido obtenidas en su mayoría del trabajo doctoral de Muñoz [80] para minimizar la existencia de errores gramaticales.

La Tabla 11.10 contiene los resultados obtenidos para cada una de las frases incluidas a partir de las respuestas de diez expertos en LSE. Al igual que en otros experimentos de respuesta cerrada, próximas evaluaciones deberán permitir una posible respuesta para manifestar disconformidad con las opciones planteadas. Aunque la tasa de aciertos es elevada, se deberá aumentar la dificultad de la evaluación incrementando la complejidad de las frases propuestas y permitiendo respuestas libres.

	Frase	Aciertos
1	mi nombre es F-E-R-N-A-N-D-O	100 %
2	la semana pasada vi el cuadro	90
3	tengo 25 años	100 %
4	llevo un abrigo pero tengo frío	90 %
5	el libro que leí es interesante	90 %
6	tu coche corre tanto como el mío	100 %
7	tú bebes zumo de naranja	100 %
8	él es médico	90 %
9	yo soy feliz por pasear contigo	100 %
10	la persona que está allí sentada es mi hermano	90 %
	Promedio	95

Tabla 11.10: Esta tabla contiene los resultados del test de frases en LSE, expresados en porcentajes de acierto.

⁵Como se verá en el informe de la FCNSE, el hecho de que se utilice español en la evaluación ha sido muy criticado. Las próximas evaluaciones deberán omitir el uso de español y sustituirlo por imágenes o secuencias de LSE glosado.

11.4.2. Clasificadores

Este experimento se ha planteado para evaluar la capacidad del sintetizador para generar las animaciones requeridas para la representación de los *clasificadores*. Este experimento se centra en la evaluación de los *clasificadores icónicos* y *afijales*. Dado que estas unidades representan de manera abstracta escenas del mundo real, se ha optado por definir las posibles respuestas mediante imágenes en vez de descripciones textuales⁶. En cada una de las preguntas, el avatar representaba una de las cinco escenas visibles en imágenes, de las que sólo una era correcta. Los resultados obtenidos de once expertos en LSE se resumen en la Tabla 11.11, la imagen representa la escena correcta.

Las escenas que mayor dificultad han supuesto para los usuarios han sido las relacionadas con los coches. Al parecer, cuando aparece un coche en una construcción clasificatoria, la punta de los dedos del signante representa los faros delanteros del coche. Este detalle no ha sido tenido en cuenta cuando se han generado las animaciones. La pregunta 14 presenta la tasa más baja de todo el experimento. En esta pregunta se presentan tres coches aparcados en paralelo, el problema que han encontrado los usuarios se justifica en los distintos puntos de vista representados. Aunque tanto la imagen como la animación del avatar representan tres coches en paralelo, la representación del avatar los coloca como si el signante estuviera viéndolos desde atrás, mientras que la imagen los presenta desde un lado. Esto demuestra que cuando se afronte la generación automática de clasificadores, no sólo hay que obtener la descripción de la escena, también se debe situar al narrador en dicha escena.

11.4.3. Variaciones emocionales

En este experimento se propone validar la síntesis emocional de la LSE. Para ello se han planteado dos frases en cinco niveles distintos de intensidad (muy relajado, relajado, normal, tenso y muy tenso). La diferencia entre la relajación, normalidad y tensión se representa en el grado de tensión de los PFQ *configuración* y *movimiento* y diferencias en el PFQ *no manual*, principalmente representado con distintas expresiones faciales y movimientos de cabeza. Otro de los parámetros que consideramos que modifica el grado de tensión en un signo es el tiempo de transición entre distintos signos y la velocidad a la que se representan, los valores utilizados se recogen en la Tabla 11.12. La diferencia entre las alternativas ‘relajada’-‘muy relajada’ y ‘tensa’-‘muy tensa’ es la elección de la gaussiana (Figura 9.2) del algoritmo de Cinemática Inversa. Las alternativas extremas utilizan las gaussianas descritas como relajada y

⁶Esta aproximación ha sido aplaudida en el informe de la FCNSE.

Apartado 11. Evaluaciones subjetivas

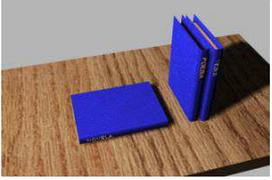
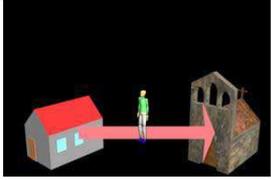
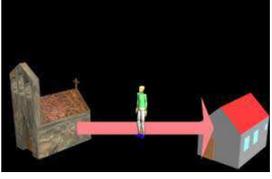
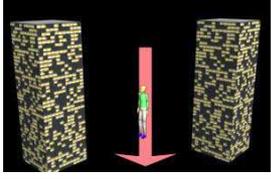
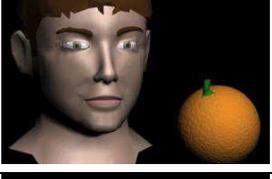
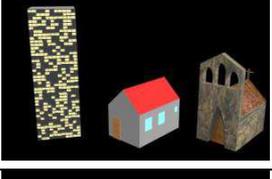
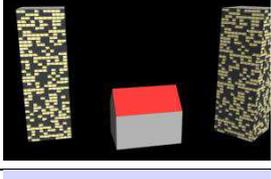
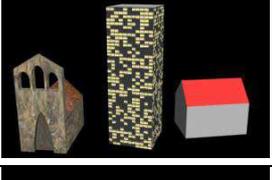
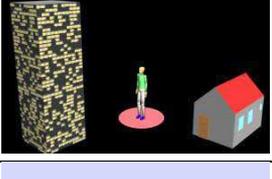
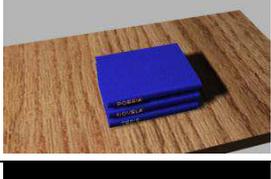
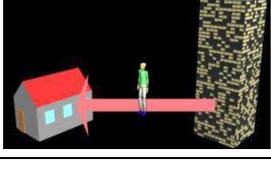
	Escena	Aciertos		Escena	Aciertos
1		100 %	9		100 %
2		90,9 %	10		100 %
3		90,9 %	11		85,7 %
4		100 %	12		100 %
5		100 %	13		100 %
6		100 %	14		50 %
7		75 %	15		100 %
8		100 %	16		83,3 %
Resultado Promedio					92,2 %

Tabla 11.11: Resultados de los test realizados para los clasificadores, esta tabla muestra el porcentaje de aciertos para cada una de las cuestiones planteadas.

tensa, mientras que las alternativas moderadas utilizan la misma gaussiana que la aproximación normal.

Variante	Tiempo de transición	Velocidad promedio
Relajado	400ms	83,43 %
Normal	300ms	100,00 %
Tenso	200ms	134,80 %

Tabla 11.12: Esta tabla muestra el tiempo de transición entre signos y la velocidad de representación de los signos para las tres variaciones emocionales.

En este experimento se han utilizado dos frases distintas: «Yo te he esperado aquí durante tres horas.» (frase 1) y «Yo quiero que tú arregles el telescopio hoy.» (frase 2); cada una de las cuales se ha presentado en las cinco variaciones distintas. Los resultados obtenidos para este experimento tras encuestar a seis expertos en LSE se muestran en las Tablas 11.13 y 11.14, en los que se muestra cómo las variaciones en la tensión han sido correctamente percibidas, aunque no en su intensidad.

Variante Representada	Opción Marcada				
	Muy Relajado	Relajado	Normal	Tenso	Muy Tenso
Muy Relajado	33,3 %	50 %	16,7 %		
Relajado		66,7 %	33,3 %		
Normal			100 %		
Tenso			50 %	50 %	
Muy Tenso				66,7 %	33,3 %

Tabla 11.13: Esta tabla contiene el porcentaje de respuestas para cada una de las categorías en las preguntas realizadas para la frase 1.

Variante Representada	Opción Marcada				
	Muy Relajado	Relajado	Normal	Tenso	Muy Tenso
Muy Relajado	16,7 %	50 %	33,3 %		
Relajado	16,7 %	66,7 %	16,7 %		
Normal			100 %		
Tenso			16,7 %	50 %	33,3 %
Muy Tenso				33,3 %	66,7 %

Tabla 11.14: Esta tabla contiene el porcentaje de respuestas para cada una de las categorías en las preguntas realizadas para la frase 2.

Los usuarios han percibido que ciertas modificaciones introducidas en la expresión facial de los signos, con el objeto de enfatizar el estado emocional, han alterado su morfología de manera indebida. Por lo tanto, el uso de distintas expresiones faciales para representar distintos estados emocionales debe ser estudiado con mayor profundidad.

11.5. Informe de la Fundación CNSE

Para concluir este apartado de evaluaciones subjetivas, hemos considerado oportuno incluir el informe realizado por la Fundación CNSE (FCNSE) sobre las evaluaciones realizadas. Este informe, que se reproduce íntegro en el Apéndice E, presenta de manera detallada la opinión de los expertos en el «Área de Investigación en LSE» de la FCNSE.

La lectura detallada de dicho informe muestra la aceptación del sintetizador propuesto por parte de este grupo de expertos. En los siguientes puntos resumimos una serie de apreciaciones obtenidas de los comentarios de dicho informe:

- El orden en el que se realizan los experimentos es importante. Esta opinión es compartida por los expertos de la FCNSE, ya que, por error, se interpretó que el orden en el que se debían realizar coincidía con el identificador de los tests.
- Es necesaria la intervención de un experto en LSE en la definición/validación de los signos. El documento contiene muchas referencias a modificaciones que se pueden incluir en la definición de ciertos signos. Estas modificaciones permiten mejorar la naturalidad de los signos, pero sólo un experto puede asegurar que dicha modificación no interferirá con el significado de los signos.
- Se debe fomentar la flexibilidad en la definición de los signos. Los comentarios presentes en este informe sobre la morfología de los signos muestran que las modificaciones que se deben realizar sobre los mismos requieren gran precisión y libertad. Por esta razón se debe incluir estas cualidades en las herramientas definidas.
- Existen errores de sintaxis en algunas de las frases. Aunque estos experimentos no tenían por objetivo evaluar la corrección de las estructuras gramaticales de las frases en LSE, sí que es conveniente que su sintaxis sea correcta. Por esta razón se debe intentar que toda frase evaluada haya sido validada en sus aspectos sintácticos.
- Este conjunto de experimentos ha permitido validar, en cierta medida, los movimientos labiales del avatar. Aunque este no era el objetivo de los test que han realizado los expertos de la FCNSE. Los comentarios presentes en el informe valoran positivamente el movimiento de los labios del avatar.
- Aunque los resultados obtenidos en la evaluación de síntesis emocional son aceptables, este informe remarca que se debe evitar el uso de las expresiones

faciales para la representación de emociones. Se deberá estudiar con mayor profundidad las variaciones que se producen en la expresión facial durante los distintos estados emocionales.

11.6. Resumen de los resultados

A continuación se muestra una tabla con el resumen de los resultados obtenidos en las distintas evaluaciones recogidas en este apartado. La Tabla 11.15 contiene el resumen de los experimentos no relacionados con la síntesis de LSE. La Tabla 11.16 presenta los resultados obtenidos al evaluar la inteligibilidad del avatar y otros aspectos relativos a la síntesis de LSE.

Experimento	Usuarios	Perfil usuarios	Resultados
11.1.1 Lectura de labios: uso de alófonos.	nº: 11 edad: 10–88	Personas no expertas en lectura de labios.	Med: 57 % Var: 14 %
11.1.1 Lectura de labios: uso de grafemas.	nº: 11 edad: 22–54	Personas no expertas en lectura de labios.	Med: 64 % Var: 11 %
11.1.2 Movimiento labial con sonido, primer plano. ^a	nº: 40 edad: 22–54	Público en general.	Med:3,7 Var: 0,2
11.1.2 Movimiento labial con sonido, plano general.	nº: 40 edad: 18–57	Público en general.	Med:3,5 Var: 0,21
11.1.2 Movimiento labial sin sonido, alófonos.	nº: 11 edad: 22–37	Público en general.	Med:2,9 Var: 0,1
11.1.2 Movimiento labial sin sonido, grafemas.	nº: 11 edad: 22–37	Público en general.	Med:2,8 Var: 0,1
11.1.3 Expresiones faciales.	nº: 45 edad: 18–62	Público en general.	Med: 57,15 % Var: 21,91 %
11.2.1 Tensión de la mano. ^b	nº: 20 edad: 20–57	Público en general.	Med: 79,17 % Var: 16,39 %
11.2.2 Tensión en el hombro.	nº: 21 edad: 20–62	Público en general.	Med: 59,12 % Var: 11,5 %

^aLa media y la varianza se muestran en un rango de 1 a 5 para todos los experimentos 11.1.2.

^bSe consideran sólo aciertos absolutos.

Tabla 11.15: Resumen de las evaluaciones subjetivas.

Experimento	Usuarios	Perfil usuarios	Resultados
11.3.1.1 Signos Aislados, respuesta libre: avatar Yuli.	nº: 6 edad: 29-40	Profesores de LSE, tres de ellos nativos.	Med: 67,5 % Var: 9,2 %
11.3.1.2 Signos Aislados, respuesta libre: avatar Daniel.	nº: 7 edad: 24-50	Expertos en LSE, que la utilizan diariamente.	Med: 90,7 % Var: 13 %
11.3.1.2 Signos Aislados, respuesta libre relajada: avatar Daniel.	nº: 7 edad: 24-50	Expertos en LSE, que la utilizan diariamente.	Med: 96,4 % Var: 6,1 %
11.3.2 Signos Aislados, respuesta cerrada: avatar Daniel.	nº: 7 edad: 24-50	Expertos en LSE, que la utilizan diariamente.	Med: 98,6 % Var: 2,7 %
11.4.1 Frasas completas, secuencias de signos.	nº: 10 edad: 24-50	Expertos en LSE, que la utilizan diariamente.	Med: 95 % Var: 5 %
11.4.2 Frasas completas, construcciones clasificatorias.	nº: 11 edad: 24-50	Expertos en LSE, que la utilizan diariamente.	Med: 92,2 % Var: 12,3 %
11.4.3 Frasas completas, prosodia. ^a	nº: 6 edad: 24-50	Expertos en LSE, que la utilizan diariamente.	Med: 83,4 % Var: 13,3 %

^aConsideramos aciertos en la variación prosódica, sin importar la intensidad presentada.

Tabla 11.16: Resumen de las evaluaciones sobre síntesis de signos.

Capítulo IV

Herramientas y Aplicaciones Desarrolladas

Apartado 12

Aplicaciones para la generación de contenido

El proceso para la definición del contenido de la base de datos relacional es especialmente sensible, ya que el resultado de la síntesis depende en gran medida de la precisión y corrección de la información almacenada. La diversa naturaleza de los datos y la participación de expertos en áreas dispares justifica la necesidad de desarrollar distintas aplicaciones para llevar a cabo esta labor, en concreto dos: la primera centrada en la gestión de los niveles tercero y cuarto de la base de datos relacional y la segunda para los niveles primero y segundo.

Las dos aplicaciones, presentadas en este apartado, modifican directamente el contenido de la base de datos relacional. Gracias a la arquitectura del sintetizador (ver Sección 5.1) con una definición de los signos centralizada, las modificaciones pueden realizarse simultáneamente por distintas personas y estar disponible para la síntesis de manera inmediata.

12.1. Contenido de niveles tercero y cuarto

El objetivo de esta aplicación es la gestión del contenido de los niveles tercero y cuarto de la base de datos relacional. Recordemos que el tercer nivel está constituido por las unidades de cada uno de los PFQ; a su vez, cada unidad se define mediante una cadena de caracteres. El cuarto nivel contiene las definiciones de cada una de las unidades descritas en el tercer nivel y que pueden realizarse mediante *Quaternions*, *Nombre de Huesos* o *Coordenadas*. Las unidades cuya definición se basa en que el *Nombre de Hueso* no requieren el uso de una aplicación especial, ya que sus valores

pueden ser introducidos manualmente en la base de datos relacional. En el caso de los *Quaternions* o de las *Coordenadas* si que es preciso contar con una herramienta para la definición de cada unidad, lo que afecta directamente a cinco de los siete PFQ (*orientación*, *configuración*, *movimiento*, *no manual* y *plano*).

La aplicación permite obtener la definición de los *Quaternions* y de las *Coordenadas* de dos maneras diferentes: la primera cargando un fichero «m3g» creado por otra aplicación (en nuestro caso 3ds Max 8 [4]) que contenga una animación de la que se pueda extraer la información requerida; la segunda es mediante la definición de los valores en la interfaz de la aplicación. En la Figura 12.1 se muestra el panel de la aplicación que permite la definición del PFQ *configuración*: cada control deslizante gestiona la rotación de una de las articulaciones de la mano, lo que permite definir fácilmente todas las unidades de este PFQ y sus variaciones de tensión. La misma aproximación se ha seguido para la *orientación*, *movimiento* y *plano*.

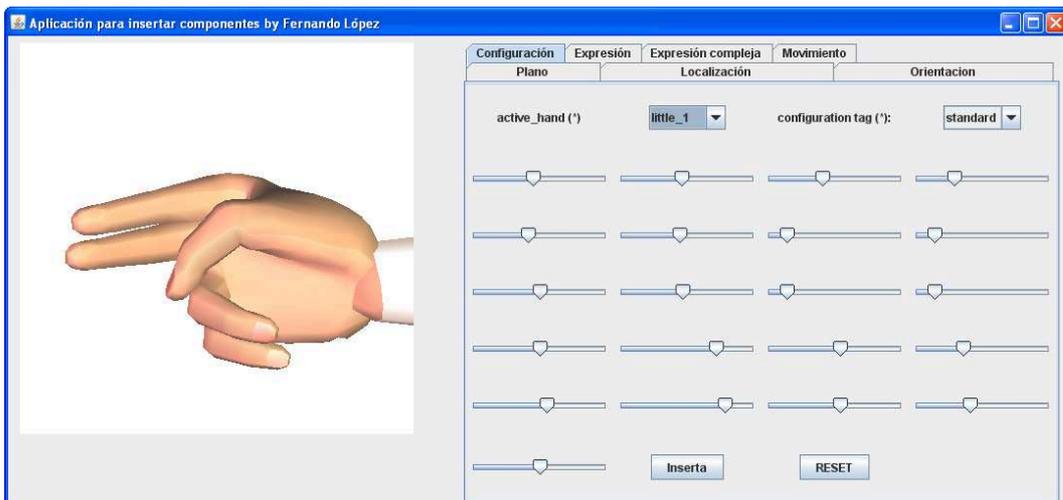


Figura 12.1: Aplicación de captura para los PFQ.

Esta aplicación también contempla la gestión específica que requiere la particular definición del PFQ *no manual*. Recordemos que este PFQ se ha descrito en dos niveles: ‘expresiones simples’ y ‘expresiones complejas’. Esta aplicación permite definir tanto las ‘expresiones simples’ que asignan un *Quaternion* a un grupo de huesos preestablecido, como las ‘expresiones complejas’, estableciendo para ello el conjunto de ‘expresiones simples’ que lo conforman.

12.2. Contenido de niveles primero y segundo

El objetivo de la segunda aplicación es proporcionar una interfaz de trabajo que permita la definición de los signos *establecidos* acorde a la estructura de almacenamiento de la base de datos relacional. El primer nivel de la base de datos relacional define las entradas para cada uno de los signos *establecidos* y el segundo define las secuencias de cada PFQ que describen al signo. Recordemos que la escala temporal de estas secuencias se define mediante porcentajes de la duración del signo.

Para ello se ha diseñado una interfaz gráfica que permite describir un signo utilizando únicamente el ratón y el método *Drag&Drop* para todos los PFQ. La Figura 12.2 muestra una captura de esta herramienta correspondiente a la descripción del PFQ *localización*. El usuario arrastra desde la imagen superior a la barra de tiempos de la mano correspondiente el valor definido para este PFQ. La Figura 12.3 muestra los paneles correspondientes a otros PFQ; el tratamiento se realiza de manera idéntica para todos ellos. La gestión del PFQ *no manual* permite usar tanto ‘expresiones complejas’ como ‘expresiones simples’ en la descripción del signo. En la definición de PFQ *orientación* se ha utilizado un conjunto de iconos similares a la definición SignWriting debido a que se trata de una notación bastante intuitiva.

Gracias a la arquitectura del sistema es posible que varias personas trabajen colaborando en la generación de un signo. Además, en el trabajo realizado ha quedado patente que una persona sin conocimientos de la LSE puede definir un signo utilizando un vídeo como referencia, aunque la definición del signo puede contener fallos que sólo un experto en LSE podría detectar. Utilizando esta aplicación también se pueden realizar las modificaciones necesarias sobre la definición existente, lo que permite repartir el trabajo de definición de un signo entre distintas personas, liberando así al experto de carga de trabajo, y permitiéndole que se concentre en la validación de la descripción.

En el diseño de la aplicación se ha tenido en cuenta además, el hecho de que un conjunto importante de potenciales usuarios sean sordos y, posiblemente, nativos de la LSE. Esto implica minimizar la información presentada en texto ya que, para este grupo de personas la capacidad de comprensión del texto es reducida, potenciando un diseño de interfaz basado en imágenes. En la versión actual los botones con las acciones más relevantes se describen usando texto, pero se prevee su sustitución por iconos en futuras modificaciones.

La experiencia obtenida al instruir a diversas personas, con distinto grado de conocimientos técnicos, en el uso de esta herramienta indica que el periodo medio

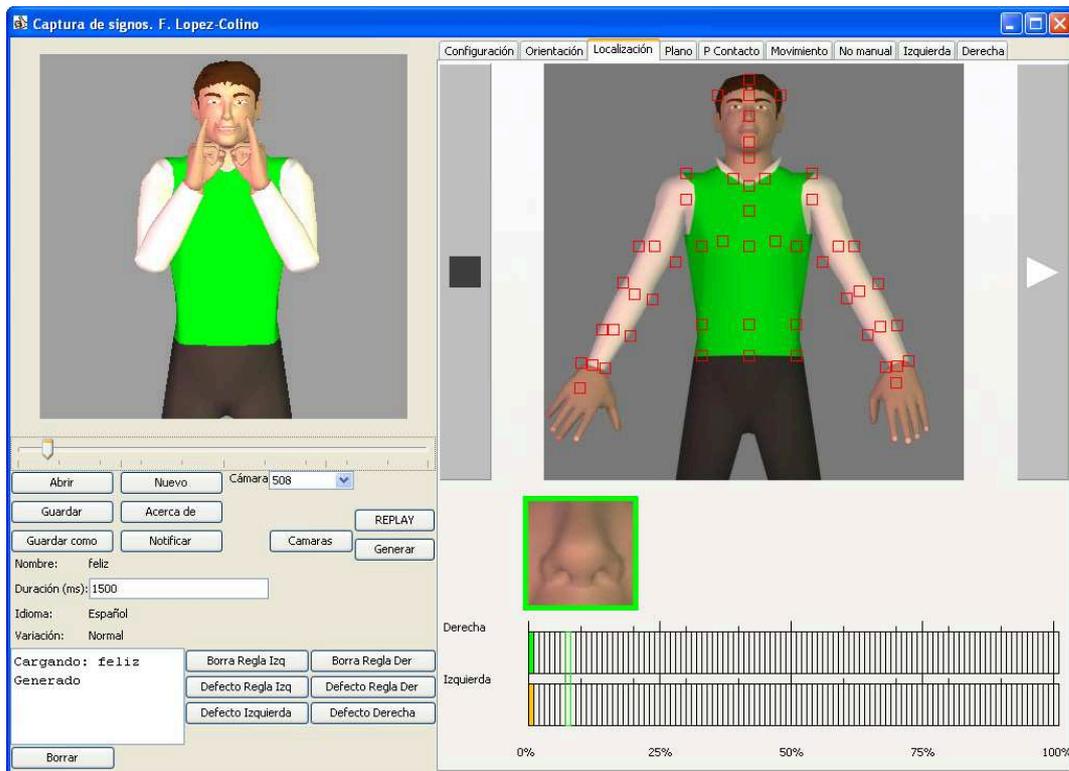


Figura 12.2: Vista general de la aplicación utilizada para la definición del contenido de la base de datos relacional para los niveles uno y dos.

de aprendizaje es de aproximadamente cinco minutos. La diferencia entre distintos usuarios se refleja en el tiempo necesario para describir un signo, dependiendo de su habilidad y experiencia, observándose que la mayor parte del tiempo se invierte en ajustes para mejorar la calidad del signo y no en la definición en sí. En la Tabla 12.1 se muestra el tiempo medio de distintos usuarios en la definición de los signos.

Usuario	Tiempo
1	4,7min.
2	10,0min.
3	10,3min.
4	2,5min.

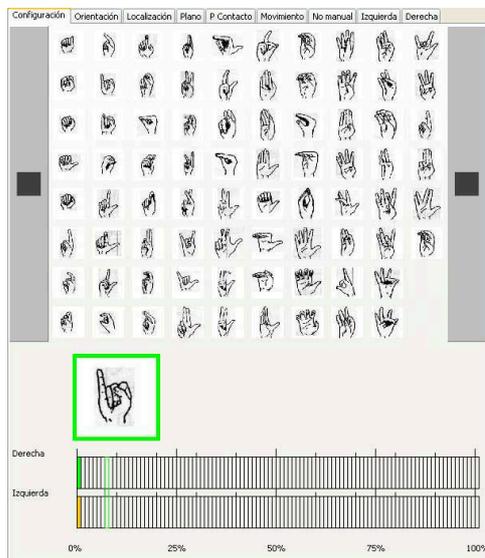
Tabla 12.1: Tiempo medio de definición de un signo con la aplicación desarrollada. Se debe tomar como cota inferior los valores del usuario 4, creador de la herramienta.

Aunque esta aplicación se ha desarrollado para introducir la definición de los signos en la base de datos relacional, se plantea su uso en otros ámbitos:

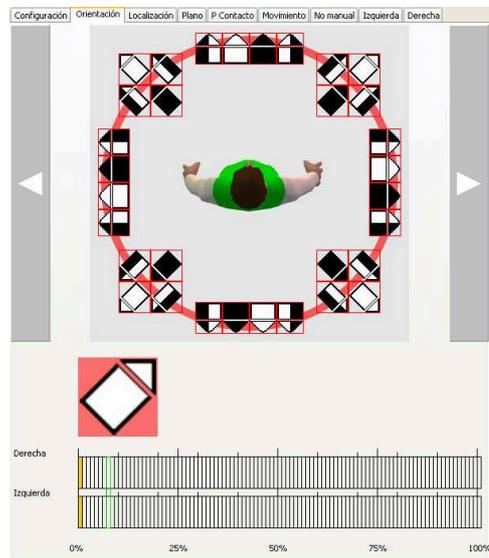
- *Educación.* Los usuarios que utilizaron esta aplicación para definir signos han demostrado haber aprendido con gran facilidad los signos con los que estuvieron

trabajando, mientras que otros signos que sólo habían sido observados no fueron recordados. Esto permite aventurar que esta aplicación puede ser adaptada para la generación de aplicaciones educativas, incluyendo herramientas de ejercicios y herramientas de evaluación.

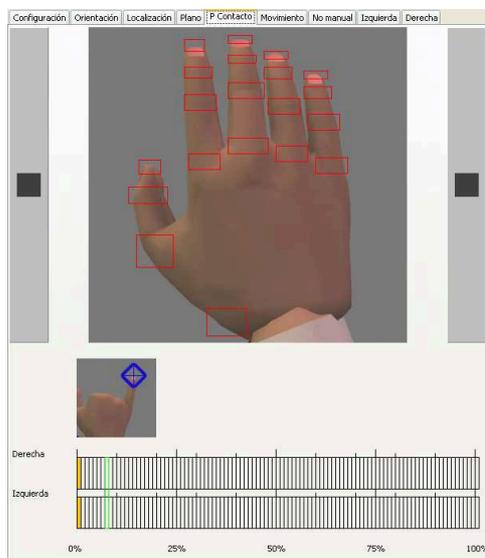
- *Investigación y Colaboración.* Al igual que el proyecto de Crasborn *et al.* [26], el uso de una base de datos relacional que contenga la definición de un concepto en distintos dialectos facilitará su comparación y estudio. En la generación del *Diccionario Normativo de la LSE* [36] se ha dedicado una gran parte del tiempo a la obtención de una versión unificada de los signos para todas las regiones de España. El uso de esta herramienta podría facilitar la colaboración entre los distintos equipos de lingüistas en las primeras fases de trabajo.



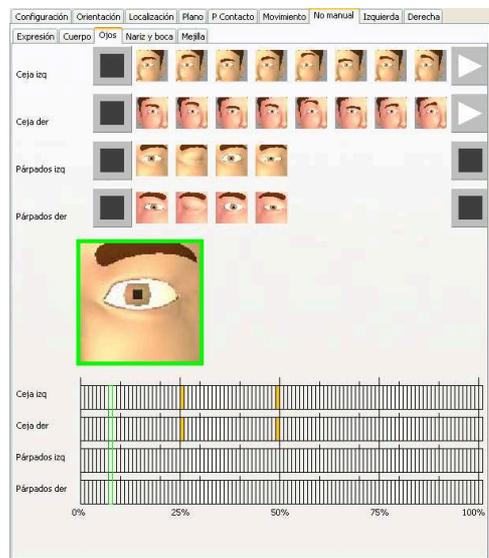
(a) Panel de configuración.



(b) Panel de orientación.



(c) Panel de punto de contacto.



(d) Panel de no manual.

Figura 12.3: Ejemplo de distintos paneles utilizados para la definición de diferentes PFQ en la aplicación utilizada para la definición del contenido de la base de datos relacional para los niveles uno y dos.

Apartado 13

Herramientas auxiliares

A lo largo del periodo de trabajo ha sido necesario el desarrollo de dos herramientas para labores puntuales: la primera durante el desarrollo del algoritmo de Cinemática Inversa (IK) para probar y validar su funcionamiento, la segunda para corregir el problema observado en la transición entre distintas *orientaciones*. En este apartado se presentan brevemente estas dos aplicaciones.

13.1. Aplicación IK y control de colisiones

La primera de las aplicaciones auxiliares desarrolladas se creó inicialmente con el objetivo de probar y validar el algoritmo de IK que se ha implementado para este trabajo (Figura 13.1). Esta validación contempla dos aspectos: el primero es la verificación de que el algoritmo es funcional, es decir, que sitúa la muñeca en la posición correcta; el segundo es comprobar que se cumplen las restricciones anatómicas definidas para evitar que articulaciones como el hombro o la muñeca adquieran rotaciones imposibles. Para ello, la aplicación auxiliar permite definir, para ambas manos la posición en la que se desea situar la muñeca y la *orientación* que se le aplica, realizando los cálculos de cinemática inversa, mostrando en el cuadro de información superior de la aplicación todos los valores definidos en el algoritmo de IK (ver Sección 9.2). En relación con la IK, esta aplicación también permite modificar el valor del ángulo Φ , actualizando igualmente los cálculos relativos.

En una etapa posterior se incluyó el cuadro de información inferior de la aplicación con el objetivo de validar el algoritmo de detección de colisiones. En este cuadro de información se recoge información para conocer si el hueso de la muñeca o del codo izquierdos están en colisión con el cuerpo o la cabeza.

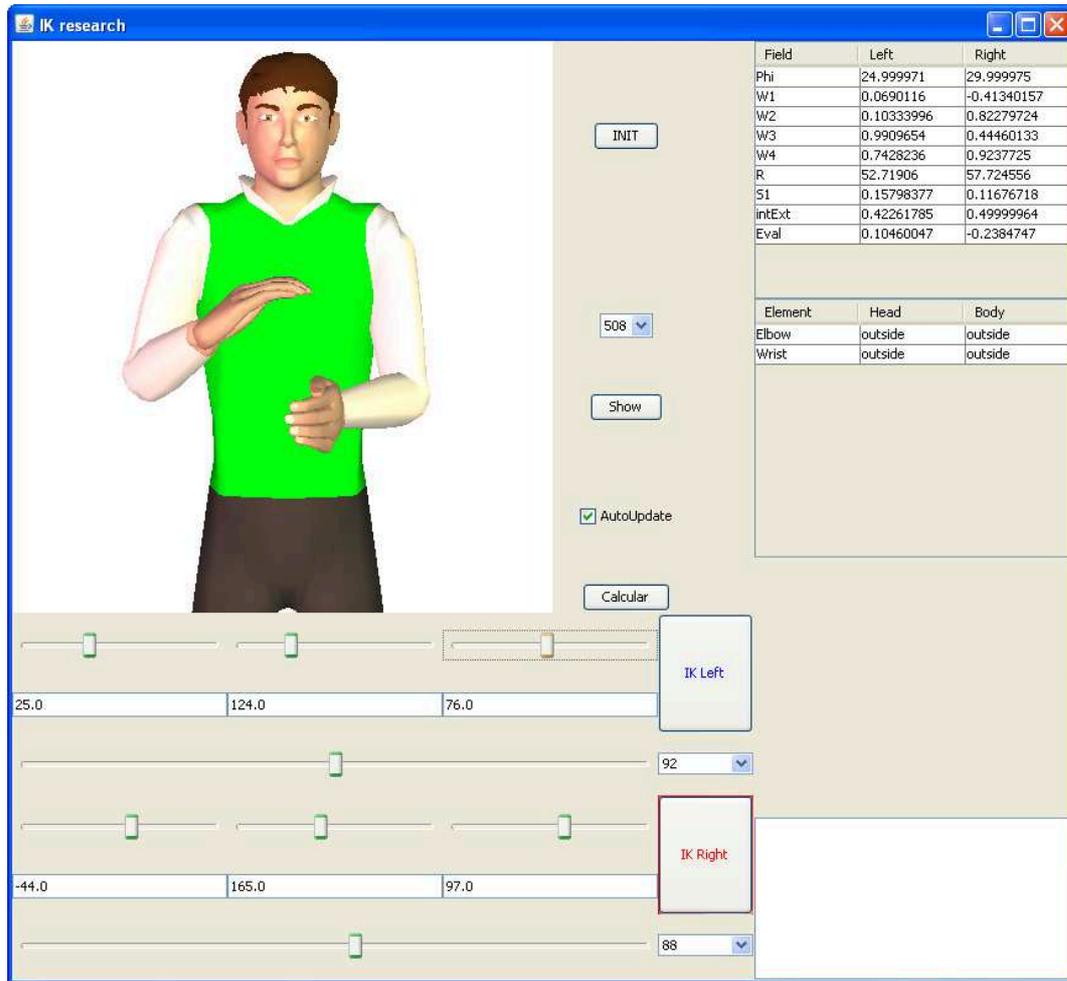


Figura 13.1: Aplicación desarrollada para el estudio del algoritmo de IK planteado en este trabajo.

13.2. Definición de transiciones en el PFQ *orientación*

La segunda aplicación auxiliar (Figura 13.2) se plantea con el objetivo de resolver los fallos en la transición entre *orientaciones* derivado del diseño de la muñeca (ver Punto 6.1.2.1); al desvincular la orientación de la muñeca de los valores heredados del brazo y del antebrazo, las transiciones entre *orientaciones* no tienen en cuenta las restricciones anatómicas de esta articulación, ya que la transición se realiza interpolando los *Quaternions* a lo largo del arco circular más pequeño (transición directa). La solución planteada consiste en definir las transiciones incorrectas así como el arco que se debe describir para realizar la transición entre dos orientaciones. Esta nueva transición se describe, bien a lo largo del arco mayor entre las dos orientaciones (transición indirecta) o bien especificando una *orientación* intermedia. La

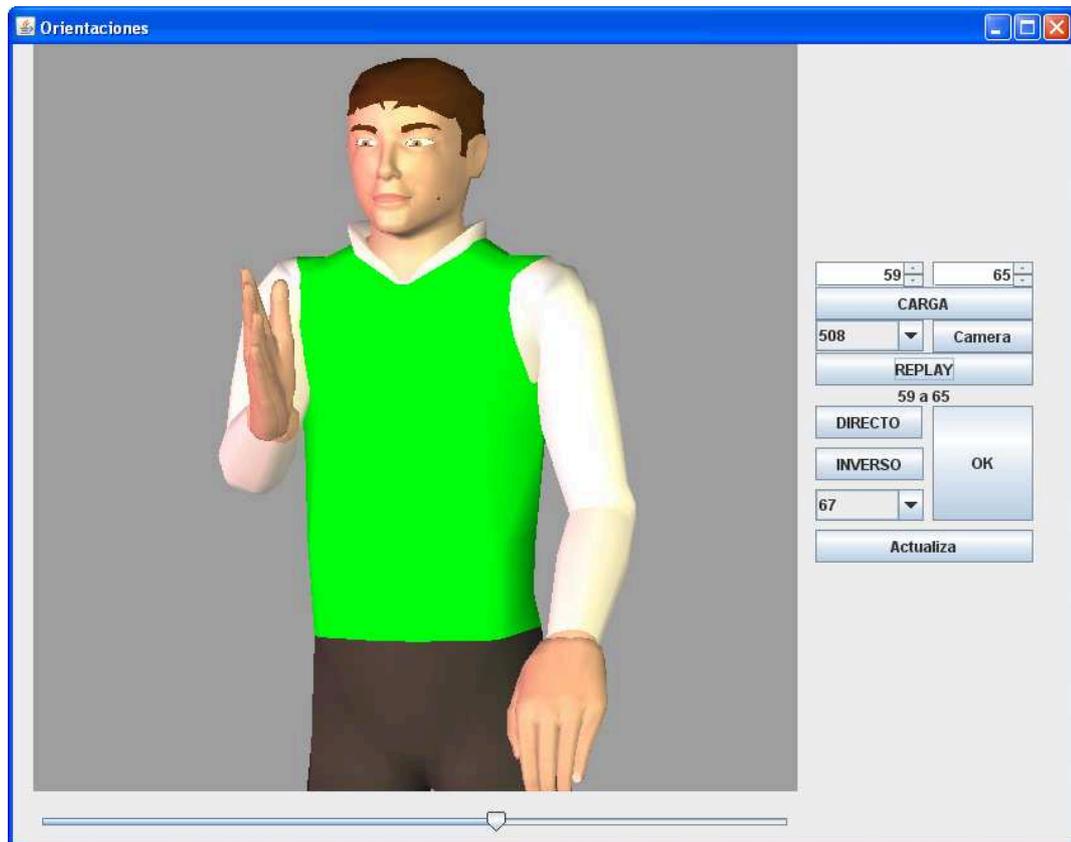


Figura 13.2: Aplicación desarrollada para verificar y resolver la transición entre dos *orientaciones*.

Tabla 13.1 muestra el número de transiciones entre dos *orientaciones*, clasificadas según la solución utilizada.

Tipo de transición	Número de transiciones
Directas	7511
Indirectas	3279
Usando orientación intermedia	26
Total	10816

Tabla 13.1: Transiciones entre orientaciones.

Apartado 14

Aplicaciones de usuario

El último grupo de programas desarrollado lo componen aquellas aplicaciones utilizadas para generar la síntesis, es decir, las destinadas a los usuarios finales del sintetizador. Se han creado aplicaciones para dos plataformas distintas: PC y PocketPC, que además han permitido verificar la viabilidad de los escenarios distribuidos α , δ y γ (ver Sección 5.2); para probar la viabilidad del escenario β se han utilizado además, aplicaciones de visualización de vídeo creadas por terceros.

14.1. Aplicación para PC

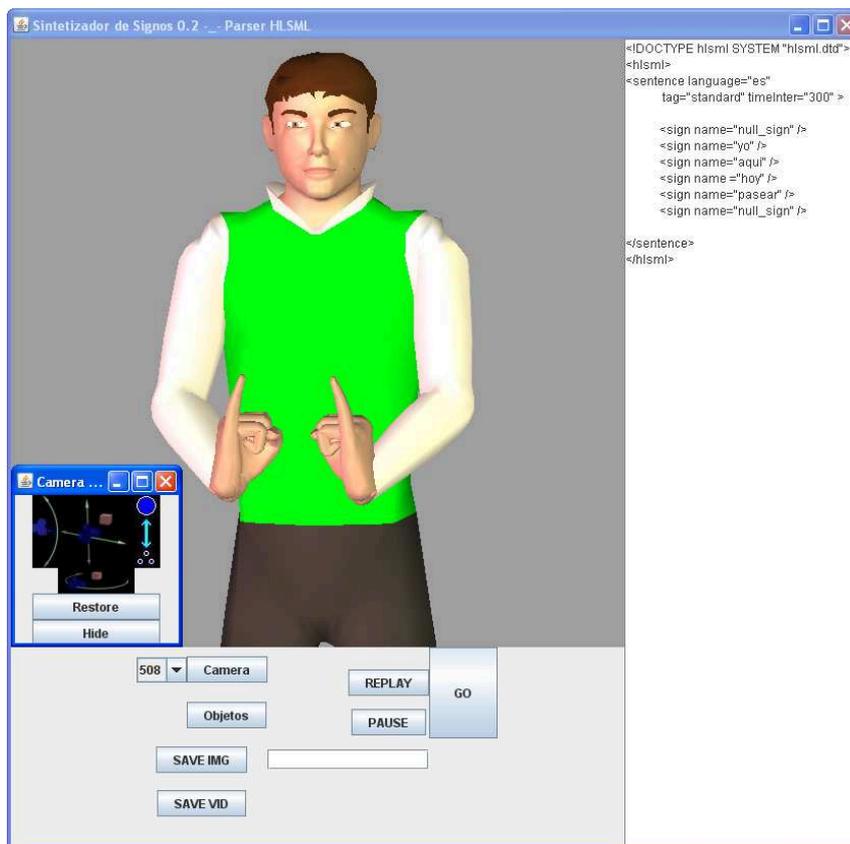
La aplicación creada para PC (Figura 14.1) utiliza el planteamiento del escenario α , en el que todo el proceso de síntesis se realiza en el dispositivo del cliente. En el escenario δ , las pistas de animación se definen en el servidor y a continuación, se envían al cliente que las asocia con los huesos correspondientes; la aplicación desarrollada para PC establece este proceso de manera interna para la generación de la animación final. Esta aplicación se ha implementado en Java, utilizando la API JSR-184 para J2SE desarrollada por la empresa Hybrid [52]. La versión 0.1 de esta aplicación permitía alternativamente la síntesis de secuencias de signos *establecidos* y el deletreo de palabras. Posteriormente, se incorporó un módulo que permite la gestión de un mensaje definido en HLSML y la síntesis; contemplando las características descritas en dicha notación, esta mejora dio lugar a la versión 0.2.

La versión 0.2 de este programa se ha utilizado para las medidas descritas en las evaluaciones realizadas en el Apartado 10, así como para la generación de los vídeos correspondientes a los casos de prueba necesarios para las evaluaciones del Apartado 11.

Sintetizador Paramétrico Multidispositivo de LSE



(a) Aplicación de síntesis 0.1.



(b) Aplicación de síntesis 0.2.

Figura 14.1: Aplicación diseñada para la síntesis de signos en un PC.

14.2. Aplicación para PocketPC

La aplicación desarrollada para la plataforma PocketPC (Figura 14.2) se ha creado siguiendo la aproximación del escenario γ . Este programa reproduce una secuencia de animación contenido en un fichero «m3g». Esta aplicación ha sido desarrollada en J2ME y requiere la instalación previa en el dispositivo PocketPC de una máquina virtual de Java; en este caso se ha optado por utilizar el desarrollo de Esmertec [31], ya que incluye una implementación del estándar JSR-184.

La generación del fichero «m3g» es una característica de la versión 2.0 del estándar JSR-184, actualmente en desarrollo. Los ficheros que se reproducen en esta aplicación han sido creados manualmente.



Figura 14.2: Aplicación diseñada para la visualización de ficheros «m3g» en una PDA.

14.3. Visualización en Nintendo DS Lite

El escenario δ se aplica para dispositivos en los que sólo se puede visualizar un vídeo. Para demostrar la utilidad de este escenario se ha utilizado una Nintendo DS Lite. Este dispositivo tiene conectividad WIFI (802.11b) y permite visualizar vídeos

en ambas pantallas, con una resolución de 256×192 píxeles. Se ha utilizado un cartucho «M3 DS REAL» junto con el reproductor «Moonshell» para la visualización de vídeos obtenidos como resultado de la síntesis. Estos vídeos deben estar en formato DPG; para esta codificación se utilizó «dpgtools 1.71». Este dispositivo tiene suficiente capacidad para mostrar los vídeos a 20 fps en cualquiera de sus dos pantallas (Figura 14.3).



Figura 14.3: Visualización de la síntesis en una Nintendo DS Lite.

Capítulo V

Conclusiones y Trabajo Futuro

Apartado 15

Conclusiones

15.1. Resumen de aportaciones

En este trabajo hemos presentado un sintetizador de Lengua de Signos Española que combina algunas de las características de los sintetizadores actuales, tales como la definición paramétrica de los signos o el uso de un avatar virtual [63, 83], con una serie de nuevas características: i) el diseño modular (que permite la adaptación del proceso de síntesis a dispositivos con diferentes capacidades de procesamiento), ii) el diseño del avatar signante, iii) el modo en que se definen y almacenan los signos, iv) la notación que describe el mensaje a sintetizar, v) la síntesis con modificaciones emocionales y vi) el desarrollo de un conjunto de herramientas para facilitar el proceso de definición de los signos sin tener que utilizar un lenguaje de modelado 3D.

15.1.1. Diseño modular del sintetizador

El sintetizador se ha desarrollado utilizando una arquitectura basada en módulos independientes, lo que permite distribuir el proceso de síntesis entre distintos dispositivos. El objeto de este diseño ha sido facilitar la adaptación del complejo proceso de síntesis a los recursos disponibles en el dispositivo que utiliza el usuario para acceder al contenido signado en cada momento, así como a la capacidad y coste de la conexión. Modificando la asignación de la ejecución de los distintos módulos entre un servidor de síntesis y el propio dispositivo del cliente se han establecido cuatro escenarios distintos, con diferentes niveles de exigencia para la red de comunicaciones y el dispositivo cliente.

15.1.2. Estructura del avatar

El diseño que se ha seguido en este trabajo al crear el avatar incluye un conjunto de características novedosas, desarrolladas específicamente para la síntesis de la lengua de signos:

- La inclusión de un conjunto de huesos auxiliares en la estructura esquelética del avatar que facilitan la definición y la síntesis de la lengua de signos:
 - Hueso de la muñeca: Este hueso facilita la gestión de PFQ *orientación* al independizar la orientación de la muñeca de los valores heredados del brazo y del antebrazo.
 - Huesos de localización: Estos huesos permiten obtener la ubicación de todas las referencias anatómicas que son necesarias para definir el PFQ *localización* sin tener que utilizar la malla del avatar.
 - Huesos de colisión: Estos huesos definen un volumen en torno a determinadas partes del avatar y permiten simplificar el proceso de detección y solución de colisiones. Al igual que los anteriores, eliminan la dependencia de la malla en este proceso.
 - Huesos direccionales: Son una variante de los huesos de animación, la característica de estos huesos es que su orientación se define a partir de la posición de un objeto auxiliar de la escena.
- La gestión de las expresiones faciales se realiza mediante el uso de animación esquelética, lo que permite unificar toda la animación del avatar mediante un único método.

15.1.3. Base de datos relacional

Este trabajo es el primero que se plantea, para la síntesis de lengua de signos, la utilización de una base de datos relacional donde se almacenan y describen, de manera, independiente los siete PFQ. La descripción de un signo que puede realizarse en una base de datos relacional es superior a la que se obtiene utilizando SignWriting, HamNoSys, SEA o Szczepankowski, principalmente porque estas notaciones no permiten definir la evolución temporal del mismo. Además, podemos almacenar y recuperar de forma simple distintas variaciones y alternativas, ya sea de la representación de un signo o de la de distintos PFQ, lo que permite, como se ha mencionado anteriormente, definir en la misma base de datos relacional diferentes variaciones dialectales y emocionales.

15.1.4. Notación de entrada: HLSML

Otra diferencia importante entre el sintetizador desarrollado aquí y otros existentes en la literatura es la definición de la notación de entrada HLSML, con un nivel de abstracción superior a notaciones como SiGML o SWML. Estas notaciones representan el único recurso para la descripción de los signos a sintetizar, lo que da lugar a mensajes con un alto nivel de detalle. En nuestra arquitectura, al estar la descripción almacenada en la base de datos relacional, se libera al mensaje de entrada de contener la información, con lo que se simplifica su descripción. HLSML permite describir de una forma simple sentencias en lengua de signos que contengan signos *establecidos*, secuencias de deletreo y construcciones clasificatorias. Así mismo permite definir un conjunto de parámetros para modificar el resultado de la síntesis, con el objeto de resaltar, en el grado requerido, distintas secciones de la frase a sintetizar.

15.1.5. Síntesis de emociones

Este trabajo ha realizado la primera aproximación a la síntesis de lengua de signos con modificaciones emocionales. Las variantes que se deben realizar en la síntesis afectan a la definición de determinados PFQ, a la descripción de los signos (en cuanto a su duración total y a la aceleración en la transición de los PFQ), y a las expresiones faciales independientes de la definición del signo. También son necesarias alteraciones en el proceso de Cinemática Inversa (IK) para definir la posición del codo así como la modificación del tiempo de transición entre dos signos consecutivos. Se ha evaluado si las variaciones en el PFQ *configuración* y las diferentes expresiones faciales que adopta el avatar son percibidas por los usuarios, siendo el resultado positivo aunque mejorable. Además hemos comprobado que el usuario es capaz de apreciar distintos grados de tensión en el brazo según los valores del ángulo Φ , utilizado al definir una solución para el algoritmo de IK. A partir de los resultados obtenidos, se ha comprobado que, con la aplicación de este conjunto de modificaciones, los usuarios perciben correctamente las distintas variaciones en la síntesis de signos.

15.1.6. Herramientas desarrolladas

La complejidad de la información necesaria para la síntesis de signos ha requerido el diseño de un conjunto de aplicaciones. Estas aplicaciones se distribuyen en tres grupos según su función: auxiliares, definición de signos y aplicaciones de usuario.

De entre todas las aplicaciones desarrolladas, la más relevante es la aplicación descrita en la Sección 12.2. Esta aplicación permite definir y modificar de manera

sencilla las descripciones almacenadas en la base de datos relacional de los signos *establecidos*.

15.1.7. Resultados obtenidos

Los resultados obtenidos en las distintas evaluaciones realizadas nos han permitido corroborar la validez del planteamiento y comprobar la corrección y el potencial del desarrollo realizado. No obstante, también han mostrado ciertas carencias en el sintetizador que deben ser mejoradas para incrementar la calidad de la síntesis.

Junto con los resultados obtenidos en los experimentos, se debe considerar la evaluación realizada por los expertos de la Fundación CNSE y que se ha incluido en el Apéndice E. Este informe recoge tanto un conjunto de sugerencias y errores detectados en el sintetizador, como la valoración positiva del trabajo realizado.

15.2. Publicaciones

Como es costumbre, para terminar el apartado de conclusiones presentamos una lista de las publicaciones relacionadas con el trabajo realizado:

15.2.1. Publicaciones nacionales

F. López-Colino, J. Beskow y J. Colás: *Mobile SynFace: Ubiquitous visual interface for mobile VoIP telephone calls*, en: Proceedings of Swedish Language Technology Conference (SLTC-08), volumen 1, pags. 57–58 (Stockholm (Sweden) 2008).

F. López-Colino y J. Colás: *HLSML, una notación de alto nivel para la descripción de mensajes sintéticos en LSE*, en: Proceedings of III Congreso Nacional de Lengua de Signos Española (CNSE - UNED, Madrid, Spain 2009).

F. López-Colino y J. Colás: *Nueva herramienta interactiva para la enseñanza de vocabulario de la lengua de signos española*, en: Proceedings of III Congreso Nacional de Lengua de Signos Española (CNSE - UNED, Madrid, Spain 2009).

15.2.2. Publicaciones internacionales

F. López, J. Tejedor, D. Bolaños y J. Colás: *Intérprete de Lenguaje de Signos en Español Multidispositivo*, en: Proceedings of IADIS-CIAWI, pags. 293–296 (IADIS 2006).

F. López, J. Tejedor, D. Bolaños y J. Colás: *Virtual Signer for Symmetrical E-Learning in Spanish Sing Language*, en: Proceedings of SIIE, volumen 1, pags. 403–407 (ADIE 2006).

F. López, J. Tejedor, J. Garrido y J. Colás: *Use of a Hierarchical Skeleton for Spanish Sign Language 3D Representation over Mobile Devices*, en: Proceedings of INTERACCION, VII International Conference of Human-Computer Interaction, pags. 565–568 (AIPO 2006).

J. Tejedor, **F. López**, D. Bolaños y J. Colás: *Augmented Service for Deaf People Using a Text to Sign Language Translator*, en: Proceedings of INTERACCION, VII International Conference of Human-Computer Interaction, pags. 601–604 (AIPO 2006).

F. López-Colino, J. Beskow y J. Colás: *Mobile Synface: Talking head interface for mobile VoIP telephone calls*, en: Proceedings of INTERACCION, X International Conference of Human-Computer Interaction, pags. XXX–XXX Editado por P. Latorre (AIPO, Barcelona, Spain 2009).

F. López-Colino, J. Garrido y J. Colás: *Description and Synthesis of Spanish Sign Language Classifiers*, en: Proceedings of INTERACCION, X International Conference of Human-Computer Interaction, pags. XXX–XXX Editado por P. Latorre (AIPO, Barcelona, Spain 2009).

F. López-Colino, J. Tejedor, J. Garrido y J. Colás: *Distributed Spanish Sign Language Synthesizer Architecture*, en: Proceedings of INTERACCION, X International Conference of Human-Computer Interaction, pags. XXX–XXX Editado por P. Latorre (AIPO, Barcelona, Spain 2009).

J. Tejedor, **F. López-Colino**, J. Garrido y J. Colás: *Integration of a Talking Head into a Spanish Sign Language Synthesizer*, en: Proceedings of INTERACCION, X International Conference of Human-Computer Interaction, pags. XXX–XXX Editado por P. Latorre (AIPO, Barcelona, Spain 2009).

Apartado 16

Trabajo futuro

En este apartado presentamos las principales líneas a seguir para continuar el trabajo iniciado en esta tesis. La síntesis de la lengua de signos es un tema de investigación que abarca disciplinas tan dispares entre sí como:

- Estudios lingüísticos de la propia lengua de signos.
- Generación de gráficos por ordenador.
- Estudio de las arquitecturas y recursos de los dispositivos móviles, para permitir el acceso al contenido en lengua de signos desde los mismos.
- Conocimientos de psicología para la generación de variaciones emocionales.

Esto hace que las posibles futuras líneas de investigación abarquen áreas muy dispares de conocimiento.

Para facilitar el acceso al resultado de la síntesis de lengua de signos desde distintos dispositivos, en este trabajo se han propuesto cuatro escenarios posibles, verificándose que todos son viables. Sin embargo, como se comentó en su momento, algunas de las características necesarias para automatizar completamente el proceso de adaptación al dispositivo no estuvieron disponibles durante el desarrollo, como es el caso del escenario γ , donde no se dispuso de la función necesaria para exportar la escena que contiene la animación generada. Actualmente, la versión 2.0 de este estándar [60] se encuentra en fase de desarrollo y se espera que incluya dicha funcionalidad, o en su defecto se llevaría a cabo el desarrollo necesario para exportar las escenas generadas. De igual modo, se ha verificado el funcionamiento del escenario δ sin embargo, aún no se ha iniciado el desarrollo en otras arquitecturas 3D, como VRML o XNA, cuyo

principal objetivo es permitir el uso de aplicaciones desarrolladas con otras librerías 3D.

Por otra parte, los experimentos realizados en la Sección 10.2 muestran que los cálculos relativos a la Cinemática Inversa y a la detección de colisiones son los que mayor tiempo de proceso requieren. Por tanto una modificación a realizar en estos dos procesos es la optimización con el objeto de disminuir el tiempo necesario para sintetizar un signo.

En este trabajo se han utilizado los últimos estudios disponibles sobre la estructura y composición de la lengua de signos. Es de esperar que, según se avance en el estudio de la LSE y se consolide una teoría que describa las lenguas de signos, se precisen modificaciones en el proceso de síntesis, descripción del HLSML, generación de los *clasificadores*, etc., para recoger nuevas características. De igual modo, es necesario profundizar en la composición generada a partir de la definición que realiza el mensaje HLSML para alterar un signo y en la descripción contenida en la base de datos relacional, ya que un signo incluido en la misma habrá sido validado previamente por un experto en LSE, pero en dicha definición no se ha tenido en cuenta el contexto en que se desarrolla el mensaje, ni los movimientos corporales que pueden acompañarlo. Por otra parte, aunque el mensaje HLSML se genera con conocimiento del contexto, incorporando modificaciones prosódicas en el mensaje, cabe la posibilidad de que alguna de las modificaciones propuestas entren en conflicto con la definición del signo, alterando su significado de manera no deseada. Según avancen los estudios de la LSE, se amplíe el contenido de la base de datos relacional y se realicen más pruebas con usuarios podremos proponer una solución para una correcta composición entre la descripción realizada en el mensaje HLSML y la contenida en la base de datos relacional.

Los resultados obtenidos en las evaluaciones realizadas han demostrado que la propuesta actual del sintetizador es válida, pero también que existen ciertos aspectos tales como las expresiones faciales, el movimiento labial y la representación de la tensión que, aunque percibidas correctamente por los usuarios, pueden y deben ser mejoradas en cuanto a calidad y precisión.

Este trabajo ha presentado una primera aproximación a la síntesis de lengua de signos que permite añadir información prosódica y emocional. Aunque los resultados obtenidos son prometedores, deben ser considerados únicamente como el comienzo de una nueva línea de investigación. Como se planteó en la Sección 4.3, actualmente sólo se representan las variaciones emocionales a través de modificaciones en las articulaciones del avatar simulando alteraciones en la tensión muscular; estas variaciones en la tensión muscular afectarían al flujo de sangre en una persona real, lo

que produciría variaciones en la apariencia de la piel. La definición de la malla del avatar dispone de los requisitos necesarios para contemplar estas modificaciones, sin embargo, el desarrollo y evaluación de estos cambios no se ha realizado todavía y por lo tanto constituyen parte de futuras mejoras.

Otro aspecto que se debe tratar sobre las modificaciones prosódicas es cómo se realiza su definición. Actualmente, las variaciones emocionales se describen de manera manual. Consideramos que puede ser interesante, una vez se alcance un número considerable de signos definidos en las tres alternativas realizar un análisis de las modificaciones realizadas para tratar de extrapolar un patrón que pueda ser aplicado de manera automática sobre nuevos signos y así obtener una primera aproximación a las variaciones emocionales.

El programa desarrollado para describir los signos en la base de datos relacional (ver Sección 12.2) ha demostrado tener un campo de aplicación superior al inicialmente considerado. Cuando se inició la colaboración con distintos expertos, vieron, en este programa, la posibilidad de generar una herramienta de apoyo a la enseñanza de la lengua de signos; se constató cómo personas que utilizaban este programa demostraban una mayor retención de los signos por ellos introducidos en la base de datos relacional en comparación con otros signos que simplemente habían observado. Esto nos lleva a plantearnos la posibilidad de generar una versión del programa para complementar la docencia tradicional, permitiendo la evaluación de los conocimientos del estudiante.

La base de datos relacional y el conjunto de herramientas que hemos desarrollado en esta tesis permiten la extensión de este sintetizador para su uso en otros idiomas. Una futura línea de acción debería cubrir la posible adaptación de estas herramientas para su uso en otros países y presentar una alternativa que permita la generación de contenidos signados en distintos idiomas. Esta colaboración permitirá mejorar en gran medida los resultados obtenidos por este sintetizador.

Apéndice

Apéndice A

Glosario de abreviaturas

ASL	American Sign Language
BSL	British Sign Language
DGS	German Sign Language
HamNoSys	Hamburg Notation System
HLSML	High Level Signing Markup Language
HML	HamNoSys Markup Language
IK	Cinemática Inversa
LS	Lengua de Signos
LSE	Lengua de Signos Española
PFQ	Parámetro Formativo Quinésico
SEA	Sistema de Escritura Alfabética
SiGML	Signing Gesture Markup Language
SLN	Sign Language of the Netherlands
SWML	SignWriting Markup Language

Apéndice B

Quaternions

Los quaternions son elementos de un espacio de cuatro dimensiones en el que están definidas tres operaciones: suma, producto por un escalar y producto vectorial, siendo esta última no conmutativa. Los quaternions fueron descritos por primera vez por Hamilton [41]. Dado que parte del vector de un quaternion pertenece a R^3 , la geometría de R^3 se refleja en los quaternions, tal es el caso de las rotaciones.

Shoemake [94] considera ventajoso usar los quaternions para representar la orientación de un cuerpo en el espacio, particularmente para aplicaciones de animación. La ventaja radica en que la animación de la orientación de un cuerpo es más fácil, dado que la interpolación entre dos quaternions transcurre a lo largo del arco esférico que une dichas orientaciones, característica que no ocurre siempre con otras representaciones de la orientación (ver Figura B.1):

Sean dos orientaciones A y B descritas mediante los quaternions $Q_A = (X_A, Y_A, Z_A, W_A)$ y $Q_B = (X_B, Y_B, Z_B, W_B)$ las orientaciones intermedias Q de la forma $Q = iQ_A + (1 - i)Q_B$ siendo $i \in [0, 1]$ definen la circunferencia c . La interpolación entre Q_A y Q_B se realizará por el arco mayor o menor de c según el signo de $(Q_A \cdot Q_B)$

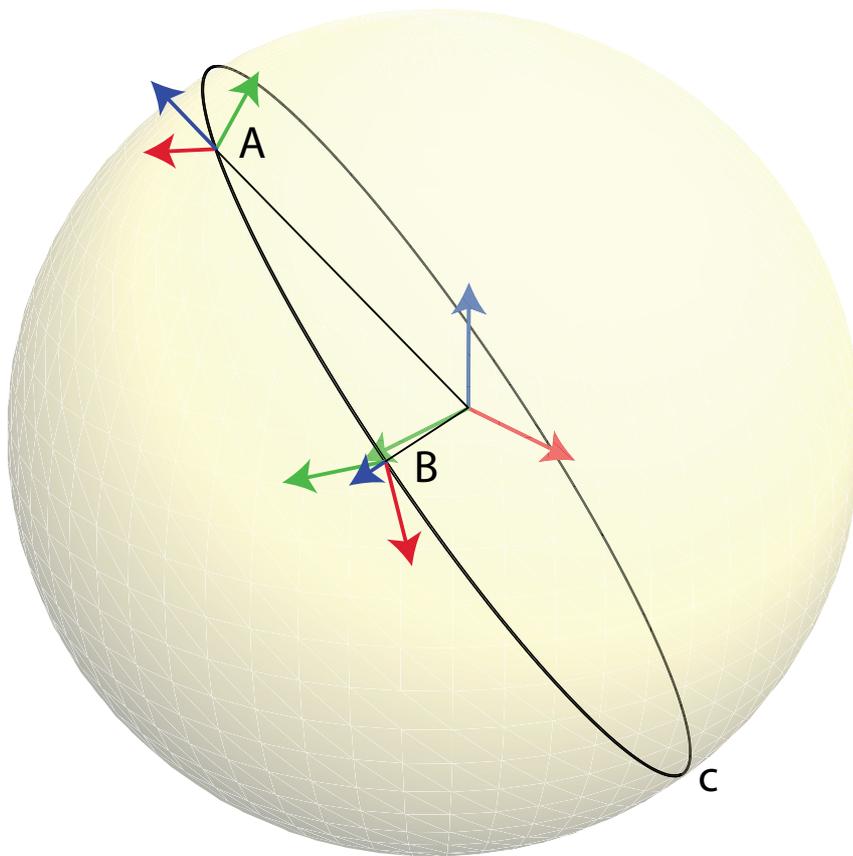


Figura B.1: Resultado de la interpolación de cuaterniones.

Apéndice C

Diagrama detallado de la base de datos relacional

En este apéndice presentamos el diagrama de la base de datos relacional con mayor detalle que en la Figura 7.2, facilitando su estudio.

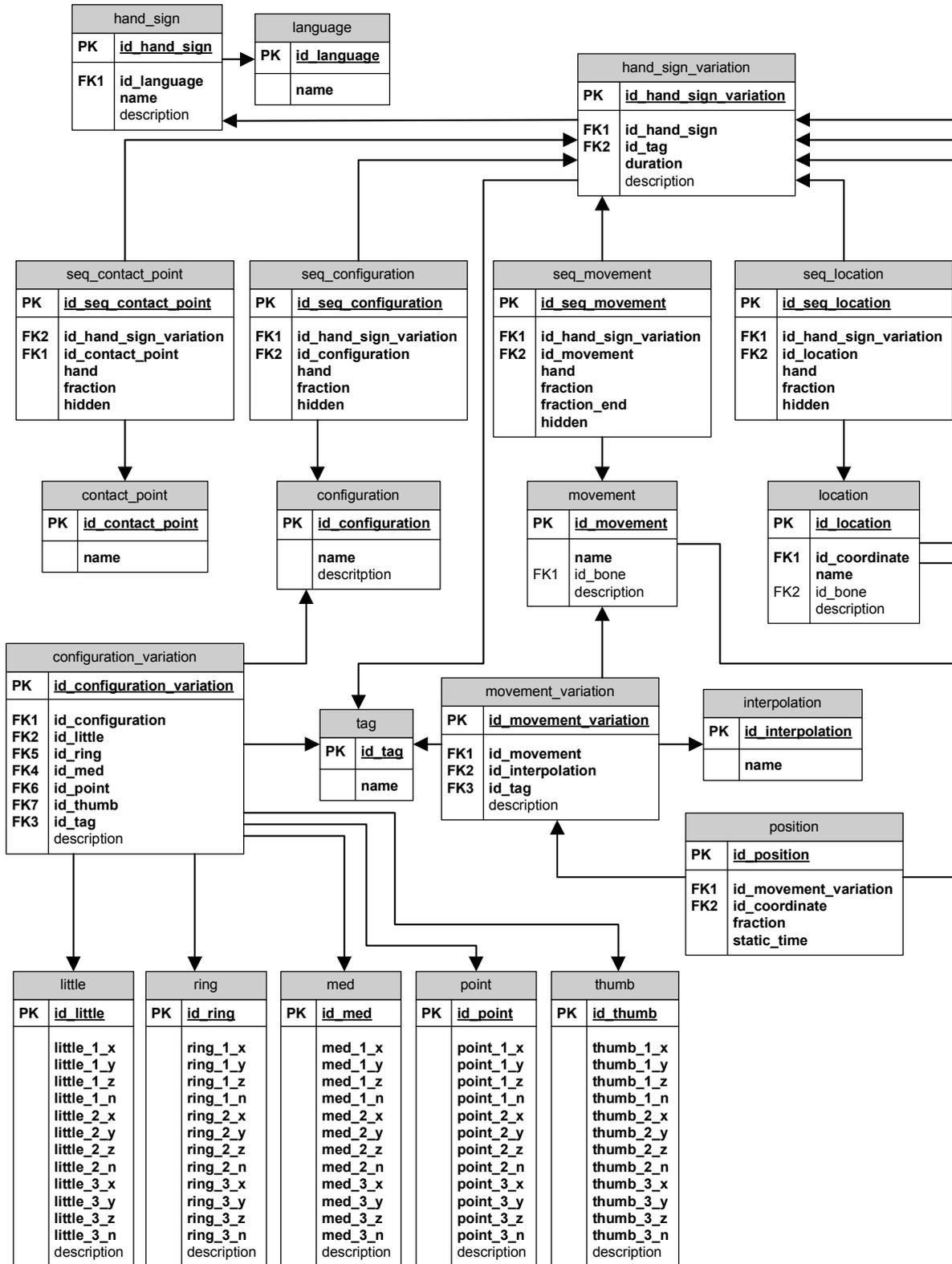
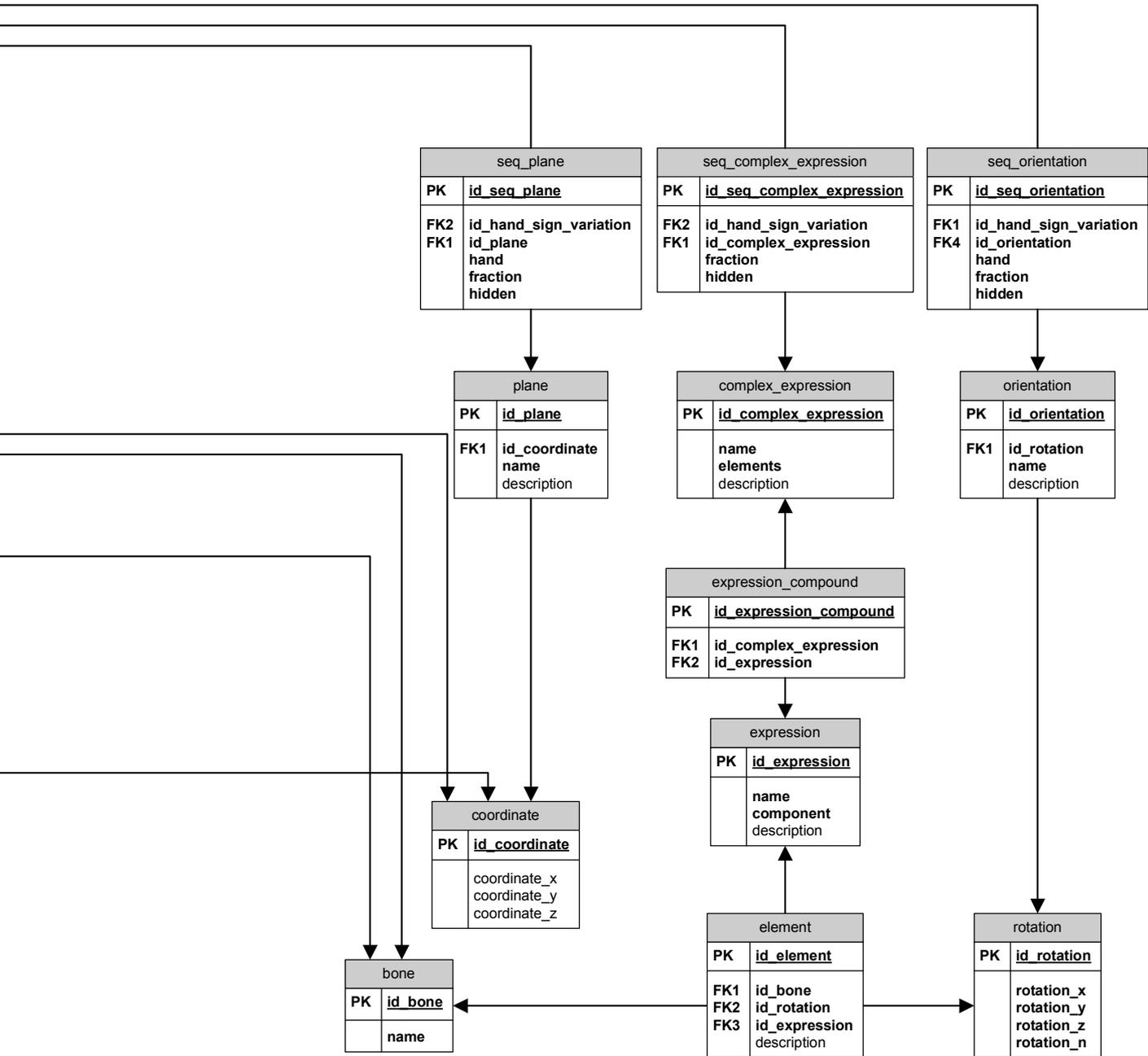


Figura C.1: Diagrama detallado de la base de datos relacional.

Apéndice C. Diagrama detallado de la base de datos relacional



Apéndice D

Jerarquía de los elementos definidos en HLSML

En este apéndice se muestra el árbol jerárquico de los elementos que componen la notación HLSML.

```

<hlsml>
  *<sentence>
    ↪  *<sign>
      ↪  ↪  *<spellSign/>
      ↪  ↪  *<signMod/>
      ↪  ↪  *<talk>
      ↪  ↪  ↪  *<viseme/>
      ↪  ↪  ↪  *<signClassifier>
      ↪  ↪  ↪  *<element/>
      ↪  ↪  ↪  *<sign>
      ↪  ↪  ↪  *<coordinate/>
      ↪  *<spellSign/>
      ↪  *<globalMod/>
      ↪  *<classifierSequence>
      ↪  ↪  *<classifier>
      ↪  ↪  ↪  *<hand>
      ↪  ↪  ↪  ↪  *<configuration>
      ↪  ↪  ↪  ↪  ↪  *<element/>
      ↪  ↪  ↪  ↪  ↪  *<orientation>
      ↪  ↪  ↪  ↪  ↪  *<element/>
      ↪  ↪  ↪  ↪  ↪  *<coordinate/>
      ↪  ↪  ↪  ↪  ↪  *<contactPoint>
      ↪  ↪  ↪  ↪  ↪  *<element/>
      ↪  ↪  ↪  ↪  <headLookAt>
      ↪  ↪  ↪  *<element/>
      ↪  ↪  ↪  *<coordinate/>
      ↪  ↪  ↪  *<eyesLookAt>
      ↪  ↪  ↪  *<element/>
      ↪  ↪  ↪  *<coordinate/>
      ↪  ↪  ↪  *<nonHand>
      ↪  ↪  ↪  *<element/>
      ↪  ↪  ↪  *<viseme/>
      ↪  ↪  *<classifierTransition/>
      ↪  ↪  *<classifierSequence>

  *<signDefinition>tabsize=4
    ↪  *<signMod/>
    ↪  *<configurationSequence>
    ↪  ↪  *<fraction>
    ↪  ↪  ↪  *<element/>
    ↪  *<orientationSequence>
    ↪  ↪  *<fraction>
    ↪  ↪  ↪  *<element/>
    ↪  *<locationSequence>
    ↪  ↪  *<fraction>
    ↪  ↪  ↪  *<element/>
    ↪  *<planeSequence>
    ↪  ↪  *<fraction>
    ↪  ↪  ↪  *<element/>
    ↪  *<contactSequence>
    ↪  ↪  *<fraction>
    ↪  ↪  ↪  *<element/>
    ↪  *<movementSequence>
    ↪  ↪  *<fraction>
    ↪  ↪  ↪  *<element/>
    ↪  *<nonHandSequence>
    ↪  ↪  *<fraction>
    ↪  ↪  ↪  *<element/>
    ↪  *<talk>
    ↪  ↪  *<viseme/>
  *<compound>
    ↪  *<sentence>
    ↪  *<nonHandSequence>
    ↪  ↪  *<fraction>
    ↪  ↪  ↪  *<element/>
    ↪  *<talk>
    ↪  ↪  *<viseme/>
    ↪  *<classifierSequence>

```

Apéndice E

Informe de la Fundación CNSE

A continuación se presenta el informe realizado por la Fundación CNSE sobre los experimentos subjetivos relativos a la síntesis de LSE. Este informe es propiedad de la Fundación CNSE.

La correspondencia entre los identificadores de test utilizados en el informe y las secciones de este documento se presenta en la Tabla E.1.

Experimento	Sección
9	11.4.3
10	11.4.2
11	11.3.1.2
12	11.3.2
13	11.4.1

Tabla E.1: Correspondencia de los tests y las secciones de este documento.



Área de Investigación
en LSE

INFORME DE EVALUACIÓN

"SINTETIZADOR PARAMÉTRICO MULTIDISPOSITIVO DE LSE"

Realizado por el Área de Investigación en LSE de la Fundación de la Confederación Estatal de Personas Sordas. 25 Junio 2009. Madrid.

Lo primero, quisiéramos felicitarnos por la idea del proyecto en sí, y la forma de llevarlo a cabo. Estamos ilusionados con las perspectivas que plantea.

Entendemos que es una primera piedra en lo que será la construcción de un sistema de "transcripción de la LSE", el cual irá ampliándose y mejorándose con revisiones y aportaciones de los usuarios signantes. Pero por eso mismo, porque los cimientos son una parte imprescindible para la sostenibilidad futura, estamos especialmente motivados a la hora de colaborar con nuestra aportación en forma del informe que ahora tenéis entre manos.

El proyecto AVATAR, con sus ventajas de ser concebido para ser ampliado de manera sencilla mediante una base de datos y con la opción de elegir diferentes dispositivos, nos hace pensar que pronto llegará de manera real a usuarios que podrán beneficiarse de este sistema.

Como lingüistas - y pensando ahora en la estandarización de la LSE - nos parece un recurso muy bueno. Aun que aún falte mucho por hacer.



Área de Investigación
en LSE

RESPECTO A LOS TESTS

Las preguntas iniciales, donde piden los datos al usuario, nos resultan sencillas y cómodas de entender. Valoramos positivamente que tengan iconos visuales que ayuden a su fácil comprensión. Al igual que, una vez en el test, el procedimiento para completarlo sea guiado, fácil y cómodo. Que la mayoría de los experimentos tengan preguntas cerradas agiliza y facilita la consecución del test.

Es cierto que la primera impresión más destacable y extendida al ver al “muñeco” es una sensación de frialdad. La falta de expresividad que se transforma en falta de información, hace relativamente incómoda la labor de entendimiento – para un usuario signante requiere más esfuerzo entender al avatar que a una persona signante, obviamente -, por eso el objetivo de querer que la expresión facial y los movimientos corporales estén presentes nos parece más que acertado, ya que completa información que de otro modo faltaría (eso sí, es importante que dé la información adecuada y en consonancia con el mensaje; ya que si es incorrecta puede variar el contenido del mensaje, haciéndolo contradictorio o incluso ininteligible).

En general a la hora de realizar los tests, vemos que su duración es adecuada, excepto quizás el test 9, que al repetir continuamente la misma información sin apenas variación, se hace un poco más.

Creemos que las preguntas son precisas, claras y concisas. Nos parece positivo que el usuario pueda repetir ver la pregunta tantas veces como quiera.

Tenemos una propuesta respecto a las respuestas; al ser obligatorio escoger una respuesta para pasar a la siguiente pregunta, se ha dado el caso de no estar de acuerdo con ninguna respuesta de las propuestas, pero tener que seleccionar “la más parecida”, que no correcta, para poder seguir con la consecución del test [es el caso de las preguntas 3 y 14 del test 10, por ejemplo]. Como solución, proponemos que entre las posibles respuestas aparezca la opción de “ninguna de las anteriores” “ns/nc” o cualquier otra que dé la posibilidad de manifestar que ninguna de las respuestas es adecuada según el parecer del usuario, o que no lo ha entendido.

Inscrita en el Registro del Ministerio de Educación, Cultura y Deporte. LE0572-12-081. Fundación Documental asociada a la Ley 30/1994. NIF: G-89115274.

www.fundacioncnse.org

C/ Islas Aleutianas, 28. 28035 Madrid. Tel: 91 376 85 60. Fax: 91 376 85 64
e-mail: invesmat@fundacioncnse.org



Área de Investigación
en LSE

RESPECTO A LA LSE

Queríamos ser lo más precisos y exhaustivos posible en esta parte, así que al revisar y evaluar una por una las preguntas de cada test, hemos descrito pormenorizadamente cada fallo encontrado, con su explicación y propuesta de cambio. Hemos estructurado esta información en tres áreas; morfológica, léxica y sintáctica.

Aclaración; dado que la sintaxis varía dependiendo de la frase, situación y mensaje que se está contando (modificando en algunas ocasiones la configuración y/o localización y/o movimiento del signo), hemos puesto todos los cambios en todas y cada una de las frases de un mismo test (por eso muchos de ellos están repetidos). Sin embargo al corregir una configuración, localización o movimiento en la morfología, no lo hemos vuelto a repetir en siguientes preguntas, ya que NO cambia de una frase a otra. Por ejemplo; el signo "YO: El signo es un poco exagerado, parece que el avatar *se disloca la muñeca* para hacerlo. Para solucionarlo, el avatar podría girar el tronco levemente (aunque eso a lo mejor produce "ruido visual", al ser maquinal y no natural...)" solo lo hemos indicado la primera vez que sale en el test 9 en la pregunta 1, y aunque el signo YO se repite en otras ocasiones como en ese mismo test todas las otras preguntas, o en la pregunta 2, 7, 9, 10, 16 del test 10, o en la pregunta 13 del test 11, o en las preguntas 1, 2, 3, 4, 5 y 9 del test 13, no lo volvemos a explicar de nuevo, sino que ponemos "Ídem", que se refiere a la primera explicación ya dada.



Área de Investigación
en LSE

EXPERIMENTO 9. FRASES CON "EMOCIÓN" [10 PREGUNTAS]

Una observación previa; desconocemos el objetivo concreto de la traducción al castellano escrito en la parte inferior del vídeo, pero creemos que es un tanto *peligrosa* en el sentido de que puede orientar la comprensión de la lengua signada, y viciar el objetivo real de los experimentos, que suponemos será si la persona signante usuaria comprende el avatar sin necesidad de apoyo en la lengua castellana. Proponemos eliminar toda referencia explícita al castellano - la traducción al castellano escrito en este caso - para la comprensión del avatar en la medida de lo posible. Si fuese realmente necesario aportar más información que la puramente signada, proponemos una transcripción de la LSE que, aun que se escribe con palabras concretas castellanas, mantiene la gramática de la LSE siendo esta forma más neutral y aséptica a la hora de comprender el mensaje signado que poniendo directamente la frase en castellano escrito. Así, os escribimos en cada caso nuestra propuesta glosada.

SINTAXIS

Preguntas 1, 2, 4, 7 y 10:

Frase en castellano escrita en el test < Yo te he esperado aquí durante tres horas >
Transcripción signada de lo que signa el avatar < YO TÚ AQUÍ HORA TRES DEBER ESPERAR >

Aun que, en realidad, la frase signada no es correcta. Para signar el contenido de la frase en castellano escrita en el avatar, el avatar debería signar < YO A TI ESPERAR-asp.durativo HORAS-TRES AQUÍ>

Preguntas 3, 5, 6, 8 y 9:

Frase en castellano escrita en el test < Yo quiero que tú arregles el telescopio hoy >
>
Transcripción signada de lo que signa el avatar < YO QUERER TÚ ARREGLAR TELESCOPIO HOY >

Aun que, en realidad, la el orden sintáctico de la frase signada no es correcta. Para signar el contenido de la frase en castellano escrita en el avatar, el avatar debería signar < YO QUERER HOY TÚ TELESCOPIO ARREGLAR >

www.fundacioncnse.org

C/ Islas Aleutianas, 28. 28035 Madrid. Tel: 91 376 85 60. Fax: 91 376 85 64
e-mail: invesmat@fundacioncnse.org



Área de Investigación
en LSE

SEMÁNTICA / LÉXICO

1, 2, 4, 7 y 10:

DEBER: Aquí el concepto está equivocado; la frase se entiende mal, ya que DEBE implica obligación. Suponemos que la confusión está motivada por querer traducir del castellano "he tenido que esperarte". En LSE no se dice así, si no < YO A TI ESPERAR asp.durativo HORAS-TRES >

TRES: En LSE los cuantificadores varían dependiendo de a qué se refieran; no es lo mismo - ni se signa igual - tres horas, que tres minutos, que tres días...

El signo correcto sería HORAS-TRES (haciendo con la mano dominante activa un movimiento circular de muñeca en sentido de las agujas del reloj)

MORFOLOGÍA

1, 2, 4, 7 y 10:

TIEMPO/HORA: Avatar da 3 golpes en la muñeca. El signo correcto tiene una repetición de solo 2 golpes.

DEBER: La extensión del movimiento es un poco exagerada [aunque, como hemos dicho, el signo DEBER en esta frase debería ser eliminado].

ESPERAR: La localización del signo no es correcta (las manos están demasiado juntas en el pecho, haciendo que la producción de ese signo sea incómoda, y la comprensión del mismo dificultosa)

YO: El signo es un poco exagerado, parece que el avatar *se disloca la muñeca* para hacerlo. Para solucionarlo, el avatar podría girar el tronco levemente - como en el caso de la pregunta 17 de este mismo test - (aunque eso a lo mejor produce "ruido visual", al ser maquinal y no natural...)

3, 5, 6, 8 y 9:

YO: ídem.

ARREGLAR: El avatar hace correctamente la configuración y localización de este signo, pero el movimiento no; mueve solo una, en vez de ambas manos con movimiento asimétrico entre sí, que es como se *pronuncia* correctamente el signo en LSE.

Inscrita en el Registro del Ministerio de Educación, Cultura y Deporte. I.E.O.E. 22.12.091. Fundación Docentes Asociada a la Ley 30/1994. N.I.E. G.89112224.

www.fundacioncnse.org

C/ Islas Aleutianas, 28. 28035 Madrid. Tel: 91 376 85 60. Fax: 91 376 85 64
e-mail: invesmat@fundacioncnse.org



Área de Investigación
en LSE

TELESCOPIO: El avatar hace correctamente la configuración y localización de este signo, pero el movimiento no; mueve solo una, en vez de ambas manos con movimiento asimétrico entre sí, que es como se *pronuncia* correctamente el signo en LSE.

HOY: El avatar signa con tres golpes, mientras que el signo correcto es con un solo golpe.

Componentes no manuales (CNM)

- Expresión facial

<< A continuación se presentan dos vídeos de la misma frase, el de la izquierda es neutro y deberá decir si el de la derecha le parece: *mucho más tenso, más tenso, igual, menos tenso o mucho menos tenso.* >>

Creemos que hay un error de concepto en la función que cumplen los componentes no manuales dentro de la oración, ya que – según el título del test – se busca la expresión de emoción, cuando en realidad en estos casos concretos la información que deberían transmitir las cejas y boca sería más sintáctico-discursiva que emocional.

6. En la pantalla de la derecha, el patrón labial “FFF” es correcto [aplicárselo al resto de situaciones donde se signa QUERER]

Inscrita en el Registro del Ministerio de Educación, Cultura y Deporte. (BOE 29-12-98). Fundación Docentes Asociada a la Ley 30/1994. NIF: G-89115274.

Propiedad de CNSE

www.fundacioncnse.org

C/ Islas Aleutianas, 28. 28035 Madrid. Tel: 91 376 85 60. Fax: 91 376 85 64
e-mail: invesmat@fundacioncnse.org



Área de Investigación
en LSE

EXPERIMENTO 10. CLASIFICADORES EN LS [16 PREGUNTAS]

En este test vemos que no hay traducción al español, y lo agradecemos.

Las frases se entienden muy bien, y gramaticalmente son correctas en su mayoría.

El concepto de clasificador en LS está claro, si bien requiere de alguna matización que comentaremos oportunamente, aparte de estos dos apuntes generales;

- Los clasificadores requieren siempre una expresión facial (si el elemento clasificado es grueso, fino, si es rápido o lento, si está alejado o cerca... Por ejemplo en la pregunta 11, donde sí se cambia el tamaño del clasificador "coger un libro", pero la expresión facial no cambia nada... es incorrecto; debería "morderse el labio inferior" si el libro es grueso, y "sacar la puntita de la lengua" si el libro es fino.)
- Cuando se producen varios clasificadores para ubicar dos o más objetos representados en el espacio – si es relevante en el mensaje – es necesario que se utilicen las dos manos para ubicarlos ambos objetos al mismo tiempo, al menos durante un momento de toda la producción de la frase. También es recomendable dejar un clasificador de referencia (el de la mano no dominante en forma pasiva) para facilitar la comprensión, como tan bien se ha hecho en las preguntas 6 y 7, por ejemplo.

SINTAXIS

1. El avatar signa TRES LIBRO. El orden correcto es LIBRO TRES.

El avatar signa de derecha a izquierda, y debería ser desde la izquierda hacia la derecha.

2. Al hacer el clasificador de iglesia y casa, hay que ubicarlas en el espacio a la vez en algún momento (mejor si se deja un clasificador como referente – el de la mano no dominante en forma pasiva - , como tan bien se ha hecho en las preguntas 6 y 7)

www.fundacioncnse.org

C/ Islas Aleutianas, 28. 28035 Madrid. Tel: 91 376 85 60. Fax: 91 376 85 64
e-mail: invesmat@fundacioncnse.org



Área de Investigación
en LSE

- 3.** El avatar signa TRES COCHE. El orden correcto es COCHE TRES. El avatar explica una imagen que no encontramos entre las posibles respuestas (aunque la más “parecida” es la respuesta d).
- 4.** Frase correcta. Expresión facial adecuada. Buen uso del espacio.
- 5.** Frase correcta. Expresión facial adecuada. Buen uso del espacio.
- 6.** Frase correcta. Expresión facial adecuada. Buen uso del espacio. El mantener el clasificador de casa a la izquierda como referente mientras se signa EDIFICIO para situar luego el clasificador de edificio al lado del clasificador de casa es muy correcto.
- 7.** Expresión facial adecuada. Buen uso del espacio. El mantener el clasificador de casa durante toda la frase es muy correcto.
- 8.** El avatar signa TRES COCHE. El orden correcto es COCHE TRES. El avatar explica una imagen que no encontramos entre las posibles respuestas (aunque la más “parecida” es la respuesta a).
- 9.** La frase sería más clara si utilizásemos las dos manos para describir visualmente cómo están colocados los libros. También sería más clara si dejásemos como referente un clasificador (mano no dominante en forma pasiva), como tan bien se ha hecho en las preguntas 6 y 7.
- 10.** Frase correcta. Expresión facial adecuada. Buen uso del espacio. El mantener el clasificador es muy correcto.
- 11.** El avatar signa TRES LIBRO. El orden correcto es LIBRO TRES. Es correcto que mantenga el clasificador del segundo libro mientras con la otra mano añade la información de que hay un tercer libro pegado a este primero. CNM. Expresión facial; Dependiendo del tamaño del libro, el avatar cambia la configuración del clasificador “coger un libro” (más amplio si el libro es grueso, más



Área de Investigación
en LSE

reducido si el libro es fino) es muy correcto. Sin embargo es incorrecto que no modifique en absoluto la expresión facial; debería "morderse el labio inferior" si el libro es grueso, y "sacar la puntita de la lengua" si el libro es fino.

12. Para explicar que hay "tres libros apilados unos sobre otros", la localización de la configuración tiene que ser más visual y *natural*. Es decir; los clasificadores de "libro" tienen que tener contacto total explicitando la posición real que tienen (ya que, como lo signa el avatar - las configuraciones no se tocan completamente - , parece que estén unos y otros haciendo como un "castillo de naipes", más que apilados unos sobre otros)

13. Frase correcta. El orden "CASA MÍO" es muy correcto. Buen uso del espacio. Los elementos EDIFICIOS y CASA tienen que ubicarse en el espacio "a la vez" en algún momento, para eso debe quedarse un clasificador como referente (el de la mano no dominante en forma pasiva), como tan bien se ha hecho en las preguntas 6 y 7.

14. El avatar signa TRES COCHE. El orden correcto es COCHE TRES. El avatar explica una imagen que no encontramos entre las posibles respuestas (aunque la más "parecida" es la respuesta a). Ya que signa los clasificadores como si los coches estuvieran los tres enfrente a la misma distancia del espectador, cuando en realidad la imagen muestra que el espectador tiene un coche más cerca, seguido de otro al lado, seguido de un último, que estaría más lejos todavía.

15. El avatar signa TRES LIBRO. El orden correcto es LIBRO TRES. Como en la pregunta 12, el clasificador de "dos libros apilados uno encima del otro", la localización de las configuraciones tiene que ser más visual y *natural*. Es decir; los clasificadores de "libro" tienen que tener contacto total explicitando la posición real que tienen (ya que, como lo signa el avatar - las configuraciones no se tocan completamente - , parece que estén haciendo como un "castillo de naipes", más que apilados uno encima del otro)

La configuración, localización y forma del tercer clasificador de "libro inclinado encima de los otros dos libros apilados" es muy correcta.

www.fundacioncnse.org

C/ Islas Aleutianas, 28. 28035 Madrid. Tel: 91 376 85 60. Fax: 91 376 85 64
e-mail: invesmat@fundacioncnse.org



Área de Investigación
en LSE

16. Sintaxis correcta. Buen uso del espacio y los clasificadores.

SEMÁNTICA / LÉXICO

3. En la configuración de coche, los faros *están* en la yema de los dedos y no mirando hacia la muñeca (como parece que muestra el avatar en esta pregunta concreta)

La configuración del clasificador es un poco *floja* (los dedos deben estar completamente juntos, incluido el pulgar)

CNM. Expresión facial: es adecuada, pero un poco exagerada (no debe salir tanto la lengua)

3, 8, 12, 14, 15. TRES: La orientación de la palma de la mano al signar TRES en estos casos es hacia el signante (y no hacia el espectador, como signa el avatar)

7. Cuando expresa que está entre medias de la casa y el edificio, el avatar signa señalando con el dedo índice hacia el suelo. Lo correcto en LSE es utilizar el clasificador de "persona" (señalando con el dedo índice hacia arriba) y colocarlo en el medio (en este caso, al lado del clasificador de casa que ha quedado a la izquierda como referente durante toda la frase).

MORFOLOGÍA

1. **LIBRO:** El signo es correcto, pero quizás abre demasiado las manos (en un ángulo de 180º) mejor abrirlo menos (130º).

La configuración del clasificador de libro está en poco floja, debería estar más rígida.

2. **YO:** Ídem.

PASEAR: La extensión y duración del movimiento es un poco exagerada. La extensión correcta es más corta, y la repetición del movimiento es circular y más rápida".

4. **NARANJA:** el contacto debe ser con la yema de los dedos de ambas manos, simultáneamente.

Inscrita en el Registro de Educación, Cultura y Deportes. I.E.O.E. 292.12.081. Fundación Docencia Asociada a la Ley 30/1994. N.I.F. G. 89115274.

www.fundacioncnse.org

C/ Islas Aleutianas, 28. 28035 Madrid. Tel: 91 376 85 60. Fax: 91 376 85 64
e-mail: invesmat@fundacioncnse.org



Área de Investigación
en LSE

5. EDIFICIO: Un poco "lento y largo", pero bien, se entiende.

6. EDIFICIO: Ídem.

7. YO: Ídem.

EDIFICIO: Ídem.

8. En la configuración del **clasificador de "coche"**, los faros *están* en la yema de los dedos y no mirando hacia la muñeca (como parece que muestra el avatar en esta pregunta concreta)

9. La configuración del **clasificador de "libro"** no es tan abierta, es más cerrada (como cuando agarras un libro)

La configuración del **clasificador de "libro caído"** es un poco *floja* (los dedos deben estar completamente juntos y rígidos, incluido el pulgar)

YO: Ídem.

10. EDIFICIO: Ídem.

CORRER: Los brazos deben estar más pegados al tronco y no hay que subir los hombros (como se corre de manera natural, quizás moviendo ligerísimamente el tronco, si se pudiera)

YO: Ídem.

11. LIBRO: Ídem.

12. LIBRO: Ídem.

13. EDIFICIO: Ídem.

En la configuración del signo MÍO, el pulgar debe verse (estar extendido, no pegado al resto de los dedos)

14. La configuración del **clasificador "coche"** es un poco *floja* (los dedos deben estar completamente juntos, incluido el pulgar)

15. LIBRO: Ídem.

16. EDIFICIO: Ídem.

YO: Ídem.

CORRER: Ídem.



Área de Investigación
en LSE

EXPERIMENTO 11. Reconocimiento de signos aislados [20 preguntas – abiertas -]

El enunciado del experimento contiene un error conceptual, dice "Instrucciones: Escriba el signo." La LSE no tiene sistema de escritura consensuado y difundido todavía. Como propuesta, diremos que sería más correcto poner "Instrucciones: Escriba en castellano el signo que realiza el avatar", o similar.

Al ser preguntas abiertas requiere un poco más de esfuerzo de comprensión por parte del usuario.

Vemos positivo que escribir algo no sea un "valor requerido" (ya que así, si el usuario no entiende lo que es, no tiene por qué poner nada como respuesta)

SEMÁNTICA

2. AGUA/BEBER: En este caso concreto, mientras que en castellano las palabras agua y beber son diferentes, en LSE el signo es exactamente el mismo, así que responder tanto AGUA como BEBER sería correcto.

3. HERMANO/IGUAL: En LSE los signos HERMANO e IGUAL se signan exactamente igual (configuración, movimiento y localización) el único rasgo diferenciador es el CNM de la oralización. En el caso de HERMANO el sería "emano" y en el de IGUAL "al"

MORFOLOGÍA

1. FELIZ: El movimiento es muy exagerado (abierto), debe ser más pequeño y cerrado. El CNM de la expresión facial es correcto, pero exagerado.

2. AGUA/BEBER: Signo correcto.

3. HERMANO/IGUAL: Signo correcto.

4. PERO: El signo es correcto, excepto que los pulgares deben estar extendidos, no pegados al resto de los dedos. La expresión facial es exagerada.

5. TIEMPO/HORA: Ídem.

Inscrita en el Registro de Educación, Cultura y Deporte. I.E.O.E. 29-12-98. Fundación Docentes Asociada a la Ley 30/1994. NIF. G-89115274.

www.fundacioncnse.org

C/ Islas Aleutianas, 28. 28035 Madrid. Tel: 91 376 85 60. Fax: 91 376 85 64
e-mail: invesmat@fundacioncnse.org



Área de Investigación
en LSE

6. VENIR: El signo es correcto, pero no se ve bien la configuración. Para solucionarlo, podría girarse ligeramente la orientación de la mano hacia el espectador.

7. COCHE: Signo correcto.

8. LIBRO: Ídem.

9. PASEAR: Ídem.

10. LEER: El signo es correcto. El CNM de mover la cabeza es correcta, pero la visión está mal. Los ojos deben mirar a la mano pasiva, pero no quedarse en blanco como hace el avatar (da un aspecto un poco psicodélico).

11. ABRIGO: Le falta flexibilidad y velocidad al movimiento (cuesta entenderlo)

12. MÉDICO/ENFERMO: La configuración de la mano activa es igual que como lo hace el avatar, solo que cambiando la posición del dedo índice por el anular. La configuración de la mano pasiva es un puño cerrado, sin movimiento.

Ambos signos son iguales (configuración, localización y movimiento) el rasgo diferenciador es el CNM de oralización de cada uno; mientras para MÉDICO la oralización es "med", para enfermo la oralización es "enf"

13. YO: Ídem.

14. DÍA/POR-LA-MAÑANA: La configuración de la mano dominante debe tener el pulgar separado del resto de dedos.

15. CASA: Signo correcto.

16. SORDO: El movimiento de este signo es tocando primero la oreja y luego la barbilla, y no arrastrando de la oreja a la barbilla, como signa el avatar.

17. ABANICO: El movimiento debe realizarse en la muñeca y no en el codo, como signa el avatar.

18. CUADRO: El movimiento no es desplazado, sino fijo.
El ángulo entre el dedo pulgar y la mano debe ser de 90°.

La oralización del avatar no se entiende muy claramente [la oralización de CUADRO es "cuado"]

19. SEMANA: Signo correcto.

20. PASADO. El avatar da dos golpes, la *pronunciación* correcta del signo en este caso es un golpe. El pulgar debe estar extendido, no pegado al resto de los dedos.



Área de Investigación
en LSE

EXPERIMENTO 12. Reconocimiento de signos aislados **[20 preguntas]**

El enunciado del experimento contiene un error conceptual, dice "Instrucciones: escoja el signo adecuado", cuando en realidad lo que hay que escoger es un PALABRA adecuada (no hay varios signos donde elegir, solo uno; el que el avatar nos muestra en el vídeo.)

SEMÁNTICA/LÉXICO

8. Interesante. CURIOSO (cosas): En las respuestas, debería sustituirse "interesante" por "curioso (cosas)", ya que el signo que propone el avatar se corresponde más con esta descripción en castellano.

MORFOLOGÍA

1. NUEVE: Correcto. Los números genéricos cardinales, como en este caso, se muestran con la palma orientada hacia el signante.

2. METRO: La configuración de la mano dominante es correcta, pero no se aprecia bien. Para solucionarlo, el movimiento de la mano dominante podría ser, en vez de completamente horizontal, un poco diagonal (sin exagerar) hacia abajo.

3. HOY: Ídem.

4. CRECER: Signo correcto. Puede añadirse un componente no manual; el patrón labial "PPP" [como si el signante pronunciase "pepepe"]

En cuanto a las palabras de las respuestas, al "crecer" debería añadirse "crecer (persona)", ya que a la palabra castellana crecer, le corresponden tres signos en LSE (persona, animal y planta)

5. ÉL/ELLA: El signo es correcto, pero debe añadirse un componente no manual; la oralización "él/ella" (según dependa, obviamente).

6. LÍNEA: Signo correcto.

7. GRANDE: Signo correcto. Componente no manual (oralización) correcto.

8. CURIOSO (cosas): Signo correcto. Componente no manual (oralización) correcto.

9. ARREGLAR: Ídem.



Área de Investigación
en LSE

10. GORDO: Signo correcto. La configuración de las manos es correcta, pero no se aprecia bien. Componente no manual; las mejillas levemente hinchadas son correctas, pero las cejas hacia abajo no (en este contexto, solo se ponen hacia abajo para expresar más cantidad - "está muy gordo", que debería acompañarse de mejillas más hinchadas - Como no es el caso, las cejas deben estar en forma normal/neutra.)

11. AGUA/BEBER: En este caso concreto, mientras que en castellano las palabras agua y beber son diferentes, en LSE el signo es exactamente el mismo, así que tanto la respuesta "a" como la "b" serían correctas. Proponemos, pues, dejar la del objetivo de la pregunta y sustituir la otra por una palabra diferente.

12. TÚ: Signo correcto.

13. CUATRO: Incorrecto. Los números genéricos cardinales, como en este caso, se muestran con la palma orientada hacia el signante, y el avatar tiene la palma orientada hacia nosotros.

14. AQUÍ: Signo correcto.

15. HERMANO: El movimiento es un poco extraño, en teoría los dedos índice y corazón se separan y vuelven a entrar en contacto un par de veces (no se frotan entre ellos, como hace el avatar)

16. ATARDECER: Debe eliminarse el signo TARDE, y dejar solo el signo ATARDECER (que es como el inicio de la pronunciación del signo NOCHE, pero más suave y corto)

17. TODO-EL-DÍA: Signo correcto.

18. CORRER: Ídem.

19. VER: Signo correcto, excepto dos detalles, el primero que la palma de la mano debe mirar ligerísimamente hacia abajo. El segundo que la distancia que alcanza la mano con respecto a la cara no tiene que ser tan alejada.

20. MÍO: El signo es correcto, excepto que el pulgar debe estar extendido, no pegado al resto de los dedos.



Área de Investigación
en LSE

EXPERIMENTO 13. Frases en LS [10 preguntas]

< Instrucciones: Escoja la frase correcta según lo que diga el avatar.> Podría ponerse <Instrucciones: Escoja la frase correcta según lo que signe el avatar.>

SINTAXIS

1. Frase correcta.

2. < SEMANA PASADA YO VER CUADRO > el orden sintáctico no es correcto. Una de las posibles formas sería "SEMANA PASADA YO CUADRO VER".

3. Frase correcta.

4. Frase correcta.

5. < YO LEER PASADO LIBRO. LIBRO CURIOSO > El orden sintáctico no es correcto, una de las posibles formas sería < AYER LIBRO YO LEER / ESE CURIOSO (cosa) > o < YO LIBRO LEER YA / CURIOSO (cosa) >

6. < COCHE TUYO CORRER IGUAL MÍO > El orden sintáctico es incorrecto. Una de las posibles formas sería < LOS-DOS(tú y yo) COCHE RÁPIDO IGUAL >

7. < TÚ BEBER ZUMO NARANJA > El orden sintáctico es incorrecto, gramaticalmente es castellano y no LSE. Una de las posibles formas sería <ZUMO NARANJA TÚ BEBER > Al ser una frase enunciativa afirmativa, debería llevar el componente no manual de asentir levemente con la cabeza.

8. < EL TRABAJAR MÉDICO > El orden sintáctico es incorrecto - es Castellano y no LSE -. En LSE bastaría con poner < EL MÉDICO >

9. < YO FELIZ MOTIVO PASEAR CON TÚ > Proponemos < YO FELIZ MOTIVO LOS-DOS(tú y yo) JUNTOS PASEAR>

Inscrita en el Registro del Ministerio de Educación, Cultura y Deporte. I.E.O.E. 29-12-98. Fundación. Documento registrado en la Ley 30/1994. N.I.F. G. 98115274.

www.fundacioncnse.org

C/ Islas Aleutianas, 28. 28035 Madrid. Tel: 91 376 85 60. Fax: 91 376 85 64
e-mail: invesmat@fundacioncnse.org



Área de Investigación
en LSE

10. < PERSONA LIBRE ALLÍ /MÍO HERMANO > Proponemos una frase más correcta en LSE < SENTAR ESE? / YO HERMANO >

SEMÁNTICA / LÉXICO

3. 25: La palma de la mano al hacer 2 y 5 debe mirar hacia el signante (y no hacia nosotros, como signa el avatar)

6. CORRER: La palabra correr en castellano sirve tanto para objetos inanimados como para seres vivos, en el caso de la LSE depende de cómo, quién o qué, de qué manera corra para que se signe de una forma u otra. En este caso, el concepto acertado en LSE sería el signo glosado como "RÁPIDO" (diferente, pues, al que signa el avatar).

IGUAL: Concepto correcto. (Es relevante, ya que hay muchas maneras de signar el "igual" castellano)

6 y 10. MÍO: El concepto de posesivo en castellano "mío" no se corresponde en LSE con un solo signo, sino con dos; MÍO y YO. Utilizándose normalmente este segundo (por ejemplo, para decir "mi hermano", se signa "YO HERMANO")

MORFOLOGÍA

1. YO: Ídem.

NOMBRE: La configuración y localización es correcta, pero el movimiento es lento.

dIFERNANDO: Dactilológico correcto, excepto la "N" (y, por similitud suponemos que "M" y "Ñ") quedan muy exageradas al mover todo el brazo para realizarlas. Con mover solo la muñeca basta.

2. YO: Ídem.

PASADO: Ídem.

CUADRO: Ídem.

3. EDAD: La localización exacta de este signo es tocando la barbilla con la yema de los dedos.

El avatar da cuatro golpes en la barbilla, la *pronunciación* correcta de este signo son dos golpes.

YO: Ídem.

www.fundacioncnse.org

C/ Islas Aleutianas, 28. 28035 Madrid. Tel: 91 376 85 60. Fax: 91 376 85 64
e-mail: invesmat@fundacioncnse.org



Área de Investigación
en LSE

4. YO: Ídem.

ABRIGO: Ídem.

PERO: Ídem.

5. PASADA: Ídem.

YO: Ídem.

LIBRO: Ídem.

Los componentes no manuales de la cara; mientras "interesante" está bien, en la acción de "leí" es exagerada, mueve exagerada y bruscamente la cabeza.

6. En cuanto a las frases propuestas, debería sustituirse "interesante" por "curioso (cosas)", ya que el signo que propone el avatar se corresponde más con esta descripción en castellano.

CORRER: Ídem (aunque el concepto de correr para coches no es este. El concepto correcto sería RÁPIDO.)

TU: El signo es incorrecto. No se entiende. Tiene otra una configuración y movimiento diferente al signado por el avatar.

MÍO: El signo es correcto, excepto que el pulgar debe estar extendido, no pegado al resto de los dedos.

Anécdota; Los signos COCHE y CORRER se parecen mucho en el avatar (no tanto en la LSE producida por un signante humano).

7. NARANJA: Ídem.

8. MÉDICO: Ídem. Falta el CNM de oralización de MÉDICO es "med" (si no, podríamos entender que es ENFERMO, aunque en este caso el contexto no ayuda a entender el significado del signo).

9. YO: Ídem.

FELIZ: Ídem.

PASEAR: Ídem.

10. SENTAR: El movimiento es de muñeca y es opuesto al que realiza el avatar (en vez de "para arriba", como lo signa el avatar, es "para abajo") [de hecho, parece que signa LIBRE en vez de SENTAR]

HERMANO: Ídem.



Área de Investigación
en LSE

Problemas técnicos

Hemos entrado varias veces en distintos días a revisar los experimentos y la verdad es que nunca hemos encontrados problemas, excepto el 23 de junio de 2009 a las 13:00, con las siguientes incidencias:

"video not found: experiment/137.flv" [Test 10, pregunta 11] y finalmente no se pudo mostrar la página:

< No se puede mostrar la página. Hay demasiadas personas que intentan tener acceso al sitio Web en este momento.

Pruebe lo siguiente:

- Haga clic en el botón [Actualizar](#) o inténtelo de nuevo más tarde.
- Abra la página principal galadriel.ii.uam.es y busque vínculos a la información que desee.

HTTP 403.9 - Access Forbidden: Too many users are connected
Servicios de Internet Information Server

Información técnica (para personal de soporte)

- Información general: Este error puede producirse si el servidor Web está ocupado y no puede procesar su petición porque hay mucho tráfico.
- Más información: [Soporte técnico de Microsoft](#) >

Realizado por el Área de Investigación en LSE de la Fundación de la Confederación Estatal de Personas Sordas. 25 Junio 200

www.fundacioncnse.org

C/ Islas Aleutianas, 28. 28035 Madrid. Tel: 91 376 85 60. Fax: 91 376 85 64
e-mail: invesmat@fundacioncnse.org

Bibliografía

- [1] **N. Adamo-Villani y G. Beni:** *Sign language subtitling*, en: ACM SIGGRAPH Educators program, pag. 38 (ACM, New York, NY, USA 2005).
- [2] **B. B. Agravat:** *Inputted text to animated sign language, interactive interface, a self-learning with fun*, en: ACM SIGGRAPH educators program, pag. 9 (ACM, New York, NY, USA 2007).
- [3] **A. S. Ahmed y D. S. K. Seong:** *SignWriting on mobile phones for the deaf*, en: Proceedings of the 3rd international conference on Mobile Technology, Applications & Systems, pag. 28 (ACM, New York, NY, USA 2006).
- [4] **Autodesk Inc.:** *3ds Max 8*, www.autodesk.com/3dsmax, 2005.
- [5] **N. Badler, R. Bindiganavale, J. Granieri, S. Wei y Z. Zhao:** *Posture interpolation with collision avoidance*, en: Proceedings of Computer Animation, pags. 13–20 (1994).
- [6] **N. I. Badler, R. Bindiganavale, J. Allbeck, W. Schuler, L. Zhao, S. joo Lee, H. Shin y M. Palmer:** *Parameterized action representation and natural language instructions for dynamic behavior modification of embodied agents*, en: AAAI Spring Symposium (2000).
- [7] **B. Baham:** *Non-manual realization of agreement in American Sign Language*, Tesis Doctoral, Boston University, 1996.
- [8] **A. Bangham, S. Cox, R. Elliot, J. Glauert y I. Marshall:** *Virtual Signing: Capture, Animation, Storage and Transmission - an Overview of the ViSiCAST Project*, en: IEE Seminary on Speech and Language Processing for Disabled and Elderly People (2000).
- [9] **R. Battison:** *Toward a phonological theory of sign language*, en: Communication Center, NTID-RIT (1973).

- [10] **J. Beskow**: *Rule-Based Visual Speech Synthesis*, en: Proceedings of Eurospeech, pags. 299–302 (1995).
- [11] **J. Beskow**: *Trainable Articulatory Control Models for Visual Speech Synthesis*, en: International Journal of Speech Technology, volumen 7, n° 4 pags. 335–349, 2004.
- [12] **J. Beskow, I. Karlsson, J. Kewley y G. Salvi**: *SYNFACE - A Talking Head Telephone for the Hearing-Impaired*, en: Proceedings of ICCHP, pags. 1178–1186 (Paris, France 2004).
- [13] **R. Bindiganavale, W. Schuler, J. M. Allbeck, N. I. Badler, A. K. Joshi y M. Palmer**: *Dynamically altering agent behaviors using natural language instructions*, en: Proceedings of the 4th international conference on Autonomous agents, pags. 293–300 (ACM, New York, NY, USA 2000).
- [14] **F. Buttussi, L. Chittaro y D. Nadalutti**: *H-animator: a visual tool for modeling, reuse and sharing of X3D humanoid animations*, en: Proceedings of the 11th international conference on 3D web technology, pags. 109–117 (ACM, New York, NY, USA 2006).
- [15] **G. Caridakis y K. Karpouzis**: *Design and Implementation of a Greek Sign Language Synthesis System*, en: WSEAS Transactions on Systems, volumen 3, n° 10 pags. 3108–3113, 2004.
- [16] **G. Caridakis, K. Karpouzis, G. Sapountzaki, S.-E. Fotinea y E. Efthimiou**: *A dynamic environment for Greek Sign Language Synthesis using virtual characters*, en: Proceedings of ACM Web3D 2005 Symposium (March 29 - April 1 2005, Bangor, UK. 2005).
- [17] **N. Cherniavsky, A. C. Cavender, R. E. Ladner y E. A. Riskin**: *Variable frame rate for low power mobile sign language communication*, en: Proceedings of the 9th international ACM SIGACCESS conference on Computers and Accessibility, pags. 163–170 (ACM, New York, NY, USA 2007).
- [18] **L. Chittaro, F. Buttussi y D. Nadalutti**: *MAGE-Anim: a system for visual modeling of embodied agent animations and their replay on mobile devices*, en: Proceedings of the working conference on Advanced Visual Interfaces, pags. 344–351 (ACM, New York, NY, USA 2006).
- [19] **S. Clay y J. Wilhelms**: *Put: language-based interactive manipulation of objects*, en: IEEE Computer Graphics and Applications, volumen 16, n° 2 pags. 31–39, 1996.

- [20] **D. Cogill-Koez**: *A model of signed language 'classifier predicates' as templated visual representation*, en: Sign Language and Linguistics, volumen 3, n° 2 pags. 209–236, 2000.
- [21] **R. Cole, S. Van Vuuren, B. Pellom, K. Hacioglu, J. Ma, J. Movellan, S. Schwartz, D. Wade-Stein, W. Ward y J. Yan**: *Perceptive animated interfaces: first steps toward a new paradigm for human-computer interaction*, en: Proceedings of the IEEE, volumen 91, n° 9 pags. 1391–1405, 2003.
- [22] **S. Cox, M. Lincoln, M. Nakisa, M. Wells, M. Tutt y S. Abbott**: *The Development and Evaluation of a Speech-to-Sign Translation System to Assist Transactions*, en: International Journal of Human-Computer Interaction, volumen 16, n° 2 pags. 141–161, 2003.
- [23] **S. Cox, M. Lincoln, J. Tryggvason, M. Nakisa, M. Wells, M. Tutt y S. Abbott**: *Tessa, a system to aid communication with deaf people*, en: Proceedings of the 5th international ACM conference on Assistive Technologies, pags. 205–212 (ACM, New York, NY, USA 2002).
- [24] **B. Coyne y R. Sproat**: *WordsEye: an automatic text-to-scene conversion system*, en: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pags. 487–496 (ACM, New York, NY, USA 2001).
- [25] **O. Crasborn, H. van der Hulst y E. van der Kooij**: *SignPhon: A database tool for cross-linguistic phonological analysis of sign languages.*, en: Proceedings of the 6th International Conference on Theoretical Issues in Sign Language Research (1998).
- [26] **O. Crasborn, H. van der Hulst y E. van der Kooij**: *SignPhon: A phonological database for sign languages*, en: Sign Language & Linguistics, volumen 4, n°1-2 pags. 215–228, 2001.
- [27] **R. Elliot, J. Glauert, V. Jennings y R. Kennaway**: *An Overview of the SiGML Notation and SiGMLSigning Software System*, en: Proceedings of LREC (Lisbon 2004).
- [28] **R. Elliott, J. Glauert, R. Kennaway y I. Marshall**: *The development of language processing support for the ViSiCAST project*, en: Proceedings of the 4th international ACM conference on Assistive technologies, pags. 101–108 (ACM, New York, NY, USA 2000).
- [29] **R. Elliott, J. Glauert, R. Kennaway y K. Parsons**: *ViSiCAST Deliverable D5-2: SiGML Definition*, Documento técnico, ViSiCAST Project, 2001.

- [30] **E. Engberg-Pedersen**: Language, Gesture, and Space, capítulo 7: Point of view expressed through shifters, pags. 133–154. *Editado por K. Emmorey y J. Reilly* (Lawrence Erlbaum Associates 1995), 1995.
- [31] **Esmertec**: *Esmertec Jbed*, <http://www.myriadgroup.com/>, 2009.
- [32] **European Media Masters of Art**: *VSIGN*, <http://www.vsign.nl/>, 2002.
- [33] **P. Fabian y J. Francik**: *Synthesis and Presentation of the Polish Sign Language Gestures*, en: Proceedings of the 1st International Conference on Applied Mathematics and Informatics at Universities, pags. 190–197 (Cabcikovo, Slovakia 2001).
- [34] **F. Fernández Villalibre**: Diccionario usual de mímica y dactilología: útil a los maestros de sordo-mudos, a sus padres y a todas las personas que tengan que entrar en comunicación con ellos (Imprenta del Colegio de Sordo-mudos y Ciegos, Madrid 1851).
- [35] **J. Francik y P. Fabian**: *Animating Sign Language in the Real Time*, en: Proceedings of 20th IASTED International Multi-Conference Applied Informatics, pags. 276–281 (2002).
- [36] **Fundación CNSE**: Diccionario Normativo de la Lengua de Signos Española (Fundación CNSE 2008).
- [37] **J. Furst, K. Alkoby, A. Berthiaume, P. Chomwong, M. Jo, B. Konie, G. Lancaster, S. Lytinen, J. Mcdonald, J. Toro, N. Tomuro y R. Wolfe**: *Database design for American Sign Language.*, en: Proceedings of 15th International Conference on Computers and their Applications, pags. 427–430 (2000).
- [38] **Gobierno de España**: *Ley 51/2003, de 2 de diciembre, de igualdad de oportunidades, no discriminación y accesibilidad universal de las personas con discapacidad*, B.O.E. 3 Diciembre, 2003, (Spanish Law).
- [39] **Gobierno de España**: *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.*, B.O.E. 24 Octubre, 2007, (Spanish Law).
- [40] **A. B. Grieve-Smith**: *SignSynth: A Sign Language Synthesis Application Using Web3D and Perl*, en: Revised Papers from the International Gesture Workshop on Gesture and Sign Languages in Human-Computer Interaction, pags. 134–145 (Springer-Verlag, London, UK 2002).

- [41] **S. Hamilton**: Lectures on Quaternions (Royal Irish Academy 1853).
- [42] **T. Hanke**: *ViSiCAST Deliverable D5-1: Interface Definitions*, Documento técnico, ViSiCAST Project, 2000.
- [43] **T. Hanke**: *HamNoSys - Representing Sign Language Data in Language Resources and Language Processing Contexts*, en: Proceedings of LREC, Editado por S. B. . Heidelberg (Lisbon 2004).
- [44] **Á. Herrero Blanco**: *Escritura Alfabética de Lengua de Signos Española* (Universidad de Alicante 2003).
- [45] **Á. Herrero Blanco**: *A practical writing of sign languages*, en: Proceedings of LREC (Lisbon 2004).
- [46] **Á. Herrero Blanco**: *Una aproximación morfológica a las construcciones clasificatorias en la lengua de signos española*, en: ELUA. Estudios de Lingüística, volumen 18 pags. 151–167, 2004.
- [47] **Z. Huang, A. Eliëns y C. Visser**: *STEP: A Scripting Language for Embodied Agents*, en: Proceedings of the Workshop on Lifelike Animated Agents (Tokyo 2002).
- [48] **M. Huenerfauth**: *Spatial Representation of Classifier Predicates for Machine Translation into American Sign Language*, en: Workshop on the Representation and Processing of Signed Languages, 4th International Conference on Language Resources and Evaluation (2004).
- [49] **M. Huenerfauth**: *Generating american sign language classifier predicates for english-to-asl machine translation*, Tesis Doctoral, Computer and Information Science, University of Pennsylvania, Philadelphia, PA, USA, 2006, adviser Mitch Marcus and Martha Palmer.
- [50] **M. Huenerfauth, L. Zhao, E. Gu y J. Allbeck**: *Evaluating American Sign Language generation through the participation of native ASL signers*, en: Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility, pags. 211–218 (ACM, New York, NY, USA 2007).
- [51] **M. Huenerfauth, L. Zhao, E. Gu y J. Allbeck**: *Evaluation of American Sign Language Generation by Native ASL Signers*, en: ACM Transactions on Accessible Computing, volumen 1, n^o 1 pags. 1–27, 2008.
- [52] **Hybrid**: *Hybrid Rasteroid*, <http://www.hybrid.fi/>, 2006.

- [53] **Informatics and Telematics Institute:** *VSIGNS*, <http://vsigns.iti.gr:8080/VSigns>, 2004.
- [54] **A. Irving y R. Foulds:** *A parametric approach to sign language synthesis*, en: Proceedings of the 7th international ACM SIGACCESS conference on Computers and Accessibility, págs. 212–213 (ACM, New York, NY, USA 2005).
- [55] **ISO/IEC 14496-11:** Information technology – Coding of audio-visual objects – Part 11: Scene description and application engine (International Organization for Standardization, Geneva, Switzerland 2005).
- [56] **ISO/IEC 14772-1:1997:** Information technology – Computer graphics and image processing – The Virtual Reality Modeling Language (VRML) – Part 1: Functional specification and UTF-8 encoding (International Organization for Standardization, Geneva, Switzerland 1997).
- [57] **ISO/IEC 14772-2:2004:** Information technology – Computer graphics and image processing – The Virtual Reality Modeling Language (VRML) – Part 2: External authoring interface (EAI) (International Organization for Standardization, Geneva, Switzerland 2004).
- [58] **ISO/IEC 19774:2006:** Information technology – Computer graphics and image processing – Humanoid animation (H-Anim) (International Organization for Standardization, Geneva, Switzerland 2006).
- [59] **Java Community Process:** *JSR-184. Mobile 3D Graphics API for J2ME*, <http://www.jcp.org/en/jsr/detail?id=184>, 2005.
- [60] **Java Community Process:** *JSR 297: Mobile 3D Graphics API 2.0*, <http://www.jcp.org/en/jsr/detail?id=184>, 2005.
- [61] **R. Kennaway:** *Synthetic Animation of Deaf Signing Gestures*, en: LNCS, volumen 2298 págs. 146–157, 2002.
- [62] **R. Kennaway, R. Elliot, J. Glauert y K. J. Parsons:** *SiGML Document Type Definition (DTD)*, Visicast deliverable, ViSiCAST Project, 2002.
- [63] **R. Kennaway, J. Glauert y I. Zwitserlood:** *Providing signed content on the Internet by synthesized animation*, en: ACM Transactions on Computer-Human Interaction, volumen 14, n^o 15 págs. 1–29, 2007.
- [64] **D. Klatt:** *Software for a cascade/parallel formant synthesizer*, en: Journal of the Acoustical Society of America, volumen 67, n^o 3 págs. 971–995, 1980.

- [65] **T. Lebourque y S. Gibet:** *A Complete System for the Specification and the Generation of Sign Language Gestures*, en: Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction, págs. 227–238 (Springer-Verlag, London, UK 1999).
- [66] **T. Lebourque y S. Gibet:** *High Level Specification and Control of Communication Gestures: The GESSYCA System*, en: Proceedings of the Computer Animation, págs. 24–35 (IEEE Computer Society, Washington, DC, USA 1999).
- [67] **P. K. Levangie y C. C. Norkin:** Joint Structure and Function: A Comprehensive Analysis, 3 edición (F.A. Davis Company 2001).
- [68] **S. K. Liddell:** *Spatial representations in discourse: Comparing spoken and signed language*, en: Lingua, volumen 98, n° 1-3 págs. 145–167, 1996.
- [69] **S. K. Liddell:** Perspectives on Classifier Constructions in Sign Languages, capítulo 9: Sources of Meaning in ASL Classifier Predicates, págs. 199–220. *Editado por K. Emmorey* (Psychology Press 2003), 2003.
- [70] **S. K. Liddell y M. Metzger:** *Gesture in sign language discourse*, en: Journal of Pragmatics, volumen 30, n° 6 págs. 657–697, 1998.
- [71] **D. Lillo-Martin:** Language, Gesture, and Space, capítulo 8: The Point of View Predicate in American Sign Languages, págs. 155–170. *Editado por K. Emmorey y J. Reilly* (Lawrence Erlbaum Associates 1995), 1995.
- [72] **J. Loomis, H. Poizner, U. Bellugi, A. Blakemore y J. Hollerbach:** *Computer graphic modeling of american sign language*, en: ACM SIGGRAPH Computer Graphics, volumen 17, n° 3 págs. 105–114, 1983.
- [73] **S. Lu, S. Igi, H. Matsuo y Y. Nagashima:** *Towards a Dialogue System Based on Recognition and Synthesis of Japanese Sign Language*, en: Proceedings of the International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction, págs. 259–271 (Springer-Verlag, London, UK 1998).
- [74] **J. Ma, R. Cole, B. Pellom, W. Ward y B. Wise:** *Accurate visible speech synthesis based on concatenating variable length motion capture data*, en: IEEE Transactions on Visualization and Computer Graphics, volumen 12, n° 2 págs. 266–276, 2006.
- [75] **O. Martínez Lazalde, S. Maddock y M. Meredith:** *A Mexican-Spanish talking head*, en: Proceedings of the 3rd International Conference on Games Research and Development, págs. 17–24 (2007).

- [76] **O. Martínez Lazalde, S. Maddock y M. Meredith:** *A constraint-based approach to visual speech for a Mexican-Spanish talking head*, en: International Journal of Computer Games Technology, volumen 2008, n° 3 pags. 1–7, 2008.
- [77] **J. McDonald, J. Toro, K. Alkoby, A. Berthiaume, R. Carter, P. Chomwong, J. Christopher, M. J. Davidson, J. Furst, B. Konie, G. Lancaster, L. Roychoudhuri, E. Sedgwick, N. Tomuro y R. Wolfe:** *An Improved Articulated Model of the Human Hand*, en: Proceedings of the 8th International Conference in Central Europe on Computer Graphics, Visualization and Interactive Digital, pags. 306–313 (2000).
- [78] **Microsoft:** *XNA Game Studio*, <http://www.xna.com/>, 2006.
- [79] **A. Moreno, D. Poch, A. Bonafonte, E. Lleida, J. Llisterri, J. B. Mariño y C. Nadeu:** *Albayzin Speech Database: Design of the Phonetic Corpus*, en: Proceedings of Eurospeech, volumen 1, pags. 653–656 (1993).
- [80] **I. M. Muñoz:** Lenguaje de Signos, Tesis Doctoral, University of Valladolid, 1990.
- [81] **I. M. Muñoz:** *La querología aplicada a la enseñanza de la LSE*, en: Revista Española de Lingüística de las Lenguas de signos, volumen 1, n° 2 pags. 1–18, 1997.
- [82] **C. Neidle, S. Sclaroff y V. Athitsos:** *SignStreamTM: A Tool for Linguistic and Computer Vision Research on Visual-Gestural Language Data.*, en: Behavior Research Methods, Instruments, & Computers, volumen 3 pags. 311–320, 2001.
- [83] **M. Papadogiorgaki, N. Grammalidis, L. Makris y M. Strintzis:** *Gesture synthesis from sign language notation using MPEG-4 humanoid animation parameters and inverse kinematics*, en: Proceedings of the International Conference on Intelligent Environments, volumen 1, pags. 151–160 (2006).
- [84] **M. Papadogiorgaki, N. Grammalidis, N. Sarris y M. Strintzis:** *Synthesis of Virtual Reality Animation from SWML using MPEG-4 Body Animation Parameters*, en: Proceedings of LREC (Lisbon 2004).
- [85] **L. Piegl:** *On NURBS: a survey*, en: IEEE Computer Graphics and Applications, volumen 11, n° 1 pags. 55–71, 1991.
- [86] **S. Prillwitz, R. Leven, H. Zienert, T. Hanke y J. Herming:** HamNoSys. Version 2.0; Hamburg Notation System for Sign Languages. An introductory guide (Signum-Verlag 1989).

- [87] **A. Rocha**: *Sign Writing Markup Language DTD (SWML DTD)*, Documento técnico, SignNet Project, 2003.
- [88] **H. Sagawa y M. Takeuchi**: *A teaching system of Japanese sign language using sign language recognition and generation*, en: Proceedings of the 10th ACM international conference on Multimedia, pags. 137–145 (ACM, New York, NY, USA 2002).
- [89] **R. San Segundo, R. Barra, R. Córdoba, L. F. D’Haro, F. Fernández, J. Ferreiros, J. M. Lucas, J. Macías-Guarasa, J. M. Montero y J. M. Pardo**: *Speech to sign language translation system for Spanish*, en: Speech Communication, volumen 50, n°11-12 pags. 1009–1020, 2008.
- [90] **R. San Segundo, J. M. Montero, J. Macías-Guarasa, R. Córdoba, J. Ferreiros y J. M. Pardo**: *Proposing a speech to gesture translation architecture for Spanish deaf people*, en: Journal of Visual Languages and Computing, volumen 19, n° 5 pags. 523–538, 2008.
- [91] **G. Sapountzaki, E. Efthimiou, K. Karpouzis y V. Kourbetis**: *Open-ended Resources in Greek Sign Language: Development of an e-Learning Platform*, en: Proceedings of LREC (Lisbon 2004).
- [92] **W. Schuler**: *Using model-theoretic semantic interpretation to guide statistical parsing and word recognition in a spoken language interface*, en: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics, pags. 529–536 (Association for Computational Linguistics, Morristown, NJ, USA 2003).
- [93] **L. M. Seversky y L. Yin**: *Real-time automatic 3D scene generation from natural language voice and text descriptions*, en: Proceedings of the 14th annual ACM international conference on Multimedia, pags. 61–64 (ACM, New York, NY, USA 2006).
- [94] **K. Shoemake**: *Animating rotation with quaternion curves*, en: Proceedings of the 12th annual conference on Computer graphics and interactive techniques, pags. 245–254 (ACM, New York, NY, USA 1985).
- [95] **Sociedad Bíblica Española**: *Traducción de la Biblia a la LSE©*, Vídeo DVD, 2007.
- [96] **F. Solina, S. Krapež, A. Jaklič y V. Komac**: Design and Management of Multimedia Information Systems, capítulo 13: Multimedia Dictionary and Synthesis of Sign Language, pags. 268–281 (Idea Group Publishing 2001), 2001.

- [97] **W. C. Stokoe**: *Sign language structure (Studies in Linguistics)*, en: Occasional Paper, volumen 8, 1960.
- [98] **W. C. Stokoe., D. C. Casterline y C. G. Croneberg**: *A Dictionary of American Sign Language on Linguistic Principles* (research publication 1965).
- [99] **S. A. Su y R. K. Furuta**: *VRML-based representations of ASL fingerspelling on the World Wide Web*, en: Proceedings of the 3rd international ACM conference on Assistive Technologies, pages. 43–45 (ACM, New York, NY, USA 1998).
- [100] **T. Supalla**: *Perspectives on Classifier Constructions in Sign Languages*, capítulo 11: *Revisiting Visual Analogy in ASL Classifier Predicates*, pages. 249–257. *Editado por K. Emmorey* (Psychology Press 2003), 2003.
- [101] **V. Sutton**: *Dance Writing*, <http://www.dancewriting.org/>, 1974.
- [102] **V. Sutton**: *Sign Writing*, <http://www.signwriting.org/>, 1974.
- [103] **B. Szczepankowski**: *Wyrównywanie szans osób niesłyszących (WSiP 1999)*.
- [104] **B. Szczepankowski y M. Rona**: *Szkolny słownik języka migowego* (Oficyna Wydawnicza Vinea 1994).
- [105] **J. Tejedor, D. Wang, J. Frankel, S. King y J. Colás**: *A comparison of grapheme and phoneme-based units for Spanish spoken term detection*, en: *Speech Communication*, volumen 50, n°11-12 pages. 980–991, 2008.
- [106] **D. Tolani, A. Goswami y N. I. Badler**: *Real-time inverse kinematics techniques for anthropomorphic limbs*, en: *Graphical Models*, volumen 62 pages. 353–388, 2000.
- [107] **X. Yang, D. Petriu, T. Whalen y E. Petriu**: *Hierarchical animation control of avatars in 3-D virtual environments*, en: *IEEE Transactions on Instrumentation and Measurement*, volumen 54, n° 3 pages. 1333–1341, 2005.
- [108] **L. van Zijl y D. Barker**: *South African Sign Language Machine Translation System*, en: Proceedings of the 2nd international conference on Computer graphics, virtual Reality, visualisation and interaction in Africa, pages. 49–52 (ACM, New York, NY, USA 2003).
- [109] **L. van Zijl y L. Raitt**: *Implementation experience with collision avoidance in signing avatars*, en: Proceedings of the 3rd international conference on Computer graphics, virtual reality, visualisation and interaction in Africa, pages. 55–59 (ACM, New York, NY, USA 2004).

- [110] **I. Zwiterslood, M. Verlinden, J. Ros y S. Schoot**: *Synthetic Signing for the Deaf: eSIGN*, en: Proceedings of the Conference and Workshop on Assistive Technologies for Vision and Hearing Impairment (CVHI, Granada, Spain 2004).