

**UNIVERSIDAD POLITÉCNICA DE MADRID
ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE
TELECOMUNICACIÓN**



ETSIT

**RESUMEN DEL PROYECTO FIN DE
CARRERA:
ENTORNO DE DESARROLLO PARA
LA CREACIÓN DE ANIMACIONES
CON AGENTES VIRTUALES 3D**

**AUTORA: Ana Huerta Carrillo
TUTOR: D. Rubén San Segundo Hernández
Dpto. Ingeniería Electrónica (IEL)
Grupo de Tecnología del Habla (GTH)
Fecha de lectura: 18 de Diciembre de 2006**

RESUMEN DEL PROYECTO FIN DE CARRERA: ENTORNO DE DESARROLLO PARA LA CREACIÓN DE ANIMACIONES CON AGENTES VIRTUALES 3D

1. ORIGEN, OBJETIVOS, DESARROLLO Y CONCLUSIONES

El objetivo de este proyecto fin de carrera, es la creación de un entorno de desarrollo para crear animaciones por medio de un avatar en tres dimensiones. Esta aplicación, forma parte de un sistema complejo de traducción de lenguaje natural a lengua de signos, cuya primera parte ya ha sido desarrollada en el Proyecto Fin de Carrera titulado '*Sistema De Traducción De Lenguaje Natural A Signos Gestuales*' (E. Ibáñez León). A continuación se describe el sistema completo, con el fin de entender la finalidad del proyecto realizado.

El sistema completo al que pertenece la aplicación diseñada, está centrado en un dominio concreto de aplicación: frases pronunciadas por un funcionario que atiende a las personas que desean obtener o renovar el DNI (Documento Nacional de Identidad) o el pasaporte. El sistema reconoce y traduce las frases pronunciadas por el funcionario a lengua de signos. El sistema está formado por 3 módulos principales: un reconocedor de voz, un módulo de traducción de lenguaje natural (traduce secuencias de palabras en secuencia de signos), y finalmente un módulo de representación de los signos mediante un agente animado en 3D. Este tercer módulo ha sido el objeto de este proyecto,

La lengua de signos presenta una gran variabilidad entre diferentes países e incluso en diferentes áreas de un mismo país. Por esta razón, desde 1960, la lengua de signos ha sido objeto de estudio en Estados Unidos, en Europa, África y Japón.

En España, durante los últimos 20 años ha habido varias propuestas para normalizar la Lengua de Signos Española (LSE), pero ninguna de ellas ha sido aceptada por las personas sordas. Desde su punto de vista, estas propuestas tienden a limitar la flexibilidad de la lengua de signos. En 1991 M.A. Rodríguez llevó a cabo un análisis detallado de las diferencias entre la lengua de signos utilizada por las personas sordomudas y las propuestas de estandarización. Este estudio constituye uno de los principales estudios de la LSE y ha sido una referencia importante en este trabajo.

1.1. Descripción del sistema

El sistema está formado por varias partes como se puede ver en la Figura 1.1. La aplicación desarrollada en este Proyecto Fin de Carrera implementa la última parte del sistema, la representación de signos procedentes de la traducción realizada a lo largo de los módulos anteriores.

1.1.1. Un reconocedor de voz.

Se encarga de traducir la voz en una secuencia de palabras. Este módulo permite reconocer habla en lenguaje natural (habla continua) e independiente del locutor. Para esta aplicación el vocabulario posible contiene 485 palabras, lo que permite conseguir tasas de aciertos por encima del 90%.

1.1.2. Un módulo de traducción automática.

Traduce la secuencia de palabras en una secuencia de signos de la LSE. Se ha utilizado una solución tecnológica que se basa en reglas de traducción elaboradas por una persona experta. Mediante este método, la traducción entre palabras y signos se realiza de forma manual: las reglas van combinando palabras y generando signos, que a su vez se pueden recombinar con otros signos o palabras para formar la secuencia completa final.

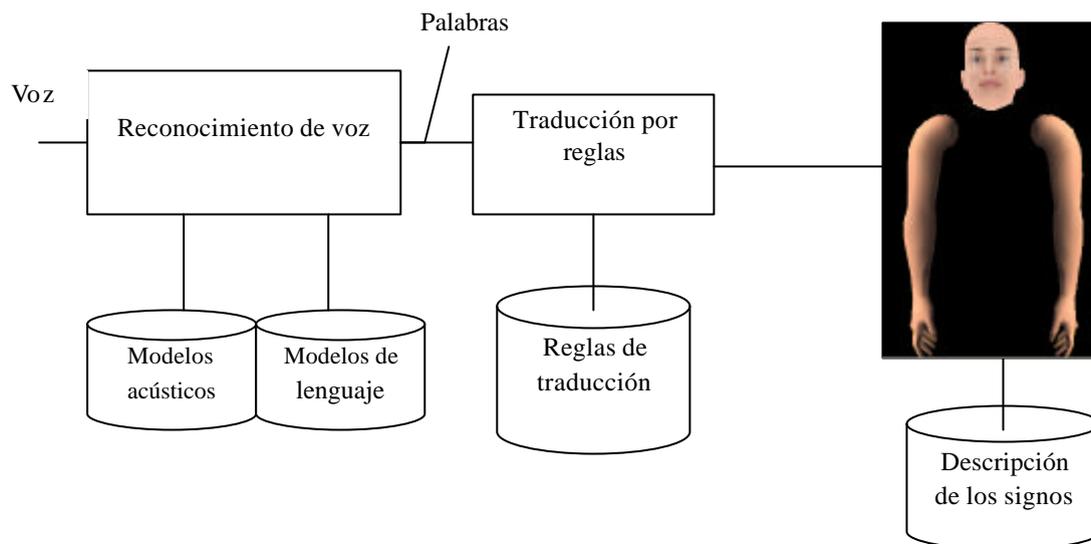


Figura 1. 1. Diagrama de módulos del sistema desarrollado.

1.1.3. Un módulo de representación de los signos.

Este módulo está basado en un agente animado en 3D. Para este módulo se dispone de un agente animado llamado ‘Sara’, que ha sido desarrollado en este Proyecto Fin de Carrera. ‘Sara’ ha sido diseñada atendiendo a las necesidades de perfeccionamiento de las partes del cuerpo determinantes a la hora de representar un signo de la lengua de signos (cabeza y brazos). Es aquí donde se integra la aplicación desarrollada en el proyecto, la cuál es capaz de general signos, posiciones y animaciones, funcionando, o bien como módulo de representación del sistema completo al que pertenece, o como sistema independiente de diseño de animaciones para cada signo, al servicio de un logopeda, con la posibilidad de incrementar el número de signos a realizar, creando bibliotecas específicas. Para ello se ha dotado al sistema gráfico de herramientas y funcionalidades de alto nivel que permiten desarrollar nuevas animaciones con cierta facilidad para el usuario.

1.2. Dominio de aplicación

Los módulos de análisis semántico y de generación de signos del sistema están diseñados para el dominio de aplicación de apoyo al servicio de solicitud/ renovación del DNI (Documento Nacional de Identidad) o pasaporte. Dicho proceso se gestiona desde el Ministerio del Interior, y por lo tanto, toda la información que se ha usado para el análisis de este proceso de solicitud y renovación del DNI se ha obtenido de la página oficial del Ministerio de Interior.

A partir de la información de dicha página se han redactado 201 frases. Se ha mantenido por lo general la estructura de las frases originales, poniendo mucho interés en generar diferentes niveles de uso, o grados de flexibilidad a la hora de realizar la traducción del español a la LSE. Si bien es cierto que, por lo general, se ha tratado de simular los hábitos de las personas sordomudas que prefieren dividir una misma oración en una serie de sentencias mucho más cortas y directas.

El objetivo del sistema es conseguir que un funcionario de la administración del Estado que no conozca en absoluto la LSE pueda explicar a una persona sorda qué necesita para poder solicitar o renovar su DNI, cuales son las utilidades de dicho documento, dónde puede viajar con él, y todo tipo de información relativa al DNI o el pasaporte.

1.3. Representación de los signos

Este módulo se centra en el desarrollo del Proyecto Fin de Carrera, basado en la creación del módulo de representación, en el cual se ha creado un avatar virtual llamado 'Sara', cuyo funcionamiento ha sido implementado por completo en el marco del proyecto.

La ventaja de utilizar el avatar 'Sara' radica en que se ha prestado especial atención a las partes más importantes a la hora de representar los signos: los brazos y la cabeza. Para ello se ha realizado un sistema basado en jerarquías de huesos, implementado sobre OpenGL. El sistema de huesos para los brazos consta de dos esqueletos (con 18 huesos cada uno), uno para cada brazo. La cabeza es un objeto en formato .3DS (3D StudioMax), que se importa como objeto aparte.

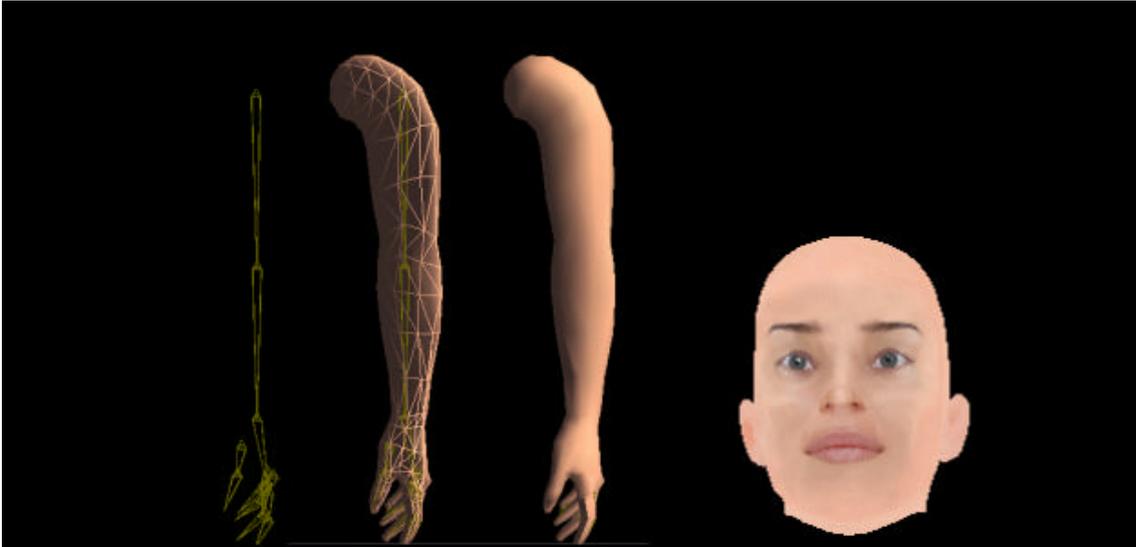


Figura 1. 2. Sistema de representación de animaciones: sistema jerárquico de huesos, con textura asociada, y cabeza del avatar 'Sara'.

1.3.1. El sistema de brazos.

El sistema de brazos consta de dos esqueletos, uno para cada brazo. Cada esqueleto (cada brazo) tendrá un hueso padre, u origen, del cual dependerán los demás. Asimismo, cada hueso tendrá un padre, que será el hueso inmediatamente superior, y uno o varios hijos. De esta forma se asegura un movimiento realista y proporcionado. Los huesos están cubiertos con una malla, a modo de piel. Dicha malla se asocia al sistema mediante un cálculo ponderado de pesos. Cada punto de la malla estará asociado a un hueso concreto (si está claro cuál es el hueso más cercano). Es decir, se tienen dos tipos de casos: puntos de la piel que deben moverse con un hueso concreto y puntos de la piel que, por estar en zonas críticas, deben moverse según el movimiento de dos huesos (zonas de unión entre huesos). Una vez definidos los esqueletos, se describe el sistema de movimiento.

Cada hueso está definido por las traslaciones tridimensionales (x,y,z) de su punto de origen (el punto de unión con el hueso padre), y las rotaciones tridimensionales. A la hora de representar un movimiento, se ha comprobado que este queda perfectamente definido por las rotaciones parciales de cada uno de los huesos únicamente. Cada vez que se mueve un hueso, se actualizan sus rotaciones, haciendo por extensión que se actualicen las rotaciones de los huesos hijos. Al final se adapta la malla (piel) a la nueva situación de los huesos.

1.3.2. La cabeza.

Es un objeto importado en el sistema en formato *.3DS. No dispone de tanta adaptabilidad como los brazos, solamente es capaz de realizar movimientos, sin expresiones. No

obstante, es suficiente para lo que se quiere implementar, ya que su principal función es acompañar o servir de referente a los movimientos de los brazos.

1.3.3. Funcionamiento del sistema.

En el procedimiento para diseñar una animación se realizan las siguientes acciones, por este orden: creación de posiciones (parciales y/o totales), creación de animaciones.

1.3.4. Creación de posiciones.

La herramienta que se va a definir a lo largo de los sucesivos capítulos ofrece la posibilidad de crear y guardar posiciones, para cargarlas posteriormente, y ‘resetear’ dichas posiciones. Se puede trabajar con posiciones del esqueleto completo, o posiciones únicamente de las manos. En ambos casos, el procedimiento es el mismo: se guardan en un archivo los datos referentes a las rotaciones de los huesos implicados (todo el esqueleto o solamente los huesos de las manos). De esta forma se pueden guardar posiciones concretas, posiciones comunes en la representación de la lengua de signos, para posteriormente combinarlas entre sí, formando la secuencia de signos deseada.

1.3.5. Creación de animaciones.

Para la creación de animaciones, se define un array de posiciones. Las animaciones se crearán como secuencias de posiciones. Para ello, se definen tres tipos de posiciones: 0, 1 y 2. Las posiciones por defecto serán de tipo 0, que significa que no han sido especificadas. Las posiciones tipo 1 son las que se establecen específicamente con una posición concreta. Las posiciones inicial y final serán siempre de tipo 1 por defecto. La creación de las posiciones intermedias se basará en la interpolación lineal, entre las posiciones definidas de tipo 1. Las posiciones obtenidas de esta interpolación se denominarán de tipo 2. La animación consistirá en la representación sucesiva de cada posición en pantalla. La creación de animaciones resulta sencilla habiendo dotado al sistema de utilidades como: guardar, cargar y ‘resetear’ posiciones, creación y borrado de interpolaciones (incluso para animaciones ya guardadas), y previsualización de resultados. Las animaciones simples pueden tener un máximo de 120 posiciones. La longitud de una animación se define según se va creando, a medida que se guardan las diferentes posiciones. La duración de un signo puede modificarse si al realizarlo queda o bien rápido o bien lento en exceso. Se dispone de dos controles: uno para duplicar el número de fotogramas y otro para reducirla a la mitad. Estas operaciones pueden realizarse tantas veces como sea preciso, siempre teniendo en cuenta que el límite superior de una animación son 120 posiciones, y que no existe límite inferior salvo el que el ojo humano pueda concluir según la calidad de la animación obtenida.

Mediante la generación de animaciones, se realiza la creación de la base de datos que permitirá representar los signos provenientes del módulo de traducción descrito anteriormente. Asimismo, se presenta la posibilidad de crear animaciones nuevas, de forma que el sistema funcione de manera independiente, al servicio de un logopeda que necesite ampliar la biblioteca de signos animados.

Para la presentación en pantalla, el sistema dispone de un generador de videos con calidad configurable, de forma que se puede guardar una animación (simple o compleja) tanto en formato de animación estándar (*.ani) como en formato de video (*.avi) que puede ser utilizado como representación en cualquier sistema, dotando así a la aplicación de mayor adaptabilidad

Para la representación de grandes secuencias de signos, se ha definido lo que se llama: animación compleja. Una animación compleja se formará concatenando animaciones con un número de posiciones variable (con un máximo de 120). Estas secuencias se unirán mediante pequeños fragmentos de animación, de un número configurable de fotogramas (10-20, tipo 0), de

forma que cuando se tenga definida la cadena de animaciones se pueda interpolar la secuencia, creando movimientos de unión entre los distintos signos a representar.

Para el tratamiento de las animaciones complejas se dispone de varias herramientas, mediante las cuales se puede modificar y ajustar la secuencia, para lograr el efecto deseado. Una de las opciones de las que se dispone para el tratamiento de animaciones complejas consiste en poder borrar y añadir posiciones en cualquier parte de la secuencia, sin más que indicar la posición en la cual se pretenden insertar las posiciones, o las posiciones entre las cuales se desea borrar.

El sistema ofrece la posibilidad de presentar en pantalla el estado de la animación simple última con la que se ha estado trabajando o la animación compleja que se está creando. De esta forma es sencillo cargar o editar pequeños fragmentos antes de añadirlos a la secuencia final. Se dispone de dos controles mediante los cuales es posible este cambio de contexto visual. Para todas estas funciones, se ofrece una interfaz gráfica completa, con todas las opciones clasificadas de forma sencilla e intuitiva para que el usuario lo utilice fácilmente.

1.5. Conclusiones.

Una vez finalizado el proyecto, se pueden destacar varias conclusiones. La primera de ellas hace referencia al estudio previo realizado sobre las herramientas comerciales. Dicho estudio, justifica que se haya elegido crear el avatar desde cero haciendo uso de las bibliotecas de OpenGL. Pese a que las distintas herramientas que se han probado ofrecen una variedad de funciones muy extensa, en muchas ocasiones sobrepasan las necesidades de este proyecto, incluyendo herramientas orientadas al diseño del comportamiento de los objetos 3D en el entorno (la escena), pero quizá resultan insuficientes para implementar las funciones básicas para la generación de animaciones de signos con tanto detalle como se ha realizado en el avatar creado en este proyecto. Las herramientas de los programas estudiados (Toon3D, Cool3D, 3DStudioMax, Lightwave, etc.), además, son complicadas de utilizar y requieren un aprendizaje previo. Por ello, se eligen las bibliotecas de OpenGL y sus herramientas para la creación de interfaces 3D, que permiten crear con detalle las partes más importantes, en este caso, las manos y los dedos, sin prestar atención a la escena del avatar. Se podría crear una escena, para hacer más amigable la ventana de presentación al usuario, pero es algo que no contribuye a mejorar la comprensión del signo que se está representando. Los programas que se han comentado anteriormente, están más orientados a la creación de videojuegos y aplicaciones complejas.

La utilización de OpenGL y C++ para la creación de la aplicación partiendo de cero, permite disponer de un software completo para la animación de gestos con control completo sobre el sistema. Esta característica es fundamental para el desarrollo del proyecto, ya que se necesita poder modificar, implementar funciones nuevas y controlar en general cada detalle en el proceso de creación de las animaciones. Estas funciones no se pueden lograr con ninguna de las posibilidades estudiadas previamente (no se dispone del código).

Mediante el desarrollo de este proyecto se ha conseguido implementar un entorno completo para la generación de posiciones y animaciones que permiten representar los signos de la LSE (Lengua de Signos Española). Sin embargo, es posible generar signos/gestos de todo tipo (se pueden ver en la memoria ejemplos para la Lengua de Signos Americana (ASL), lo cuál hace de la aplicación un entorno completamente funcional. Dicho entorno permite la generación y gestión de posiciones, creación y gestión de animaciones de diferentes longitudes, pudiéndose enlazar entre ellas y crear secuencias complejas de signos, que permitan representar frases, conceptos, palabras, etc. La mayoría de los parámetros característicos para las animaciones son configurables, lo que hace de la aplicación una alternativa completamente adaptable a las necesidades de la animación que precise crear. La gestión de las animaciones se realiza a través de una interfaz sencilla de utilizar para el usuario, muy intuitiva, de forma que cualquier persona pueda ser capaz de generar posiciones, y crear animaciones.

Para mejorar la adaptabilidad al medio en el que se utilice, se ha implementado un módulo de generación de video, que permite guardar las animaciones en formato *.avi y representarlas directamente desde la aplicación en un reproductor multimedia. Además, se ofrece la posibilidad de elegir la codificación de video deseada, en función de las necesidades del usuario (más resolución o menor tamaño).

Como se vio en la introducción, el sistema forma parte de un sistema más complejo, en el cuál, la aplicación desarrollada permite representar los signos que resultan de la traducción realizada previamente, de voz a lengua de signos. Sin embargo, el sistema puede funcionar de forma independiente, al servicio de un logopeda que necesite crear e implementar nuevos signos, creando su propia biblioteca de animaciones. El entorno desarrollado puede, por lo tanto, utilizarse como módulo de representación para el sistema complejo descrito, y a su vez, servir de herramienta para la creación de posiciones y animaciones.

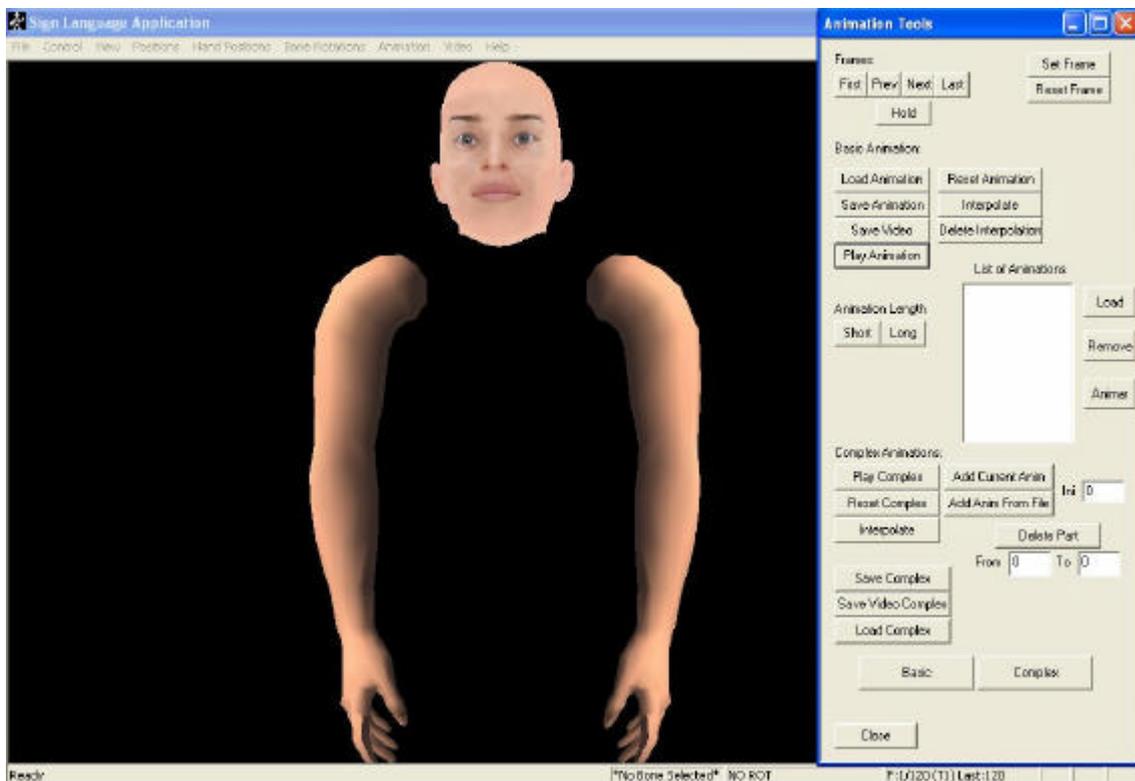


Figura 1. 3. Interfaz gráfica de la aplicación para la representación de signos mediante el avatar 'Sara'.

ORIGINALIDAD DEL PROYECTO FIN DE CARRERA: ENTORNO DE DESARROLLO PARA LA CREACIÓN DE ANIMACIONES CON AGENTES VIRTUALES 3D

Originalidad de la aplicación

En la actualidad son muchas las herramientas disponibles para el diseño de objetos en 3D, por el gran desarrollo que se ha producido en la creación de videojuegos, que cada vez son más reales, más complejos y ofrecen más opciones de interacción. Esta complejidad siempre depende de la finalidad de la aplicación, de qué grado de detalle se necesite y de qué grado de realismo se pretende alcanzar.

Entre las opciones disponibles, para el diseño de la aplicación objeto de este proyecto, se realizaron pruebas con diferentes herramientas de software. Algunas de ellas partían ya de objetos previamente creados, o incluso algunas disponían ya de un avatar completo. Se ha tratado también la opción de fabricar un avatar completo mediante el diseño de cada una de sus partes, por medio de herramientas de creación de gráficos. Las herramientas estudiadas han sido: 3DStudioMax, Lightwave 3D, Cool3D, Toon3D, It's Me.

Tras estudiar las distintas opciones comerciales, se puede concluir que ninguna de ellas se adapta a las necesidades del proyecto. Existen algunas ventajas que podrían aprovecharse satisfactoriamente para el desarrollo de la aplicación pero se presentan diversos inconvenientes que impiden su utilización.

Ventajas

Algunos de los sistemas descritos (3D Studio Max, Lightwave y Cool3D) presentan una gama de herramientas y funciones que permiten realizar proyectos de alto nivel, en cuanto a la creación de objetos y la animación de los mismos. Las interfaces son muy ergonómicas, se trata de programas muy profesionales, que ofrecen al usuario las posibilidades de implementación más avanzadas en el campo del diseño en 3D.

Son programas muy comercializados, que se utilizan en la industria de los videojuegos y la creación de películas de animación. Con ellos, se puede crear prácticamente todo tipo de objetos y escenas tridimensionales (algunos de ellos incluso pueden trabajar en dos dimensiones para algunas de las características).

Otra gran ventaja es que estos programas trabajan con lenguajes interpretados para el manejo de los objetos creados, es decir, por medio de scripts. Esto dota a los sistemas de una mayor abstracción.

En el caso del software Toon3D la principal ventaja es que se dispone del código fuente completo, y del manual de referencia escrito por el autor del programa. Las animaciones conseguidas presentan movimientos de gran realismo y las escenas creadas se pueden definir con bastante detalle. La iluminación y la modificación del punto de observación dotan a la animación de mayor dinamismo. La posición de la cámara puede cambiarse a medida que avanzan los fotogramas, de forma que se acentúe el movimiento del avatar creado. Esto permite acercar y alejar la representación creando efecto de película.

Inconvenientes

La complejidad de los sistemas descritos hace en muchos casos que la aplicación sea poco manejable para conseguir el objetivo del proyecto. Muchas de las funciones y herramientas no son útiles para el trabajo que se realiza, y lo que hacen es complicar la implementación. No son herramientas que un logopeda pueda utilizar para implementar animaciones. Debe ser algo

intuitivo para un usuario sin conocimientos de programación, que pueda utilizar el entorno de desarrollo de forma natural y sencilla. Los avatares predefinidos que proveen los programas como 3D Studio Max, Cool 3D y Toon 3D tienen un nivel de detalle adecuado para la creación de videojuegos o películas, pero no para el caso que se trata en este proyecto. Lo que se necesita es un avatar que muestre con especial precisión los movimientos de las manos, y en concreto, de los dedos. La parte inferior del avatar, por tanto, no es necesaria. Es más, es preferible que el avatar esté cerca del usuario, mostrando únicamente las partes del cuerpo implicadas en la representación de gestos de la lengua de signos, para que el observador del gesto se centre en lo representado.

Otro inconveniente de los programas descritos es que todos ellos utilizan formatos de archivo propietarios, que precisan de herramientas concretas para su utilización. Esto es un problema, que resta adaptabilidad a los objetos creados ya que no se pueden exportar a otras aplicaciones creadas por el usuario fácilmente. Igualmente, los programas comentados, al ser de uso profesional, tienen un coste muy elevado que si no se necesita la mayoría de las funcionalidades que se ofrecen, no merece la pena afrontar.

Se puede observar que las herramientas más potentes para la creación de gráficos en 3D son las que tienen un precio más elevado (3D Studio Max y Lightwave), lo cual hace inviable su utilización en este proyecto fin de carrera. Varias de ellas presentan un precio relativamente asequible, pero teniendo en cuenta el balance de ventajas e inconvenientes, se descarta su utilización.

Otro inconveniente que presentan las aplicaciones descritas (exceptuando Toon3D) es que no se dispone del código, y por lo tanto, no se pueden mejorar las herramientas ni modificar las funciones para personalizar el resultado. No son programas de libre distribución.

El caso del Toon3D es distinto, ya que, pese a que se posee el código completo (se ha adquirido) éste no funciona correctamente. Presenta tantos problemas de programación que resulta más sencillo crear una aplicación de cero a intentar comprender todo el software para solucionar los problemas encontrados. Como se ha comentado, se realizaron varias modificaciones, fruto del estudio del código durante varias semanas, hasta llegar a la conclusión de que no merecía la pena en función de los resultados que se preveía alcanzar.

Por todos estos motivos, se ha optado por la creación propia de una aplicación nueva, con un avatar propio, desarrollado completamente en el marco de este proyecto.

Originalidad del tema

Las tecnologías del habla guardan una relación cada vez más estrecha con los agentes virtuales animados. La combinación de estas dos tecnologías permite mejorar de forma importante el interfaz entre la persona y la máquina. Los usuarios pueden interactuar con el agente animado usando el lenguaje natural que es una de las vías más importantes de comunicación que tiene el ser humano con el mundo exterior. Para ello es imprescindible conocer las necesidades y preferencias de los usuarios respecto al trato y apariencia que quieren recibir de dichos agentes virtuales. En la actualidad, gran número de científicos en el mundo entero están investigando y desarrollando este tipo de sistemas que fusionan un agente virtual animado y las tecnologías del habla (reconocimiento, comprensión, síntesis de voz...). La evaluación de este tipo de sistemas está dando unos resultados tremendamente positivos hasta el momento en lo referente a humanización y capacidad de automatización del interfaz persona-máquina.

Una virtud de este tipo de sistemas es la gran variedad de servicios que ofrecen y la capacidad de adaptarse a diferentes escenarios. Uno de estos posibles escenarios es el que trata este proyecto: el uso de los sistemas tecnológicamente híbridos voz-imagen para la ayuda al colectivo de personas sordomudas. En este caso concreto resulta imprescindible la incorporación

en el sistema de un agente virtual animado como el diseñado en este proyecto, que no solo humanice el interfaz persona-máquina sino que permita literalmente la comunicación entre la máquina y la persona sorda.

A propósito del Lenguaje de Signos, es importante destacar que a pesar de los intentos desarrollados para su universalización y estandarización, lo cierto es que hoy en día no existe dicha uniformidad y en cada país existe una variedad diferente del lenguaje de signos, tremendamente influenciada por el idioma hablado en dicha región. Si las estructuras de dicho lenguaje se ven afectadas por el idioma del país en el que viva la persona sorda, el léxico de dicho idioma posee un altísimo grado de variación, no sólo entre diferentes ciudades dentro del mismo país, sino también entre diferentes escuelas, asociaciones o grupos de personas sordomudas. En base a esto, se puede destacar la originalidad del sistema en cuanto a que la biblioteca de signos que puede representar es ilimitada, pudiendo incluir la representación para palabras nuevas en cualquier momento, con sólo definir las posiciones como se indica en la descripción del sistema. De esta forma, se atiende a la necesidad de dinamismo y actualización que la Lengua de Signos requiere, unida a la gran variedad geográfica de las estructuras y el léxico de la lengua de signos que ha motivado una de las características más importantes del sistema de comunicación que se pretende desarrollar en este proyecto: la flexibilidad.

El objetivo principal de este proyecto es el apoyo a aquellas personas cuyo trabajo o dedicación exija comunicación directa con personas sordomudas. Para ello se ha llevado a cabo el desarrollo de un sistema de comunicación unidireccional que permita que dicha persona diga una frase ante un micrófono y que la persona sorda perciba dicha frase signada en el Lenguaje de Signos Español. En este proyecto concretamente se lleva a cabo la creación del agente virtual encargado de representar los signos y concatenarlos para formar frases complejas.

Es cierto, que en la actualidad hay algunos proyectos en marcha centrados también en la representación de signos con avatares. Uno de los ejemplos más representativos es el proyecto eSIGN, (Essential Sign Language Information on Government Networks) constituye uno de los esfuerzos más importantes para el desarrollo de herramientas de apoyo a la generación de contenidos para personas sordas. En este proyecto, uno de los principales resultados ha sido la generación de un avatar en 3D (VGuido) (capaz de representar cualquier signo de la lengua de signos) y un entorno visual para el desarrollo de animaciones fácilmente. Estas herramientas están orientadas a traducir contenido web. El resultado del proyecto eSIGN está funcionando en páginas web oficiales de Alemania, Países Bajos y el Reino Unido. Sin embargo el avatar desarrollado VGuido presenta restricciones a la hora de crear nuevos gestos, no es de libre distribución, con lo cual no se puede modificar. De cara al usuario, VGuido no tiene el grado de flexibilidad que sí se alcanza en la creación de 'Sara', el avatar desarrollado en este Proyecto Fin de Carrera. Uno de los aspectos más elaborados en el proyecto, es la precisión de las manos. En las figuras 1.4 y 1.5 se muestran algunas posiciones para ambas manos, que muestran el grado de precisión alcanzado.



Figura 1.4. Algunas posiciones para la mano derecha.



Figura 1.5. Algunas posiciones para la mano izquierda.

RESULTADOS DEL PROYECTO FIN DE CARRERA: ENTORNO DE DESARROLLO PARA LA CREACIÓN DE ANIMACIONES CON AGENTES VIRTUALES 3D

Como resultado del Proyecto Fin de Carrera, se ha implementado una interfaz, cuya vista simple es la siguiente:

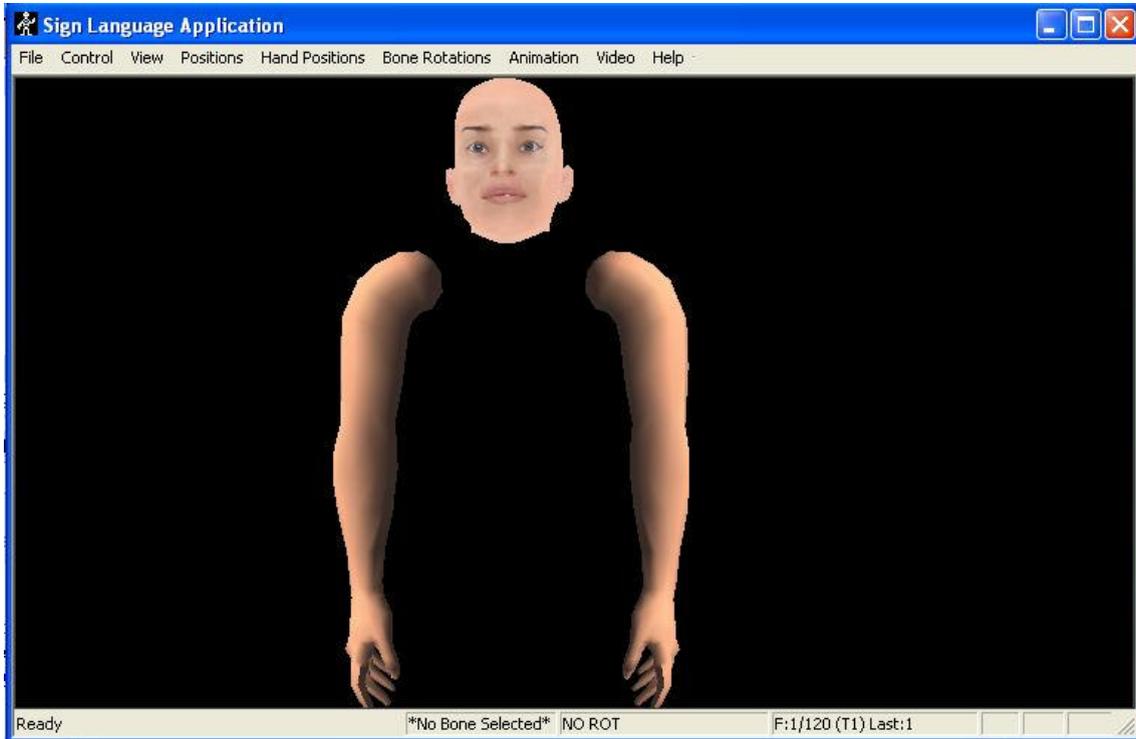


Figura 1. 6. Vista simple de la interfaz de usuario.

En la ventana principal, se presenta el modelo en la posición de reposo. Esta será la posición por defecto siempre que se inicie la aplicación. Partiendo de este inicio, se utiliza la herramienta de animaciones, que es la que se muestra en la siguiente figura:

Mediante esta herramienta, es posible crear posiciones, seleccionando los huesos de diversas formas, y asignarlas a determinados fotogramas. Una vez definidas las posiciones, se puede proceder a crear una animación simple. Esto se realiza partiendo de las posiciones creadas en determinados fotogramas, mediante una interpolación lineal. Estas animaciones pueden guardarse, reproducirse o formar parte de una secuencia mayor, lo que se denomina en el ámbito de este proyecto: 'Animación Compleja'. Las posiciones parciales de cada una de las manos, se puede guardar también, para poder reutilizarlas cuando sea necesario.

En la herramienta diseñada se pueden observar distintos bloques. Se ha estructurado de forma que se intuya fácilmente cuáles son las opciones para tratamiento de fotogramas, animaciones simples, animaciones complejas y tratamiento de resultados.

Los resultados obtenidos son las animaciones correspondientes a las secuencias de palabras de la Lengua de Signos.

Una característica importante de la aplicación diseñada es que las animaciones se guardan con toda la información referente a fotogramas, es decir, es posible cargar de nuevo una animación guardada y conocer cuáles son los fotogramas interpolados, de forma que si se desea se pueda deshacer la interpolación. Esta característica dota al sistema de una mayor eficacia a la hora de crear nuevos gestos, ya que facilita en gran medida la creación de gestos a partir de los

gestos que ya se encuentran en la biblioteca. Cuantos más gestos y animaciones se encuentren guardados, más sencillo será crear gestos nuevos, ya que en la Lengua de Signos es muy común que distintos gestos compartan posiciones similares. También contribuye a facilitar el trabajo, la opción comentada anteriormente, que permite guardar posiciones de cada una de las manos.

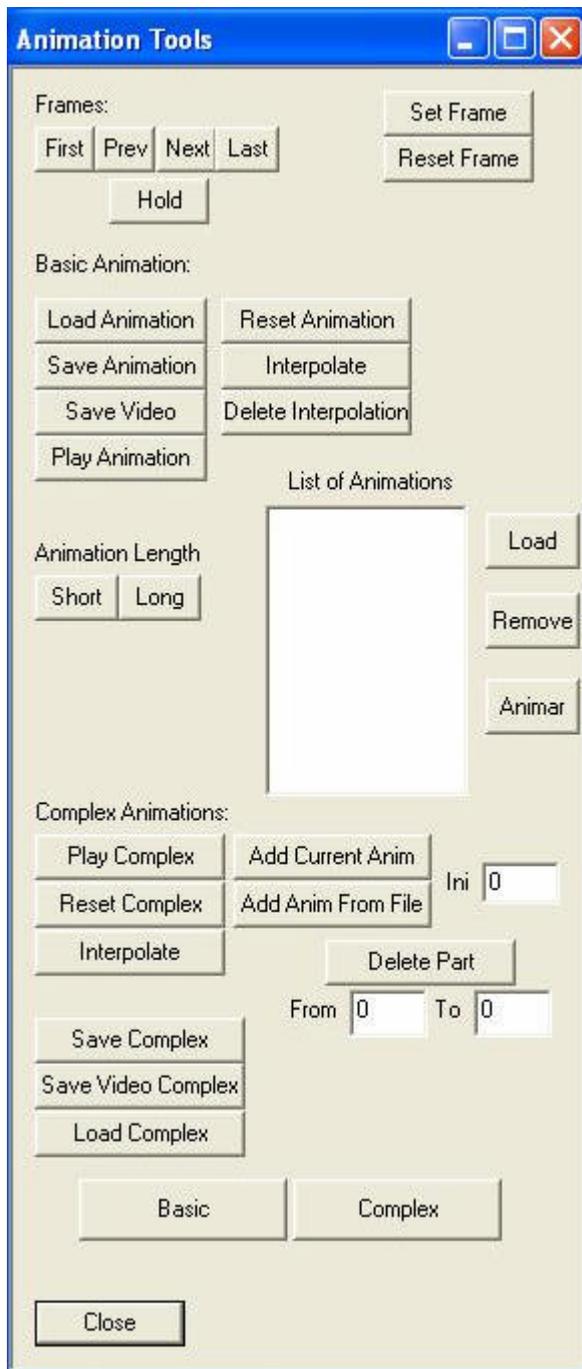


Figura 1.7. 'Animation Tools'

La herramienta desarrollada permite así el control absoluto sobre los gestos que se desean programar, así como cualquier detalle de la posición, en especial de las manos y los distintos dedos. Este era uno de los objetivos más importantes a conseguir, ya que con las opciones estudiadas sobre programas comerciales, no se disponía de una precisión suficiente para poder crear los gestos de la Lengua de Signos de una forma clara y natural. Gracias a una definición de

Gracias a esta característica del sistema, se puede crear lo que sería una biblioteca de manos, las cuales podrían utilizarse en numerosos gestos. Hay que tener en cuenta que en la Lengua de Signos, la mayoría de los signos centran su representación en la manos, por lo tanto debe ser la parte más elaborada para el avatar. Gracias a la posibilidad de rotar los huesos, se puede colocar la mano en una posición de mejor visibilidad para quien esté desarrollando la posición, de forma que una vez que se haya conseguido dicha posición, se pueda guardar la mano creada y resetear el esqueleto para poder centrarse en el resto de huesos. Una vez creada la posición para el resto de huesos, se puede importar la mano más adecuada de la biblioteca y crear así la posición deseada.

Aparte de la posibilidad de guardar las animaciones creadas en un formato que contiene toda la información de los fotogramas (*.ani), otro resultado importante es que cualquier animación que sea representada por el avatar puede guardarse como un archivo de video. El video resultante muestra la secuencia de fotogramas. Al guardar un video, se ofrece al usuario la posibilidad de elegir el formato más adecuado, en base a los codecs de los cuales se disponga.

Se ha obtenido por lo tanto, un sistema funcional, completo y de fácil comprensión, que permite definir los gestos de la Lengua de Signos y representarlos de forma natural y fluida, gracias a la posibilidad de interpolar varios gestos consecutivos, de forma que se presenten como una secuencia continua.

gran resolución para los dedos, las secuencias interpoladas presentan un grado de realismo bastante elevado.

La cabeza se ha incorporado únicamente con la posibilidad de realizar movimientos sencillos, como un objeto 3D cuasi-estático, que únicamente sirve de referencia para los gestos realizados con los brazos.

Se han planteado diversas mejoras que podrían complementar el sistema en un futuro. Entre esas mejoras se pueden enumerar las siguientes:

- Considerar diferentes tipos de interpolación, además de la interpolación lineal. En este proyecto se ha considerado que cuando se realiza la interpolación entre dos fotogramas de tipo creados por el usuario, la variación de la posición entre ambas posiciones se reparte de forma equitativa entre los fotogramas que se encuentren entre ellas, produciendo un movimiento uniforme. Las nuevas interpolaciones consisten en repartir la variación en forma de curva, de forma que el movimiento se realice más rápido al principio o al final, dependiendo del caso.
- Mejoras para el aspecto del avatar, como puede ser la incorporación de un torso. En este sentido la mejor solución es importar un objeto en formato *.3DS del mismo modo que se importa la cabeza, y definirlo en una posición fija, de forma que la parte superior de los brazos (que no permite traslaciones) quede unida al objeto importado.
- Mejoras en la unión de animaciones. En la actualidad, cuando se unen dos animaciones simples se insertan 10 fotogramas vacíos, para realizar una interpolación entre ambas y crear así un movimiento natural para pasar de una a otra. Una forma más eficiente de implementar esta inserción de fotogramas vacíos, sería considerar la distancia entre las dos posiciones a unir. De esta forma, se podría asignar el número de fotogramas basado en umbrales dependiendo de cuál sea la variación a cubrir entre el último fotograma de la primera animación y el primero de la siguiente. Esto mejoraría la velocidad del movimiento, mejorando a su vez la continuidad entre las animaciones a unir.
- Mejoras sobre el método de asignación de pesos a los puntos de la malla de piel. Actualmente, el método implementado asigna peso con valor 1 al hueso más cercano, calculada la distancia entre el punto de la malla y el punto central del hueso, y peso nulo al resto de huesos. Se puede crear una función de asignación de pesos más elaborada, en la cuál el peso para el resto de huesos no sea nulo, sino que se asignen de forma ponderada para dar más realismo a los brazos. La asignación ponderada, dotaría de un valor de peso al hueso en proporción a la distancia al punto que se esté evaluando. Así, los huesos cercanos tendrán más peso que los lejanos, en los cuales el valor del peso tenderá a cero.

Como resultado de la realización del Proyecto Fin de Carrera y gracias a los conocimientos adquiridos se han realizado varias publicaciones, que se detallan en el documento Anexo (se incluyen también las publicaciones), cuyos títulos son los siguientes:

1. SignTutor: AN INTERACTIVE SYSTEM FOR SIGN LANGUAGE TUTORING (IEEE Multimedia, 2008)
2. PROTOTIPO DE TRADUCCIÓN DE VOZ A LENGUA DE SIGNOS ESPAÑOLA (IV Jornadas de Tecnología del Habla, Zaragoza, 2006)
3. PRIMERA EXPERIENCIA DE TRADUCCIÓN AUTOMÁTICA DE VOZ EN LENGUA DE SIGNOS (Telecom I+D, 2006)
4. SIGN LANGUAGE TUTORING TOOL (Dubrovnik, Enterface 06)
5. MULTIMODAL CARICATURAL MIRROR (Mons, Enterface 05)

APLICABILIDAD DEL PROYECTO FIN DE CARRERA: ENTORNO DE DESARROLLO PARA LA CREACIÓN DE ANIMACIONES CON AGENTES VIRTUALES 3D

Uno de los principales objetivos del planteamiento de este Proyecto Fin de Carrera es que sea un sistema flexible, adaptable y aplicable a distintos entornos. Como se ha descrito anteriormente, las características del sistema permiten un alto nivel de adaptabilidad y por tanto facilitan su aplicabilidad en distintos ámbitos.

Aplicabilidad como herramienta independiente

Como se ha indicado, la herramienta puede funcionar de forma individual. Como sistema individual puede considerarse como un generador de animaciones que permite también generar videos con las mismas. El sistema de creación de nuevos gestos a partir de posiciones, permite incrementar la biblioteca de gestos guardados de forma ilimitada, por lo tanto el vocabulario disponible se encuentra en constante crecimiento.

Al crearse las animaciones a partir de cualquier posición definida por el usuario, el sistema puede representar gestos de cualquier Lengua de Signos. Un ejemplo de este hecho, son los signos que han sido implementados en el Summer Workshop Enterface '06, celebrado en Dubrovnik, en el cual se han programado diversos gestos pertenecientes a la Lengua de Signos Americana, necesarios en la implementación del sistema desarrollado. En las Figuras 1.8 y 1.9 se puede ver parte de la secuencia de la animación para representar dos gestos: 'door' y 'fast'.

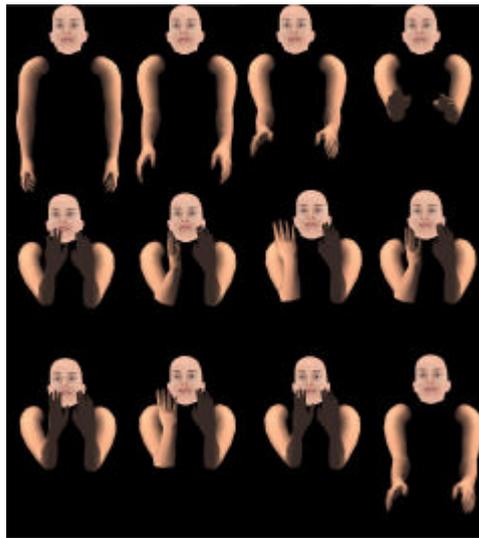


Figura 1.8. Gesto para 'door'

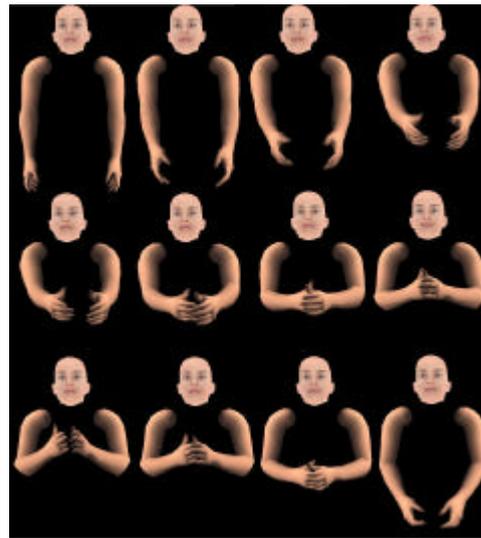


Figura 1.9. Gesto para 'fast'

Otra gran ventaja es la facilidad con la que se pueden añadir nuevas características al sistema. La aplicación es fácilmente modificable, y adaptable a cualquier plataforma, ya que al haber sido programada sobre librerías de OpenGL, bastaría con modificar la creación de ventanas simples en el código e indicar el caso correspondiente, para la plataforma que se pretenda utilizar.

Aplicabilidad como parte de otro sistema complejo

Como parte del sistema completo de traducción que se ha detallado, se trata de una aplicación que, aunque en este caso haya sido diseñada para el ámbito de renovación del Documento Nacional de Identidad, podría utilizarse para cualquier otro servicio, previo tratamiento del vocabulario que se requiera en cada caso. En este sentido, el funcionamiento de 'Sara' sería el mismo, ya que su función es tomar la secuencia de signos traducida y representarla

concatenando los sucesivos gestos. Lo único que se necesitaría sería definir el conjunto de palabras nuevas y crear los gestos que no se encuentren en la biblioteca. Como se ha comentado anteriormente, cuantos más gestos contenga la biblioteca, más sencilla resultará la incorporación de otros nuevos, gracias a la reutilización de posiciones de las manos y los brazos.

Como parte de otros sistemas, cabe destacar la incorporación de la aplicación al proyecto SignTutor: AN INTERACTIVE SYSTEM FOR SIGN LANGUAGE TUTORING. En dicho proyecto se ha realizado un sistema que pretende enseñar Lengua de Signos, ofreciendo la posibilidad de corregir e instruir al alumno. El sistema muestra un gesto, que el alumno debe imitar en frente de una cámara de vídeo. El gesto realizado por el alumno es analizado y comparado con patrones obtenidos de una gran base de datos (grabada también durante el Workshop). El gesto es comparado con el original mostrado, para tomar la decisión de si debe darse como válido o no. Del gesto del alumno se obtienen ciertos parámetros tales como la distancia entre hombros, la longitud de los brazos, el punto medio de la cabeza, que se pasan como parámetro de entrada al avatar. Este, adaptará los parámetros a sus propias proporciones físicas, ofreciendo al alumno una imitación algo exagerada del gesto realizado, de forma que pueda comparar lo que él mismo ha realizado con el original, comprendiendo el porqué de la decisión tomada por el sistema: TRUE o FALSE.

La construcción del avatar ha sido una unión de los brazos de 'Sara' y la cabeza convenientemente implementada de Xface. Gracias a la flexibilidad del código de 'Sara' se ha eliminado fácilmente la cabeza, para poderla unir a la nueva cabeza, que en este caso sí representa emociones y cierta flexibilidad de movimiento. En la figura 1.10 se puede observar la aplicación desarrollada así como la decisión tomada según el gesto realizado por la usuaria (cuadro inferior izquierdo). En el cuadro superior derecho aparece el tutor de Lengua de Signos, y en la parte inferior derecha el avatar imitando el gesto realizado.



Figura 1.10. Sign Tutoring Tool

Gracias a la posibilidad de guardar las animaciones en formato de vídeo y reproducirlo desde la interfaz, la aplicación resulta muy sencilla de incorporar a distintos sistemas. Se puede configurar el código para que siempre se guarde el video con el mismo formato, o que cada vez que se guarde una animación en formato *.ani se guarde también un video de forma automática. Este planteamiento es ventajoso siempre que no se utilice un formato de video de cuadros sin comprimir, ya que el proceso de almacenamiento del video, si este tiene un tamaño considerable puede implicar un elevado coste de tiempo innecesario. La aplicación final se ha configurado de forma que el usuario pueda acceder a las opciones 'Save Video' y 'Save Video Complex' (según

sea una animación simple o compleja), y guardar la animación en curso. Automáticamente se mostrará un menú desplegable con los formatos disponibles. Estos formatos vienen dados por los codificadores que el usuario tenga instalados. Si se desea un formato que no está disponible en la aplicación, se deberá instalar el codificador correspondiente en el sistema y no será necesario modificar nada en el módulo de video de la herramienta ya que la detección de los formatos disponibles se realiza de forma automática al ejecutar la creación de video desde la interfaz. En la figura 1.11 se puede observar esta característica.



Figura 1.11. Elección del formato de video

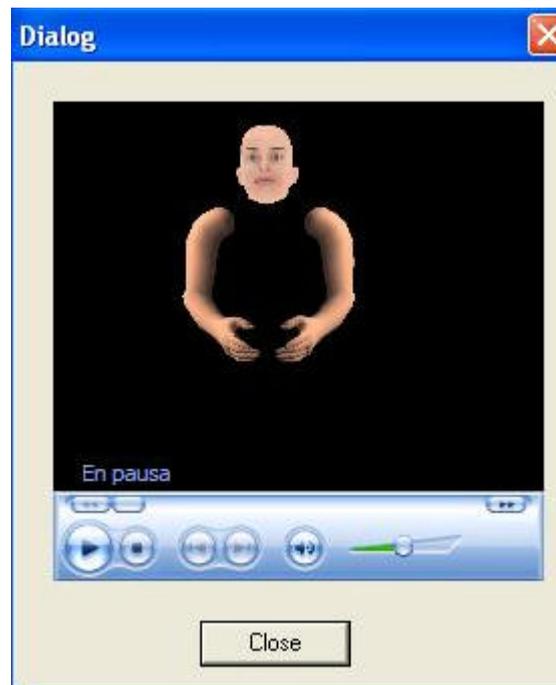


Figura 1.12. Reproductor de video de la interfaz

Como conclusión final, se puede decir que el sistema es de gran aplicabilidad para mejorar el acceso a los servicios para las personas sordas, tanto como aplicación independiente como aplicación integrada dentro de un sistema más complejo. Es fácilmente adaptable y flexible ante la posible incorporación de mejoras.

PUBLICACIONES RELACIONADAS CON EL PROYECTO FIN DE CARRERA: ENTORNO DE DESARROLLO PARA LA CREACIÓN DE ANIMACIONES CON AGENTES VIRTUALES 3D

Ana Huerta Carrillo

PUBLICACIONES:

El trabajo realizado en el Proyecto Fin de Carrera ha sido aplicado directamente en las siguientes publicaciones, cuyo contenido se adjunta.

SignTutor: AN INTERACTIVE SYSTEM FOR SIGN LANGUAGE TUTORING.

Oya Aran, Ismail Ari, Alexandre Benoit, Ana Huerta Carrillo, François-Xavier Fanard, Pavel Campr, Lale Akarun, Alice Caplier, Michele Rombaut, Bulent Sankur. IEEE MultiMedia 2008 (Aceptado, pendiente de publicación).

PROTOTIPO DE TRADUCCIÓN DE VOZ A LENGUA DE SIGNOS

ESPAÑOLA. E. Ibáñez, A. Huerta, R. San-Segundo, L.F. D'Haro, F. Fernández, R. Barra. IV Jornadas de Tecnología del Habla, Zaragoza, 2006.

PRIMERA EXPERIENCIA DE TRADUCCIÓN AUTOMÁTICA DE VOZ

EN LENGUA DE SIGNOS. A. Huerta, E. Ibáñez, R. San-Segundo, F. Fernández, R. Barra, L.F. D'Haro. Telecom I+D, 2006.

SIGN LANGUAGE TUTORING TOOL. Oya Aran, Ismail Ari, Alexandre Benoit, Ana Huerta Carrillo, François-Xavier Fanard, Pavel Campr, Lale Akarun, Alice Caplier, Michele Rombaut, Bulent Sankur. Proceedings of eINTERFACE 2006, The Summer Workshop on Multimodal Interfaces, hosted by Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia. (Dubrovnik 2006).

'Sign Language Tutoring Tool' (INTERFACE '06, Dubrovnik, Croacia), es una herramienta para enseñar la lengua de signos a un alumno. En este caso, la aplicación ha funcionado como módulo de representación para los brazos del avatar final. El alumno observa un video del tutor realizando el signo, y su imitación es grabada por una cámara. Los datos del gesto realizado por el alumno se procesan y se decide si el gesto ha sido correctamente representado o no. Los datos del signo se guardan en un archivo de texto, que el módulo de representación tomará como entrada y el signo será representado por el avatar.

MULTIMODAL CARICATURAL MIRROR. Martin O., Adell J., Huerta A., Kotsia I., Savran A., Sebbe R. Proceedings of eINTERFACE 2005, The Summer Workshop on Multimodal Interfaces, Faculté Polytechnique de Mons, Belgium.

La realización de este proyecto, ha permitido también adquirir los conocimientos necesarios para participar en otras actividades, como la colaboración en el proyecto 'Caricatural Mirror' (INTERFACE '05, Mons, Bélgica), en el cuál se desarrollaron caracterizaciones de cara y brazos, basadas en bibliotecas OpenGL, así como análisis de expresiones faciales y gestos de la lengua de signos de diferentes países. También se realizaron

colaboraciones en el proyecto 'Combined Gesture-Speech Analysis and Synthesis' (ENTERFACE '05), construyendo un avatar que generara varios gestos concretos ante la aparición de ciertos eventos.

Asimismo, los conocimientos adquiridos con la realización del proyecto, han permitido una colaboración puntual en la publicación:

COMBINED GESTURE-SPEECH ANALYSIS AND SÍNTESIS. Mehmet Emre Sargin, Ferda Ofli, Yelena Yasinnik, Oya Aran, Alexey Karpov, Stephen Wilson, Yucel Yemez, Engin Erzin and A. Murat Tekalp. Proceedings of eNTERFACE 2005, The Summer Workshop on Multimodal Interfaces, Faculté Polytechnique de Mons, Belgium.

ESTANCIAS:

Interface summer workshop (2005): Participación en el workshop de verano de la red de excelencia SIMILAR sobre Interfaces Multimodales en Mons (Bélgica). Se desarrollaron siete proyectos diferentes sobre integración de aplicaciones basadas en voz e imagen. <http://www.interface.net/>. Como resultado de esta estancia se publicaron los dos últimos artículos del apartado anterior, el primero de ellos como co-autora y el segundo como colaboración puntual.

Interface summer workshop (2006): Participación en el workshop de verano de la red de excelencia SIMILAR sobre Interfaces Multimodales en Dubrovnik (Croacia). Se desarrollaron nueve proyectos diferentes sobre integración de aplicaciones basadas en voz e imagen. <http://www.interface.net/>. Como resultado de esta estancia se publicaron tanto el artículo para IEEE como el artículo correspondiente al proyecto realizado, en ambos como co-autora.