

INTÉRPRETE DE LENGUAJE DE SIGNOS EN ESPAÑOL MULTIDISPOSITIVO

Fernando López, Javier Tejedor, Daniel Bolaños, José Colás
Departamento de Ingeniería Informática, E.P.S. Universidad Autónoma de Madrid
Carretera de Colmenar Viejo km 15, 28049 Madrid, SPAIN
fj.lopez@uam.es, javier.tejedor@uam.es, daniel.bolanos@uam.es, jose.colas@uam.es

RESUMEN

En este artículo presentamos un transcriptor de texto a lenguaje de signos distribuido y multidispositivo. La presentación al usuario final del lenguaje de signos es realizada por un personaje animado en tres dimensiones. Este transcriptor está creado para adaptar su salida a la capacidad de proceso del dispositivo receptor. Por lo que puede ser utilizado por un usuario en un ordenador personal para transcribir una página Web, o en un teléfono móvil para transcribir una conversación (utilizando un reconocedor de voz). La flexibilidad del sistema permite adaptarlo a varios idiomas o usarlo como un simple elemento para mejorar una interfaz multimedia.

PALABRAS CLAVE

Accesibilidad, Lenguaje de signos, Interprete, Multidispositivo, Distribuido.

1. INTRODUCCIÓN

Hasta la fecha se han utilizado diversos métodos para llevar a cabo la representación infográfica de un intérprete de signos. Debemos destacar el proyecto ViSiCAST (Bangham, 2000) como un proyecto de referencia para la transcripción al lenguaje de signos. Este sistema está basado en la captura de los movimientos realizados por un interprete humano y su posterior segmentación en unidades semánticas. Otra de las alternativas presentadas (Papadogiorgaki, 2004) es la realización de dicha animación utilizando una secuencia de "Body Animation Parameters" del estándar MPEG-4.

La práctica totalidad de los sistemas de conversión de texto a lenguaje de signos utilizan la representación HamNoSys (Hanke, 2004) como un paso intermedio en el proceso de traducción. Para posteriormente llevar a cabo una transformación a un formato SiGML (Elliot, 2001) o SWML (da Rocha, 2003) capaz de ser interpretado con facilidad por un ordenador y convertido en una animación 3D. Para la generación de la animación, estos sistemas utilizan un módulo basado en VRML (ISO/IEC 14772-1:1997). El sistema de animación está relacionado con la definición H-ANIM (Roehl, 1998), especializada en animación de modelos humanos.

Estos sistemas están desarrollados para funcionar sobre máquinas con capacidad gráfica media-alta. En el caso de ViSiCAST, sobre ordenadores dedicados, como es la interpretación de noticias para un telediario, o la síntesis de lenguaje de signos en una oficina de correos. Todos estos sistemas han sido creados para hacer funcionar el sistema de manera fluida. Pero los visores de VRML no son fácilmente adaptables a los dispositivos móviles y además requieren una gran potencia de cálculo para aportar la tasa de cuadros por segundo que se requiere para que la animación sea fluida.

En este documento se describe una nueva concepción de transcriptor, diseñando un nuevo sistema de almacenamiento de la información necesaria para realizar un signado básico. Este nuevo formato está destinado a poder adaptar el sistema y el modo representación de manera dinámica al dispositivo que esté utilizando el usuario. De esta manera podrá ser utilizado tanto en un ordenador personal como en un teléfono móvil.

2. OBJETIVOS

El objetivo principal de este sistema es ofrecer un intérprete virtual de lenguaje de signos (Kennaway, 2002) (Kennaway, 2004). Dada la gran variabilidad de dispositivos que se utilizan hoy en día, dicho intérprete deberá estar a la altura de todos los dispositivos sobre los que pueda ser solicitado. Por este modo se debe crear un método de almacenamiento para la definición de los signos con un nivel adecuado de abstracción para permitir su almacenamiento de manera adecuada y adaptada al dispositivo. Dicho nivel de abstracción debe permitir que la representación gráfica sea lo más directa posible para que la animación sea fluida y fácilmente entendida.

3. DESCRIPCIÓN DEL SISTEMA

Como ya se ha mencionado en el apartado de objetivos, el desarrollo del sistema implica dos módulos conceptuales distintos. El primero de estos módulos ha desarrollado un método de almacenamiento de la información del lenguaje de signos en una base de datos relacional. Dicha estructuración de la información aporta la flexibilidad necesaria para hacer posible la labor del segundo módulo. Este segundo módulo se encarga de balancear la carga de trabajo entre el servidor de traducción y el dispositivo utilizado por el usuario.

3.1. Descripción de la base de datos

Con el objetivo de hacer posible la escalabilidad del sistema se ha optado por generar una estructura relacional para la representación de los signos. Formalmente (Muñoz, 1997), los signos se definen por siete componentes. De los cuales, cinco son básicos, y dos pueden expresarse como una adecuada combinación de los anteriores.

De los cinco elementos básicos, cuatro hacen referencia a las manos. Éstos son la configuración o forma que adopta la mano y la flexión de los dedos, la orientación de la palma y dirección del dedo índice, la localización de las manos del intérprete visto desde el frente y por último el plano de articulación o la distancia horizontal de las manos del intérprete a su cuerpo. El quinto elemento básico es la expresión corporal y especialmente facial.

Los otros dos elementos formativos son el punto de contacto entre la mano dominante del intérprete y la otra mano u otra parte del cuerpo del intérprete y el movimiento o variación de los cuatro elementos básicos referidos a las manos. Estos elementos no serán representados en la base de datos como entidades, debido a que pueden descomponerse en una combinación simultánea o secuencial de los cuatro elementos básicos utilizados en las manos.

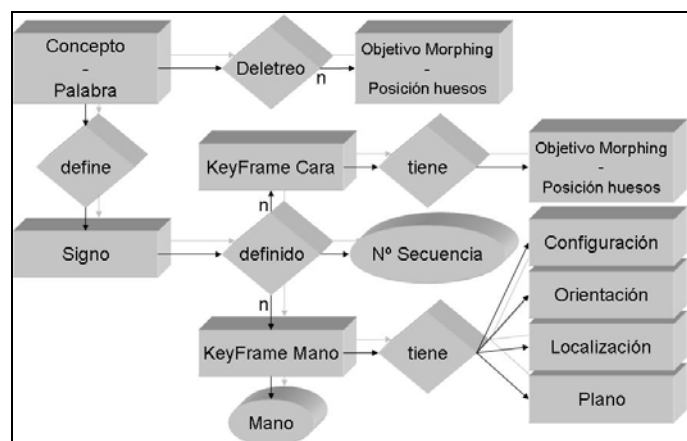


Figura 1. Esquema de la base de datos para la representación de la lengua de signos

Por las razones antes expuestas, hemos diseñado una estructura relacional para almacenar toda esta información (Figura 1). En español existen setenta y una “configuraciones” distintas. Utilizadas todas por la mano dominante del intérprete y en su mayoría también a la otra mano. Se podría definir los valores de ambas manos, ciento cuarenta y dos elementos, o modelar solo una mano y luego aplicar una transformación de simetría. La “orientación” se traduce en un ángulo de rotación a la muñeca, y la “localización” y el “plano” se traducen en las coordenadas XYZ. Con estos cuatro elementos tenemos una posición estática de la mano y el brazo. Si definimos de manera simultánea los valores para las dos manos, lograremos el elemento “punto de contacto”, y si definimos una secuencia en el tiempo, la interpolación de valores y posiciones hará que se genere el elemento “movimiento”.

La aplicación de la estructura de esta base de datos resulta inmediata sobre un esqueleto jerárquico utilizado para la animación de un modelo humano en 3D. Dicho esqueleto ha sido diseñado específicamente para su uso con esta base de datos. Este esqueleto no sigue la estructura anatómica de los huesos, sino que ha sido desarrollado para que cada una de las entidades sea asignada en exclusiva a un solo hueso, mientras que un hueso puede recibir valores de ninguna, una o varias entidades.

3.2. Escalabilidad y multidispositivo

Los principales elementos que se deben tener en cuenta para adaptar el contenido al dispositivo son, la propia capacidad del dispositivo y el ancho de banda disponible para el envío de información. Utilizando estos dos elementos, el sistema tiene dos opciones para prestar el servicio de transcripción. En la Figura 2 podemos observar la arquitectura esquematizada del sistema.

Para poder atender las peticiones de los dispositivos que soliciten dicho servicio de traducción, el sistema dispone de dos variantes. La primera de ellas consiste en utilizar una base de datos potente y realizar la generación de la animación correspondiente a la transcripción en el propio servidor. Una vez generada, o mientras se genera, dicho video es tratado como si de una entrada de video se tratara. Y es enviado utilizando el encapsulado y formato de transmisión usado en videoconferencia. El dispositivo del usuario sólo tendría que reproducir dicha secuencia de vídeo.

El segundo método propuesto, se basa en el envío de la secuencia de signos que se han transcrito al dispositivo del usuario. Este dispositivo deberá hacer uso de una base de datos, más simplificada y adaptada a la capacidad de este dispositivo, y utilizando ambos elementos generar la animación y presentarla al usuario. Para realizar la generación de la animación se propone utilizar la API “3D Java mobile” (JSR-184, 2003:a). La elección de esta API se basa en dos motivos principalmente, el formato de archivos que gestiona, el M3G (JSR-184, 2003:b), es completamente compatible con la utilización de esqueletos para animar mallas de polígonos. Y el segundo motivo es que esta API ya se encuentra implementada en muchos móviles puesto que, es la base para los juegos que incorporan.

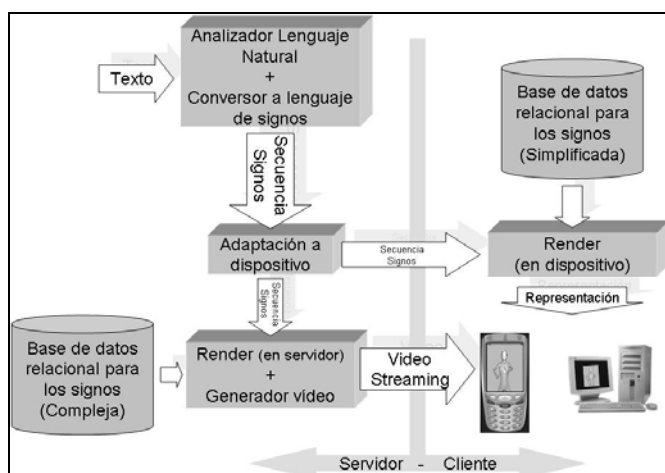


Figura 2. Descripción de la arquitectura del sistema

Como puede observarse de la descripción de las opciones, la primera de ellas permite su utilización en dispositivos con escasa capacidad de proceso para llevar a cabo los cálculos necesarios para la animación 3D. Por el contrario, hace un uso importante del ancho de banda para el envío de un “stream” de video. La segunda de las opciones es complementaria, se recomienda para su uso en entornos de red en los que el ancho de banda esté muy restringido y los dispositivos posean una buena capacidad para el proceso 3D, cualquier ordenador con aceleradora gráfica. En caso de que el servicio al usuario no esté restringido por ninguna de estas limitaciones, el sistema podrá balancear su carga para no colapsar el servidor.

4. CONCLUSIÓN Y TRABAJO FUTURO

Este transcriptor se basa en la utilización de entornos de aplicación y canales de comunicación estándares. Esto otorga al sistema una gran compatibilidad con la mayoría de los dispositivos que pueden encontrarse en el mercado. La API “3D Java mobile” se encuentra en la mayoría de los móviles de última generación, esto reduce el impacto y evita tener que solicitar al los fabricantes de dispositivos que incluyan una nueva librería gráfica. La escalabilidad del sistema permite la adaptación óptima al contexto.

El desarrollo de este sistema se está realizando en el contexto de una plataforma multimodal. El objetivo de este módulo es aportar la capacidad para generar una respuesta al usuario utilizando el lenguaje de signos. La primera versión del sistema estará preparada para soportar un diccionario básico del lenguaje de signos en español, tanto la ampliación de dicho diccionario como la adaptación a otros idiomas será tratada después.

Las pruebas de usuario comenzarán una vez terminadas las pruebas funcionales. Estas pruebas se plantean en dos grandes bloques funcionales y la segregación de tres tipos de usuarios objetivo. Las pruebas que se realizarán serán similares a las que puedan realizarse en un examen de una academia de idiomas. En una primera fase se partirá de un pequeño mensaje en texto y el usuario deberá comprobar que el avatar representa correctamente el mensaje. Estas pruebas evaluarán corrección morfológica, sintáctica y semántica. En la segunda fase de las pruebas se omitirá suministrar al usuario el mensaje previamente y se le solicitará que anote el mensaje que ha entendido. Es decir, se solicitará una traducción inversa. En cuanto a los grupos de usuarios entre los que se presentan diferencias son: El primer grupo estará conformado por profesores e intérpretes del lenguaje de signos sin ninguna discapacidad auditiva reseñable. El segundo por personas que han desarrollado una sordera o que siendo sordos de nacimiento han tenido una educación en colegios para oyentes. Y el tercero estará constituido por personas sordas de nacimiento que han sido educadas en un entorno para personas no oyentes.

El esquema de este conversor presenta la base para llevar a cabo un estudio del impacto y la aceptación de la inserción de un factor de imperfección en la representación de los signos. Es decir, los intérpretes humanos no realizan siempre de la misma forma cada uno de los signos. Dicho estudio medirá el impacto de las variaciones en la ejecución de los parámetros formativo en niveles de inteligibilidad y aceptación de los usuarios.

REFERENCIAS

- Bangham, JA. et al, 2000. Virtual Signing: Capture, Animation, Storage and Transmission – an Overview of the ViSiCAST Project. En IEE Seminar on *Speech and language processing for disabled and elderly people*. Londres, England.
- Elliot, R. et al. 2001. D5-2 SiGML definition. *Working document*, ViSiCAST Project.
- Hanke, T. 2004. HamNoSys – Representing Sign Language Data in Language Resources and Language Processing Contexts. En *Workshop on the occasion of the 4th International Conference on Language Resources and Evaluation, LREC 2004*. Lisbon, Portugal
- ISO/IEC 14772-1:1997 and ISO/IEC 14772-2:2004 — Virtual Reality Modeling Language (VRML)
- JSR-184 2003 Mobile 3D Graphics API for J2ME™ [En línea]. <http://www.jcp.org/en/jsr/detail?id=184>
- JSR-184 2003 File Format for Mobile 3D Graphics API. [En línea]. <http://www.j2medev.com/api/jsr184/file-format.html>
- Kennaway, R., 2002. Synthetic Animation of Deaf Signing Gestures. In I. Wachsmuth and T. Sowa (eds.), *Gesture and sign language in human-computer interaction. International Gesture Workshop, GW 2001*. Londres, England, 146-157.

- Kennaway, R., 2004. Experience with and requirements for a gesture description language for synthetic animation. In A. Camurri and G. Volpe (eds.) *Gesture Based Communication in Human-Computer Interaction, 5th International Gesture Workshop 2003*, Genova, Italy. 300-311.
- Muñoz Baell, I. M. 1997. La querología aplicada a la enseñanza de la LSE. Comunicación presentada en el *I Curso de Lingüística Aplicada a la LSE*. Madrid. Spain.
- Papadogiorgaki, M. et al, 2004. Synthesis of Virtual Animation from SWML using MPEG-4 Body Animation Parameters. En *Workshop on the occasion of the 4th International Conference on Language Resources and Evaluation, LREC 2004*. Lisbon, Portugal
- da Rocha, A.C., Pereira, G. 2003. Supporting deaf sign languages in written form on the web. TALN and RECITAL 2003 Vol. 2 pp. 193-202. SWML, Official Site of . <http://swml.ucpel.tche.br/swml.htm>
- Roehl, B., 1998. Specification for a Standard VRML Humanoid. *Definido como ISO/IEC FCD 19774:200x* <http://h-anim.org/Specifications/H-Anim1.1/>